

**Практические занятия**  
по курсу «Современные проблемы прикладной математики и  
наукоемкого программного обеспечения»  
модуль «Введение в безопасность систем искусственного интеллекта»  
2025 г.

## 1. Цель работы

Изучить методы устойчивого оценивания параметра сдвига распределений непрерывных случайных величин.

## 2. Содержание работы

**Уровень выполнения № 1 (минимальный), максимум 10 баллов за выполнение.**

1. Разработать программу, которая реализует:

1.1) генерацию наборов данных с заданным в варианте чистым распределением и искаженных наборов данных; искажение генерировать, заменяя в чистой выборке 10%-30% наблюдений выбросами; получить выбросы можно путем увеличения исходных наблюдений в несколько раз (несколько десятков раз), наличие выбросов необходимо контролировать визуально;

1.2) вычисление выборочных характеристик: среднего арифметического, дисперсии, коэффициентов асимметрии и эксцесса;

1.3) вычисление выборочной медианы, урезанного среднего.

2. Провести проверку генератора чистого распределения путем сравнения выборочных характеристик (см. пп. 1.2) с их теоретическими значениями на выборке большого объема ( $N$  порядка  $10^4$ – $10^7$ ); как альтернативу можно использовать какие-либо критерии согласия, в том числе с использованием стороннего программного обеспечения.

3. Для выборок с чистым и искаженным распределениями (минимум по три выборки с одинаковыми значениями всех параметров) вычислить оценки параметра сдвига в виде

- среднего арифметического;
- выборочной медианы;
- урезанного среднего с разными уровнями (как минимум три обязательных значения 0.05, 0.10, 0.15).

Рекомендуемый объем выборки  $N$ : 200–1000. Сравнить устойчивость оценок для используемых распределений по их отклонению от истинного значения, сопоставить результаты сравнения со свойствами функций влияния оценок.

4. Оформить отчет по проделанной работе. Отчет должен содержать постановку задачи, описание и результаты работы, выводы, текст программы. Для каждой выборки привести графики (значения наблюдений как функции от номеров наблюдений). Привести график с функциями влияния для среднего арифметического, выборочной медианы и урезанных средних. Можно привести «приблизительную»

функцию влияния урезанного среднего при произвольном положительном значении  $q$ .

#### 5. Защитить работу.

**Уровень выполнения № 2 (стандартный), максимум 20 баллов за выполнение.**

##### 1. Разработать программу, которая реализует:

1.1) генерацию наборов данных с заданным в варианте чистым распределением и засоренным распределением, использовать засоряющие распределения, совпадающие с чистым с точностью до значений параметров сдвига и масштаба;

1.2) вычисление выборочных характеристик: среднего арифметического, дисперсии, коэффициентов асимметрии и эксцесса;

1.3) вычисление выборочной медианы, урезанного среднего, оценки максимального правдоподобия и обобщенных радикальных оценок параметра сдвига модельного распределения;

1.4) вычисление математического ожидания, дисперсии, коэффициентов асимметрии и эксцесса случайной величины с засоренным распределением по заданным характеристикам чистого и засоряющего распределений.

2. Провести проверку генератора чистого и засоренного распределений путем сравнения выборочных характеристик (см. пп. 1.2) с их теоретическими значениями на выборках большого объема ( $N$  порядка  $10^4$ – $10^7$ ); как альтернативу можно использовать какие-либо критерии согласия, в том числе с использованием стороннего программного обеспечения.

3. Для выборок с разными видами распределений вычислить следующие оценки параметра сдвига:

- среднее арифметическое;
- выборочная медиана;
- оценка максимального правдоподобия;
- урезанное среднее с разными уровнями (как минимум три обязательных значения 0.05, 0.10, 0.15);
- обобщенные радикальные оценки с разными значениями параметра (как минимум три обязательных значения 0.1, 0.5, 1).

Использовать выборки, имеющие следующие виды распределений:

- чистое распределение;
- засоренное распределение с симметричным засорением (равные сдвиги у чистого и засоряющего распределений, масштаб у засоряющего больше в 2-4 раза, чем у чистого);
- засоренное распределение с асимметричным засорением (сдвиги у чистого и засоряющего распределений отличаются на 2-4 стандартных отклонения, масштаб у засоряющего распределения не меньше, чем у чистого).

При выборе параметров засорения ориентироваться на график чистой, засоряющей и засоренной плотностей. Рекомендуемый уровень засорения  $\varepsilon$ : 0.1–0.4. Рекомендуемый объем выборки  $N$ : 200–1000. Сравнить устойчивость оценок для распределений указанных видов (минимум по три выборки с одинаковыми значе-

ниями всех параметров) по их отклонению от истинного значения, сопоставить результаты сравнения со свойствами функций влияния оценок.

4. Оформить отчет по проделанной работе. Отчет должен содержать постановку задачи, описание и результаты работы, выводы, текст программы. Для каждого рассмотренного засоренного распределения привести график, на котором сравнить чистую, засоряющую и засоренную плотности. Привести график с функциями влияния для всех использованных оценок.

5. Защитить работу.

**Уровень выполнения № 3 (продвинутый), максимум 30 баллов за выполнение.**

Предполагает выполнение всех работ уровня № 2, но выполнение п. 3 провести не на отдельных выборках, а путем исследования всех оценок методом Монте-Карло. Исследование провести для выборок различного объема  $N$  (наибольшее значение  $N$  выбрать в зависимости от времени вычислений), и засоренных распределений с различными уровнями засорения  $\varepsilon$ .

### 3. Методические указания

#### 3.1. Оценки

В основе классических статистических процедур лежат некоторые предположения о свойствах наблюдений, например, их независимости, однородности (одинаковой распределенности). Известный метод максимального правдоподобия требует знания закона распределения наблюдений.

На практике некоторые предположения могут не выполняться, они могут выступать в роли идеализации, удобного приближения реальной ситуации. В результате соответствующие методы теряют оптимальность, становятся неустойчивыми.

Для решения этой проблемы разработаны различные подходы, приводящие к устойчивым процедурам.

Рассмотрим задачу оценивания параметра сдвига  $\theta$  симметричного распределения. Модель наблюдений имеет вид

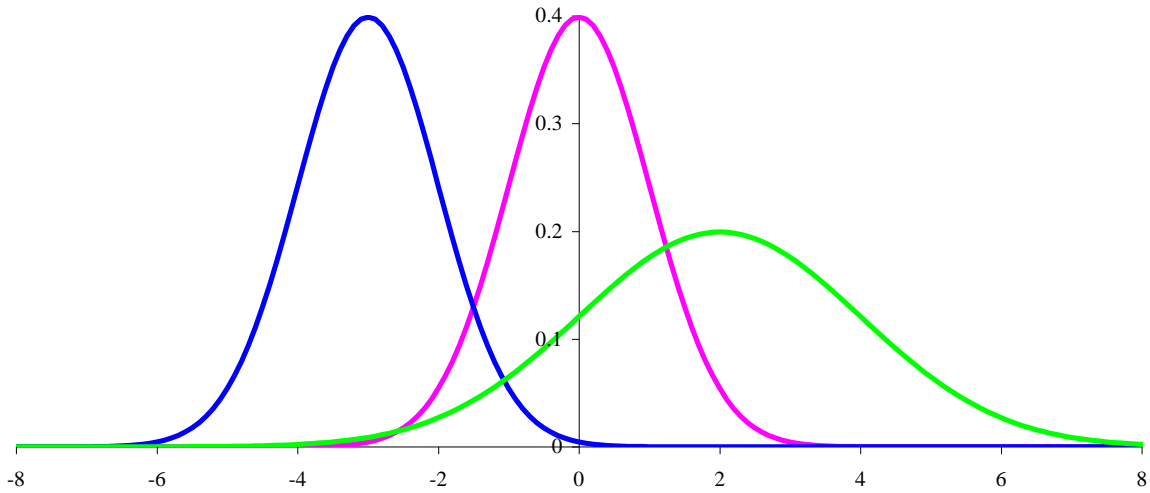
$$y_i = \theta + e_i, \quad i = 1, \dots, N,$$

где  $y_i$  –  $i$ -е наблюдение исследуемой случайной величины,  $\theta$  – параметр сдвига,  $e_i$  –  $i$ -е значение случайной величины с нулевым параметром сдвига,  $N$  – количество наблюдений.

Учтем, что помимо параметра сдвига распределение может содержать не оцениваемый параметр масштаба  $\lambda$ . В этом случае плотность распределения можно представить в виде

$$f(y, \theta, \lambda) = \frac{1}{\lambda} f\left(\frac{y - \theta}{\lambda}\right), \quad (1.1)$$

где  $f(x)$  – плотность **стандартного распределения** случайной величины, имеющей нулевой сдвиг и единичный масштаб.



**$M$ -оценка**  $\hat{\theta}$  скалярного параметра  $\theta$  находится путем минимизации функции вида

$$Q(\theta) = \frac{1}{N} \sum_{i=1}^N \rho(y_i, \theta),$$

где  $\rho$  – функция потерь.

Альтернативное определение  $M$ -оценки требует решения неявного уравнения

$$\frac{1}{N} \sum_{i=1}^N \psi(y_i, \hat{\theta}) = 0,$$

где  $\psi$  – оценочная функция.

**Среднее арифметическое** вычисляется по формуле

$$\bar{y} = \frac{1}{N} \sum_{i=1}^N y_i,$$

где  $y_i$  – значение  $i$ -го наблюдения. Среднее арифметическое часто может использоваться в качестве оценки параметра сдвига симметричного распределения. В частности, оно является оценкой максимального правдоподобия параметра сдвига нормального распределения.

Отметим также, что среднее арифметическое является  $M$ -оценкой с функцией потерь

$$\rho(y, \theta) = (y - \theta)^2,$$

и оценочной функцией

$$\psi(y, \theta) = c(y - \theta),$$

где  $c$  – ненулевая константа, с точностью до которой определяется оценочная функция ( $c$  может зависеть от параметров распределения, но не от  $\theta$ ).

Его функция влияния имеет вид

$$\text{IF}(y) = y - \theta.$$

**Медиана** последовательности определяется как такое число, что одна половина элементов не больше него, а другая – не меньше.

В статистике выборочная медиана определяется для нечетного числа наблюдений как  $[N/2+1]$ -я порядковая статистика, где  $[ \ ]$  – целая часть числа, а для четного числа наблюдений – как среднее арифметическое  $[N/2]$ -й и  $[N/2+1]$ -й порядковых статистик ( $i$ -я порядковая статистика –  $i$ -е наблюдение в выборке после упорядочения ее элементов по возрастанию).

Выборочная медиана может использоваться в качестве оценки параметра сдвига симметричного распределения. В частности, она является оценкой максимального правдоподобия параметра сдвига распределения Лапласа.

Отметим также, что выборочная медиана является  $M$ -оценкой с функцией потерь, недифференцируемой в нуле:

$$\rho(y, \theta) = |y - \theta|,$$

и оценочной функцией, разрывной в нуле:

$$\psi(y, \theta) = c \operatorname{sign}(y - \theta).$$

Её функция влияния имеет вид

$$\operatorname{IF}(y) = \frac{\lambda \operatorname{sign}(y - \theta)}{2f(0)}.$$

**Урезанное среднее** вычисляется следующим образом: отбрасывается некоторое (одинаковое) количество минимальных и максимальных наблюдений в выборке; на основе оставшихся наблюдений находится среднее арифметическое. Точнее, урезанное среднее уровня  $\alpha$  вычисляется следующим образом

$$\bar{y}_\alpha = \frac{1}{N - 2k} \sum_{i=k+1}^{N-k} y_i,$$

где  $0 \leq \alpha < 0.5$ ,  $k = [N\alpha]$ ,  $y_i$  –  $i$ -я порядковая статистика,  $[ \cdot ]$  – целая часть числа. Предельными случаями оценки при  $\alpha = 0$  и  $\alpha \rightarrow 0.5$  являются среднее арифметическое и выборочная медиана.

Урезанное среднее может использоваться в качестве оценки параметра сдвига симметричного распределения. Урезанное среднее не является  $M$ -оценкой, его функция влияния для симметричного распределения пропорциональна оценочной функции оценки Хьюбера и имеет вид

$$\operatorname{IF}(y) = \frac{1}{1 - 2\alpha} \begin{cases} -q, \frac{y - \theta}{\lambda} < -q \\ \frac{y - \theta}{\lambda}, \left| \frac{y - \theta}{\lambda} \right| \leq q, \\ q, \frac{y - \theta}{\lambda} > q \end{cases}$$

где  $q$  – квантиль порядка  $1 - \alpha$  стандартного распределения (распределения с нулевым параметром сдвига и единичным параметром масштаба), т.е.  $\int_{-\infty}^q f(t) dt = 1 - \alpha$ .

Однако, пороговая точка оценки равна лишь  $\alpha$ .

**Замечания.**

1. При вычислении функции влияния, если определение теоретических квантилей вызывает затруднения, можно воспользоваться выборочными квантилями

$$\hat{q} = \frac{y_{[(1-\alpha)N]+1} - \theta}{\lambda} \text{ для выборки большого объема из чистого распределения.}$$

2. В высокоуровневых средах разработки обычно имеются функции вычисления медианы и урезанного среднего. Для пользующихся языком C++ имеется программа вычисления этих характеристик с помощью алгоритма `nth_element` стандартной библиотеки.

**Оценка максимального правдоподобия** является  $M$ -оценкой с функцией потерь

$$\rho(y, \theta) = \rho\left(\frac{y - \theta}{\lambda}\right) = -\ln f\left(\frac{y - \theta}{\lambda}\right),$$

и оценочной функцией

$$\psi(y, \theta) = \psi\left(\frac{y - \theta}{\lambda}\right) = c \frac{f'\left(\frac{y - \theta}{\lambda}\right)}{f\left(\frac{y - \theta}{\lambda}\right)}.$$

Функция влияния оценок имеет вид

$$IF(y) = \frac{-\lambda f'\left(\frac{y - \theta}{\lambda}\right) / f\left(\frac{y - \theta}{\lambda}\right)}{\int_{-\infty}^{\infty} [f'(t)]^2 / f(t) dt}.$$

**Обобщенные радикальные оценки** (в англоязычных источниках известны как оценки минимума гамма-дивергенции, оценки минимума логарифмической дивергенции степени плотности) также являются  $M$ -оценками и задаются функцией потерь

$$\rho(y, \theta) = \rho\left(\frac{y - \theta}{\lambda}\right) = -\frac{1}{f^{\delta}(0)} f^{\delta}\left(\frac{y - \theta}{\lambda}\right)$$

и оценочной функцией

$$\psi(y, \theta) = \psi\left(\frac{y - \theta}{\lambda}\right) = c f'\left(\frac{y - \theta}{\lambda}\right) f^{\delta-1}\left(\frac{y - \theta}{\lambda}\right),$$

где параметр  $\delta > 0$  регулирует степень робастности оценки – чем больше значение  $\delta$ , тем более устойчивой является оценка.

Функция влияния обобщенных радикальных оценок имеет вид

$$IF(y) = \frac{-\lambda f'\left(\frac{y - \theta}{\lambda}\right) f^{\delta-1}\left(\frac{y - \theta}{\lambda}\right)}{\int_{-\infty}^{\infty} [f'(t)]^2 f^{\delta-1}(t) dt}.$$

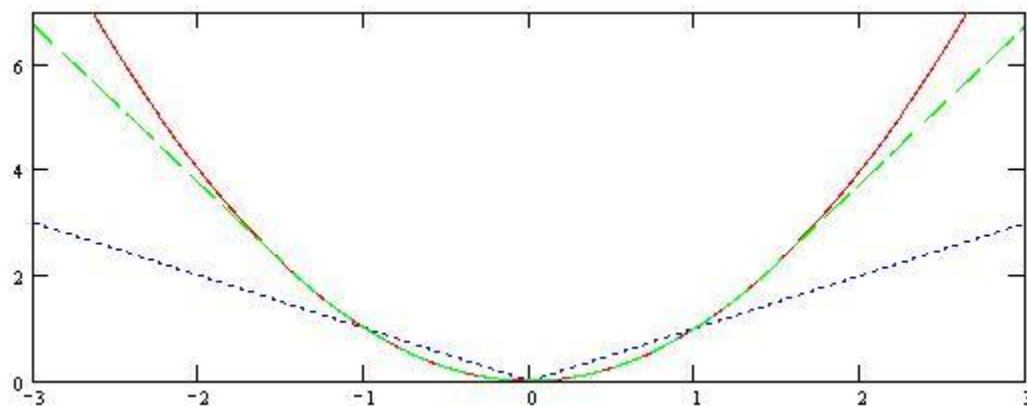
### ***Замечания о вычислениях.***

1. При вычислении оценки максимального правдоподобия и обобщенных радикальных оценок удобнее решать задачу оптимизации. Часто решение оказывается неединственным, в этом случае необходимо либо искать глобальный оптимум, либо использовать хорошее начальное приближение. В качестве последнего могут выступать либо истинное значение параметра, либо выборочная медиана.

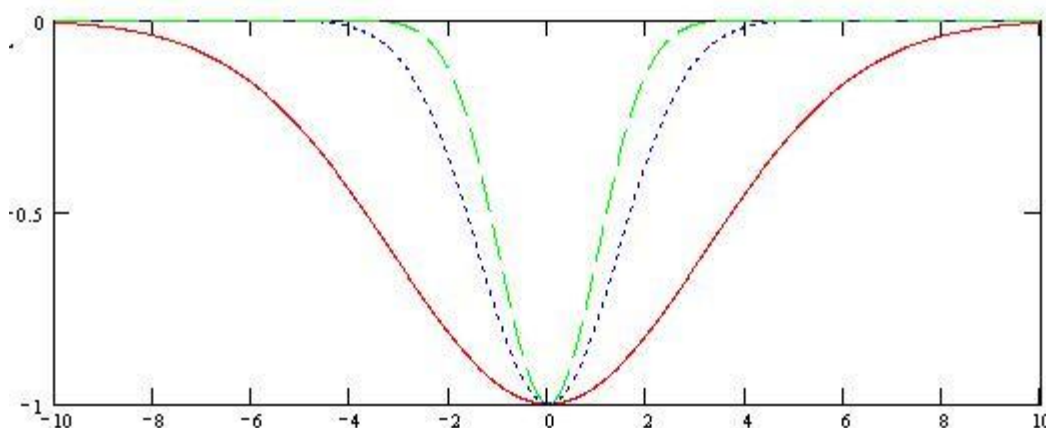
2. Нет необходимости устанавливать слишком высокую точность решения задачи оптимизации, достаточно точности  $10^{-4}$ – $10^{-5}$ .

3. Для повышения эффективности вычислений целесообразно учитывать, что плотности содержат нормирующие множители, от оцениваемого параметра  $\theta$  не зависящие, однако, нередко содержащие сложно вычисляемые специальные функции. Все это позволяет упрощать вид функций потерь и оценочных функций, что может приводить к заметному ускорению вычислений.

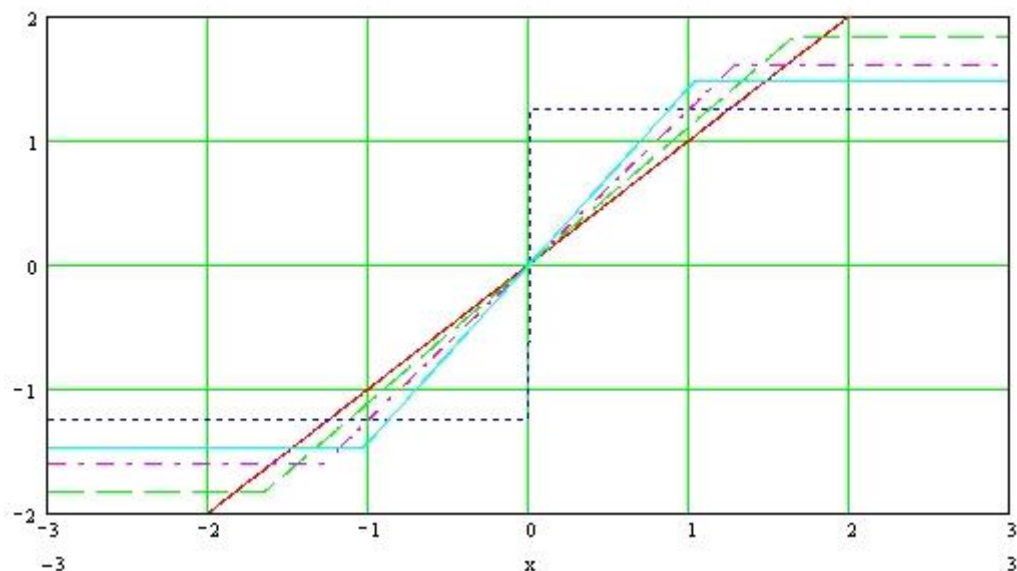
4. При построении графиков функции влияния и графиков плотностей по оси абсцисс должны откладываться произвольные значения (например, в узлах какой-то детерминированной сетки), а не значения из выборки.



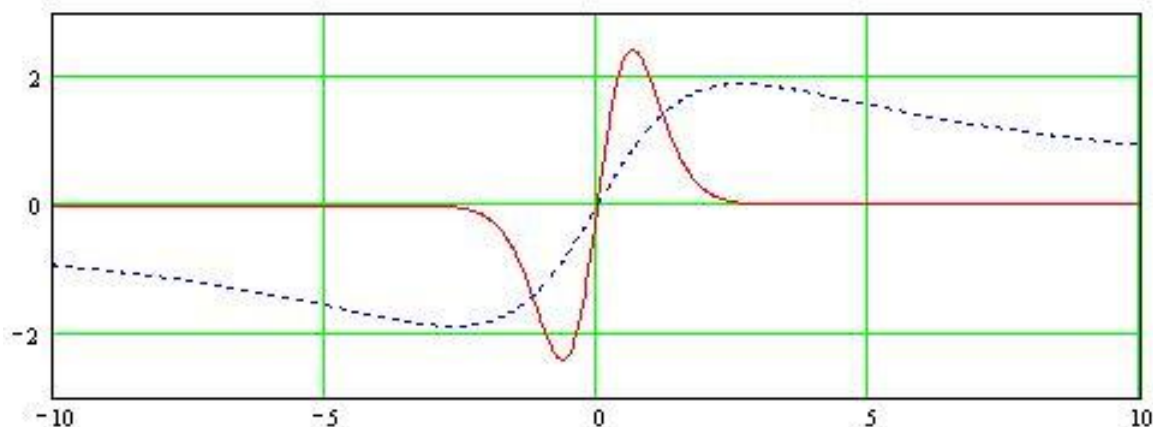
Функции потерь среднего (красная), медианы (синяя) и оценки Хьюбера (зеленая)



Функции потерь обобщенных радикальных оценок с разными значениями параметра (для красной линии минимален)



Функции влияния среднего (красная), медианы (синяя) и урезанных средних



Функции влияния обобщенных радикальных оценок с разными значениями параметра (для красной линии максимален)

### Другие выборочные характеристики распределений случайных величин:

– выборочная дисперсия

$$\hat{D} = \frac{1}{N} \sum_{i=1}^N (y_i - \bar{y})^2 ;$$

– выборочный коэффициент асимметрии (для симметричного распределения коэффициент асимметрии равен нулю)

$$\hat{\gamma}_1 = \frac{\sum_{i=1}^N (y_i - \bar{y})^3}{N \cdot \hat{D}^{1.5}} ;$$

– выборочный коэффициент эксцесса (показатель островершинности распределения, в ряде случаев может быть использован для измерения тяжести хвостов)



$$\hat{\gamma}_2 = \frac{\sum_{i=1}^N (y_i - \bar{y})^4}{N \cdot D^2} - 3.$$

### 3.2. Методы моделирования случайных величин

Для решения разнообразных задач в области вероятностно-статистических исследований необходимо получать на компьютере реализации случайных величин, имеющих заданные распределения. Такой процесс принято называть моделированием случайной величины.

#### 3.2.1. Моделирование дискретной случайной величины (стандартный алгоритм)

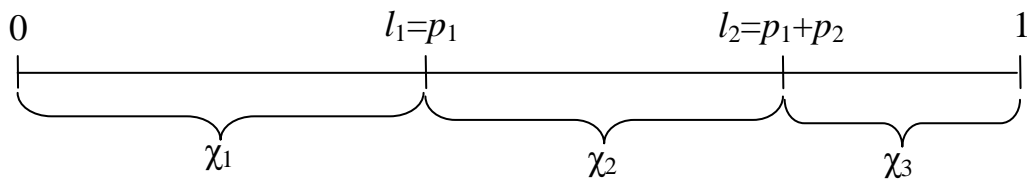
Рассмотрим моделирование дискретной случайной величины  $\xi$  с конечным числом значений  $1, 2, \dots, m$ , и

$$P\{\xi = i\} = p_i, \quad p_i > 0, \quad i=1, 2, \dots, m, \quad \sum_{i=1}^m p_i = 1.$$

В силу данных соотношений реализация того или иного значения случайной величины  $\xi$  означает розыгрыш события  $\xi = i$  с вероятностью  $p_i$ . Кроме того, из этих соотношений следует, что интервал  $(0,1)$  можно разбить на следующие полуинтервалы  $\chi_i$  длины  $p_i$ :

$$\chi_1=(0, l_1], \chi_2=(l_1, l_2], \chi_3=(l_2, l_3], \dots, \chi_{m-1}=(l_{m-2}, l_{m-1}], \chi_m=(l_{m-1}, 1),$$

где  $l_i = \sum_{j=1}^i p_j$ . Пример для случая  $m=3$  приведен на рисунке.



Если мы получим реализацию  $r$  случайной величины, равномерно распределенной в интервале  $(0, 1)$ , то будет справедливо

$$P\{r \in \chi_i\} = P\{l_{i-1} < r < l_i\} = \int_{l_{i-1}}^{l_i} dx = l_i - l_{i-1} = p_i,$$

поэтому реализацией  $\xi$  полагаем такое значение  $i$ , которое удовлетворяет соотношению

$$l_{i-1} \leq r < l_i.$$

Приходим к следующему алгоритму.

Шаг 1. Получить реализацию  $r$  случайной величины, имеющей равномерное распределение на интервале  $(0, 1)$ . Присвоить  $i = 1$ .

Шаг 2. Вычислить новое значение  $r$  по формуле

$$r = r - p_i.$$

Шаг 3. Если  $r \leq 0$ , то реализацией случайной величины будет  $i$ . Иначе присвоить  $i = i + 1$  и перейти на шаг 2.

### 3.2.2. Метод обратной функции для моделирования непрерывной случайной величины (стандартный алгоритм)

Перейдем к методам моделирования непрерывной случайной величины  $\xi$  с функцией распределения  $F(x)$  и плотностью  $f(x)$ . Область значений  $\xi$  – множество вещественных чисел  $\mathbf{R}^1$ .

В основе метода обратной функции лежит тот факт, что случайная величина  $F(\xi)$  имеет равномерное распределение на интервале  $(0, 1)$ . Поэтому, имея реализацию  $r$  случайной величины, равномерно распределенной в интервале  $(0, 1)$ , получить реализацию  $x$  случайной величины  $\xi$  можно из соотношения

$$r = F(x). \quad (2.1)$$

Здесь в первую очередь используется формула

$$x = F^{-1}(r). \quad (2.2)$$

Для ее применения желательно, чтобы функция  $F^{-1}(r)$  имела явный вид и была простой в использовании. В отсутствии явного вида у  $F^{-1}(r)$  решается уравнение (2.1), но это может оказаться слишком трудоемким. Еще более трудоемким будет случай, когда функция  $F(x)$  не имеет аналитического вида и решается уравнение

$$r = \int_{-\infty}^x f(u) du.$$

**Пример 1.** Пусть случайная величина  $\xi$  имеет экспоненциальный (показательный) закон распределения с функцией

$$F(x) = 1 - e^{-x}, \quad x > 0.$$

Найдя обратную по отношению к  $F$  функцию, получим реализацию  $x = -\ln(1 - r)$ . Так как  $1 - r$  имеет то же самое распределение, что и  $r$ , то удобнее при получении реализации случайной величины  $\xi$  пользоваться формулой

$$x = -\ln r.$$

### 3.2.3. Метод функциональных преобразований

Нередко оказывается, что целевая случайная величина  $\xi$  представима в виде функции  $g(x)$  от другой случайной величины  $\xi_1$ , имеющей удобные алгоритмы моделирования, т. е.

$$\xi = g(\xi_1).$$

В этом случае согласно методу функциональных преобразований для получения реализации случайной величины  $\xi$  можно воспользоваться функцией  $g(x)$  от реализации случайной величины  $\xi_1$ .

**Пример 2.** Случайная величина  $\xi$  получается из случайной величины  $\xi_1$  зеркальным отражением:  $g(x) = -x$ . Для зеркально отраженной экспоненциальной случайной величины, воспользовавшись результатами примера 1, получим реализацию

$$x = \ln r.$$

**Пример 3.** Сдвиг-масштабное преобразование (1.1). Пусть случайная величина  $\xi_1$  имеет некоторое стандартное распределение, т.е. распределение с нулевым параметром сдвига и единичным параметром масштаба. Случайная величина  $\xi$ , имеющая то же распределение, но с параметром сдвига  $\theta$  и параметром масштаба  $\lambda$ , получается из  $\xi_1$  линейным преобразованием  $g(x) = \theta + \lambda x$ . Тогда, имея реализацию  $x_1$  случайной величины  $\xi_1$ , получим реализацию  $x$  случайной величины  $\xi$  по формуле

$$x = \theta + \lambda x_1.$$

Метод функциональных преобразований допускает использование функций от нескольких случайных величин (например, от двух –  $g(\xi_1, \xi_2)$ ) и даже от потенциально неограниченного их количества –  $g(\xi_1, \xi_2, \dots, \xi_n, \dots)$ . Пример последнего случая будет рассмотрен в п. 3.2.5.

Вообще, метод функциональных преобразований оказывается не каким-то отдельным методом: когда все  $\xi_i$  имеют некое базовое распределение (общепринятым является использование в качестве него равномерного на интервале  $(0, 1)$ ), можно говорить об **общем принципе моделирования случайных величин** так, что все рассматриваемые здесь конкретные методы являются его реализациями. Например, в методе обратной функции, согласно формуле (2.2), используется функциональное преобразование  $g(x) = F^{-1}(x)$ .

### 3.2.4. Метод суперпозиции

Сначала рассмотрим **общий случай**. Пусть помимо целевой случайной величины  $\xi$  имеется вспомогательная случайная величина  $\eta$  с функцией распределения  $H(u)$  и областью значений  $U$ . Обозначим  $g(x|u)$  условную плотность случайной величины  $\xi$  при условии  $\eta = u$ . Тогда безусловное (маргинальное) распределение случайной величины  $\xi$  можно представить в виде смеси распределений, в частности, ее плотность имеет вид

$$f(x) = \int_U g(x|u) dH(u). \quad (2.3)$$

В этом случае реализации случайной величины  $\xi$  можно получать через моделирование случайного вектора  $(\eta, \xi)$ . Последнее сводится к двухэтапному последовательному моделированию элементов вектора: на первом этапе получаем реализацию  $u$  случайной величины  $\eta$  в соответствии с ее распределением, на втором этапе – реализацию случайной величины  $\xi$  в соответствии с условной плотностью  $g(x|u)$ .

Таким образом, метод суперпозиции полезен, если моделирование целевой случайной величины  $\xi$  по безусловному распределению менее эффективно, чем моделирование по некоторому условному распределению с предварительным моделированием вспомогательной случайной величины, определяющей условие.

Когда случайная величина  $\eta$  является непрерывной с плотностью  $h(u)$ , случайный вектор  $(\eta, \xi)$  имеет совместную плотность  $g(x|u)h(u)$ . В этом случае получаем распределение случайной величины  $\xi$  в виде *бесконечной смеси* с плотностью

$$f(x) = \int_U g(x|u)h(u)du$$

и следующий **метод интегральной суперпозиции**: сначала получить реализацию  $y$  случайной величины  $\eta$  в соответствии с плотностью  $h(u)$ , затем – реализацию случайной величины  $\xi$  в соответствии с условной плотностью  $g(x|y)$ .

Рассмотрим **метод дискретной суперпозиции**. Пусть случайная величина  $\eta$  является дискретной с распределением, определяющимся вероятностями

$$P\{\eta = i\} = p_i = h(i), \quad p_i > 0, \quad \sum_{i=1}^m p_i = 1, \quad i = 1, \dots, m, \quad (2.4)$$

$f_i(x)$  являются условными плотностями случайной величины  $\xi$  при условии  $\eta = i$ :

$$f_i(x) = g(x|\eta = i).$$

В этом случае формула (2.3) принимает вид

$$f(x) = \sum_{i=1}^m p_i f_i(x), \quad (2.5)$$

соответствующий *конечной смеси* распределений.

Имеем следующий алгоритм моделирования случайной величины  $\xi$  с плотностью (2.5):

- 1) получить реализацию  $i$  дискретной случайной величины  $\eta$  в соответствии с распределением (2.4);
- 2) получить реализацию случайной величины  $\xi$  в соответствии с плотностью распределения  $f_i(x)$ .

**Пример 4.** Моделирование случайной величины, имеющей распределение в виде смеси двух распределений – см. в п. 3.3.

Метод суперпозиции удобно использовать для **составных плотностей** вида

$$f(x) = \begin{cases} g_1(x), x \in \mathbf{X}^1 \\ g_2(x), x \in \mathbf{X}^2 \\ \dots \\ g_m(x), x \in \mathbf{X}^m \end{cases},$$

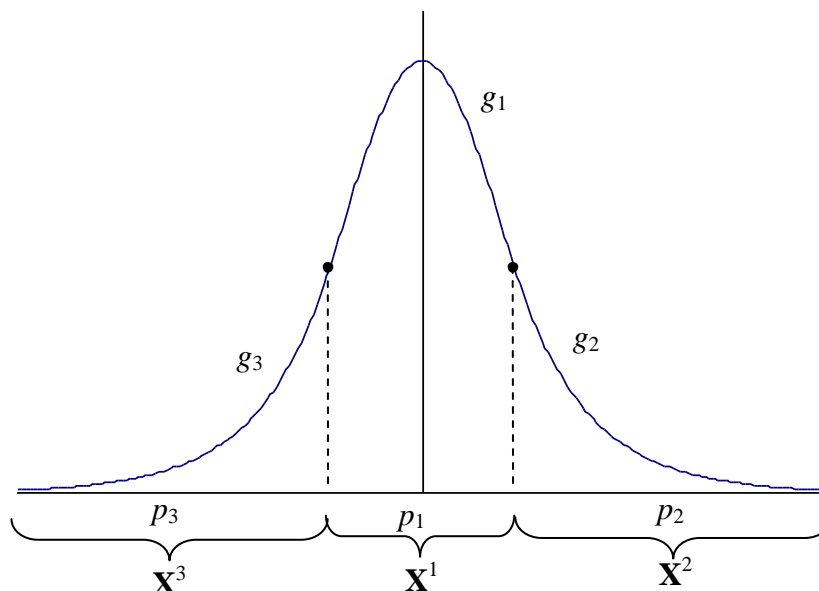
где подинтервалы  $\mathbf{X}^i$ ,  $i = 1, \dots, m$ , не пересекаются и составляют интервал  $\mathbf{X} \subseteq \mathbf{R}^1$  – область значений случайной величины, имеющей составную плотность.

Представим функции  $g_i(x)$  в виде

$$g_i(x) = p_i f_i(x), \quad i = 1, \dots, m,$$

где  $p_i = \int_{\mathbf{X}^i} g_i(x) dx > 0$ ,  $\sum_{i=1}^m p_i = 1$ ,  $f_i(x) = g_i(x) / \int_{\mathbf{X}^i} g_i(t) dt = g_i(x) / p_i$  – плотности не-

которых случайных величин, распределенных на интервалах  $\mathbf{X}^i$  и имеющих удобные алгоритмы моделирования. В результате плотность  $f(x)$  приобретает вид (2.5).



### 3.2.5. Метод исключения

В методе исключения используется вспомогательная случайная величина такая, что ее реализации, удовлетворяющие некоторому условию, отбираются в качестве реализаций целевой случайной величины, а реализации, указанному условию не удовлетворяющие, из рассмотрения исключаются. Вспомогательная случайная величина при этом должна иметь эффективный алгоритм моделирования, а ее плотность быть в некотором смысле (он будет рассмотрен ниже) близкой к плотности целевой случайной величины.

Пусть требуется смоделировать случайную величину  $\xi$ , распределенную на интервале  $\mathbf{X} \subseteq \mathbf{R}^1$  согласно плотности  $f(x)$ , которая пропорциональна заданной неотрицательной функции  $g(x)$ , т. е.

$$f(x) = \frac{g(x)}{K}, \quad K = \int_{\mathbf{X}} g(x) dx.$$

Для моделирования целевой случайной величины используется вспомогательная случайная величина  $\xi_1$ , распределенная согласно плотности

$$f_1(x) = \frac{g_1(x)}{K_1}, \quad K_1 = \int_{\mathbf{X}} g_1(x) dx.$$

При этом функция  $g_1(x)$  должна быть **мажорантой** функции  $g(x)$ , т.е.  $g(x) \leq g_1(x)$  при  $x \in \mathbf{X}$ .

Идея состоит в том, чтобы каждый раз отбирать реализацию  $x_1$  случайной величины  $\xi_1$  в качестве реализации случайной величины  $\xi$  с вероятностью  $g(x_1)/g_1(x_1)$ .

Это дает следующий алгоритм.

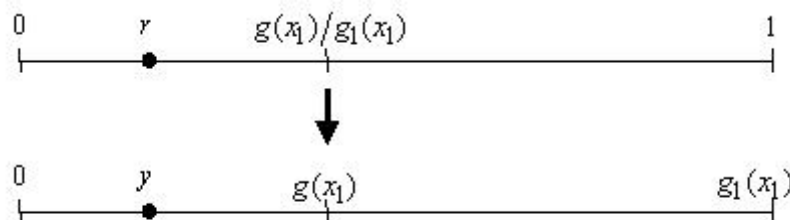
Шаг 1. Получить реализацию  $x_1$  случайной величины с плотностью  $f_1(x)$ , а также реализацию  $r$  случайной величины, равномерно распределенной в интервале  $(0, 1)$ , и вычислить  $y = r g_1(x_1)$ .

Шаг 2. Если

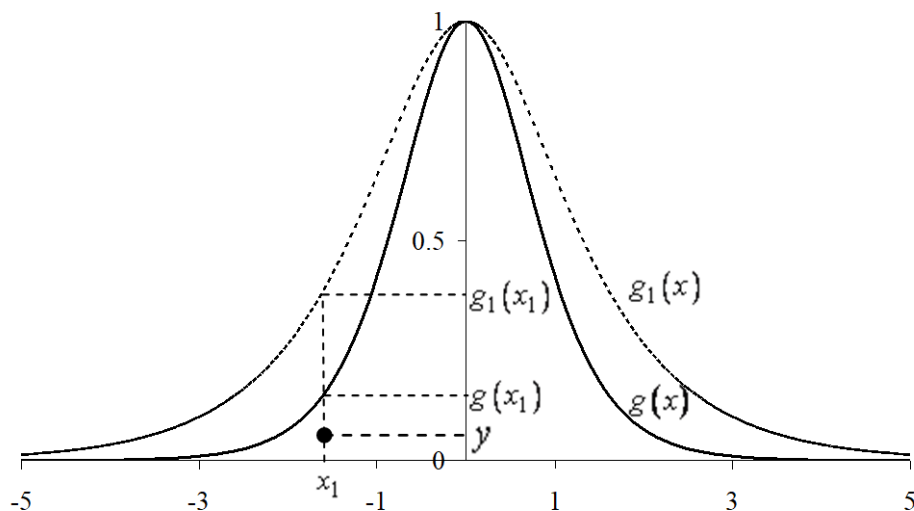
$$y < g(x_1), \quad (2.6)$$

то в качестве реализации случайной величины  $\xi$  принимаем  $x_1$ . В случае, когда неравенство (2.6) не выполнено, перейти на шаг 1.

Очевидно, неравенство (2.6) эквивалентно неравенству  $r < g(x_1)/g_1(x_1)$ , используемому при выборе из двух вариантов – отобрать или исключить – стандартным алгоритмом моделирования дискретной случайной величины.



С геометрической точки зрения изначально моделируются точки  $(x_1, y)$ , равномерно распределенные под графиком функции  $g_1(x)$ , а отбираются только те из них, которые лежат под графиком функции  $g(x)$ .



Обоснованием алгоритма служит тот факт, что условное распределение  $\xi_1$  при условии (2.6) совпадает с распределением  $\xi$ .

Метод исключения реализует функциональное преобразование случайного и потенциально неограниченного количества реализаций базовой случайной величины, поэтому важно обращать внимание на трудоемкость метода – она, как минимум, должна быть конечной.

Трудоемкостью метода исключения назовем среднее количество реализаций случайной величины  $\xi_1$ , затрачиваемых на получение одной реализации случайной величины  $\xi$ . Трудоемкость вычисляется как величина, обратная вероятности попадания точки  $(x_1, y)$  под график функции  $g(x)$ , и равна  $K_1 / K$  (значение трудоемкости

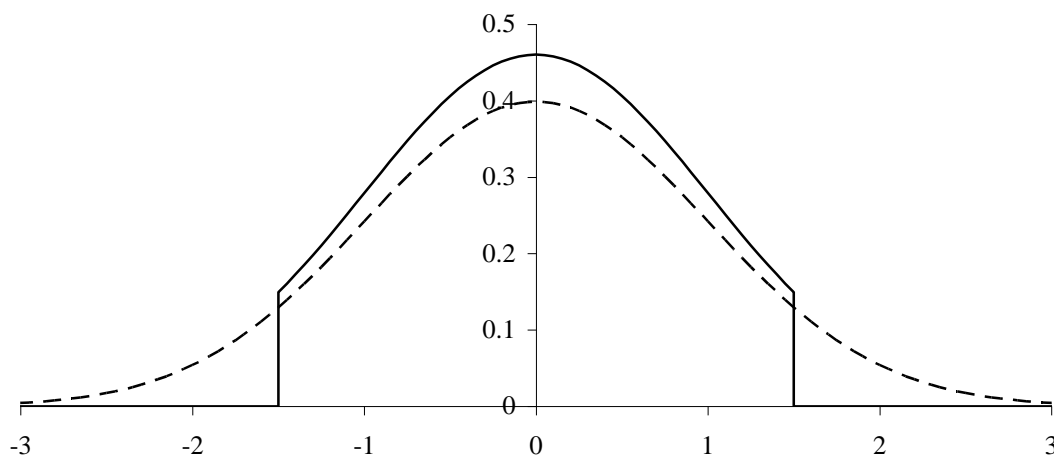
не меньше единицы). Таким образом, мажоранту  $g_1(x)$  функции  $g(x)$  следует подбирать так, чтобы площади  $K_1$  и  $K$  были близки; это выполнено при  $g_1(x) \approx g(x)$ .

**Пример 5. Усеченное распределение.**

Рассмотрим случайную величину  $\xi_1$ , распределенную на интервале  $\mathbb{X} \subseteq \mathbf{R}^1$  согласно плотности  $f_1(x)$ .

Случайная величина  $\xi$  имеет усеченное распределение величины  $\xi_1$ , если она распределена на подинтервале  $\mathbf{X} \subset \mathbb{X}$  и ее плотность распределения  $f(x)$  пропорциональна на  $\mathbf{X}$  плотности  $f_1(x)$ :

$$f(x) = \frac{1}{K} f_1(x) = \frac{f_1(x)}{\int_{\mathbf{X}} f_1(y) dy}, \quad x \in \mathbf{X} \subset \mathbb{X}.$$



В случае, когда имеется эффективный алгоритм моделирования случайной величины  $\xi_1$ , можно использовать следующий алгоритм для моделирования случайной величины  $\xi$ .

1. Получим реализацию  $x_1$  случайной величины  $\xi_1$  на интервале  $\mathbb{X}$  согласно плотности  $f_1(x)$ .

2. Если  $x_1 \in \mathbf{X}$ , то в качестве реализации случайной величины  $\xi$  принять  $x_1$ , иначе перейти на шаг 1.

Несложно понять, что данный алгоритм является частным случаем алгоритма метода исключения на интервале  $\mathbb{X}$ , на котором для функции  $f(x)$  использована мажоранта  $\frac{1}{K} f_1(x)$  при  $x \in \mathbb{X}$  (при  $x \in \mathbf{X}$  имеем  $f(x) = \frac{1}{K} f_1(x)$ , а при  $x \notin \mathbf{X}$  выполнено  $f(x) = 0 < \frac{1}{K} f_1(x)$ ). Решение – отобрать или исключить – принимается детерминировано, так как при  $x_1 \in \mathbf{X}$  неравенство (2.6) заведомо выполнено, а при  $x_1 \notin \mathbf{X}$  – заведомо не выполнено.

Трудоемкость моделирования усеченного распределения методом исключений равна  $1/K$  и, следовательно, растет с уменьшением длины интервала  $\mathbf{X}$ .

*Примечание.* Простое введение в методы Монте-Карло можно найти в брошюре [Соболев И.М. Метод Монте-Карло. – М.: Наука., 1985], более продвинутый материал – в учебниках [Соболев И.М. Численные методы Монте-Карло. – М.: Наука,

1973], [Михайлов Г.А., Войтишек А.В. Численное статистическое моделирование. Методы Монте-Карло. – М.: Издательский центр «Академия», 2006], практико-ориентированное изложение можно найти в книге [Иванова В.М. Случайные числа и их применение. — М.: Финансы и статистика, 1984].

### 3.3. Моделирование засоренных распределений

Засоренное распределение представляет собой двухкомпонентную смесь с плотностью

$$g(x) = (1 - \varepsilon)f(x) + \varepsilon h(x),$$

где  $0 < \varepsilon < 0.5$  – уровень засорения,  $f(x)$ ,  $h(x)$  – плотности чистого и засоряющего распределений.

Пусть плотности  $f(x)$  и  $h(x)$  симметричны. Обозначим  $M_i$ ,  $D_i$ ,  $\gamma_{2i}$ ,  $i = 1, 2$ , математическое ожидание, дисперсию, коэффициент эксцесса, соответствующие при  $i = 1$  и  $i = 2$  плотностям  $f(x)$  и  $h(x)$ . Тогда математическое ожидание, дисперсия, коэффициенты асимметрии и эксцесса случайной величины с засоренным распределением выражаются формулами

$$M = (1 - \varepsilon)M_1 + \varepsilon M_2,$$

$$D = (1 - \varepsilon)(M_1^2 + D_1) + \varepsilon(M_2^2 + D_2) - M^2,$$

$$\gamma_1 = \frac{1}{D^{1.5}} \left\{ (1 - \varepsilon) \left[ (M_1 - M)^3 + 3(M_1 - M)D_1 \right] + \varepsilon \left[ (M_2 - M)^3 + 3(M_2 - M)D_2 \right] \right\},$$

$$\begin{aligned} \gamma_2 = \frac{1}{D^2} \left\{ (1 - \varepsilon) \left[ (M_1 - M)^4 + 6(M_1 - M)^2 D_1 + D_1^2 (\gamma_{21} + 3) \right] + \right. \\ \left. + \varepsilon \left[ (M_2 - M)^4 + 6(M_2 - M)^2 D_2 + D_2^2 (\gamma_{22} + 3) \right] \right\} - 3 \end{aligned}$$

в предположении, что все величины в правых частях равенств существуют.

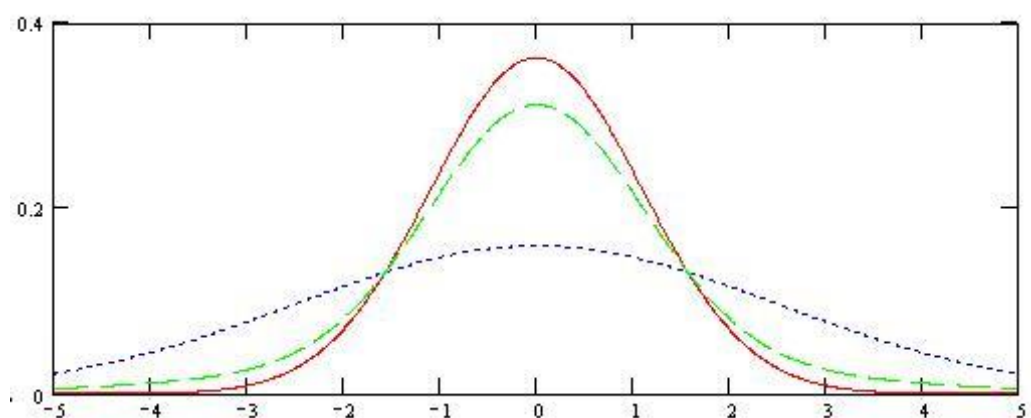
Для моделирования случайной величины с засоренным распределением можно использовать следующий алгоритм, реализующий метод дискретной суперпозиции.

Шаг 1. Получить реализацию  $r$  случайной величины, имеющей равномерное распределение на интервале  $(0, 1)$ . Если  $r \leq 1 - \varepsilon$ , перейти на шаг 2, иначе перейти на шаг 3.

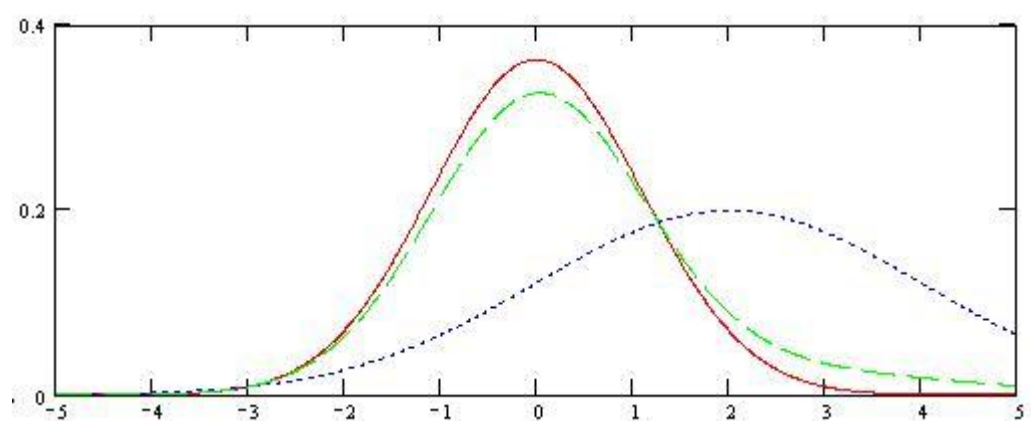
Шаг 2. Получить реализацию случайной величины с плотностью  $f(x)$ . Она и будет реализацией целевой случайной величины.

Шаг 3. Получить реализацию случайной величины с плотностью  $h(x)$ . Она и будет реализацией целевой случайной величины.

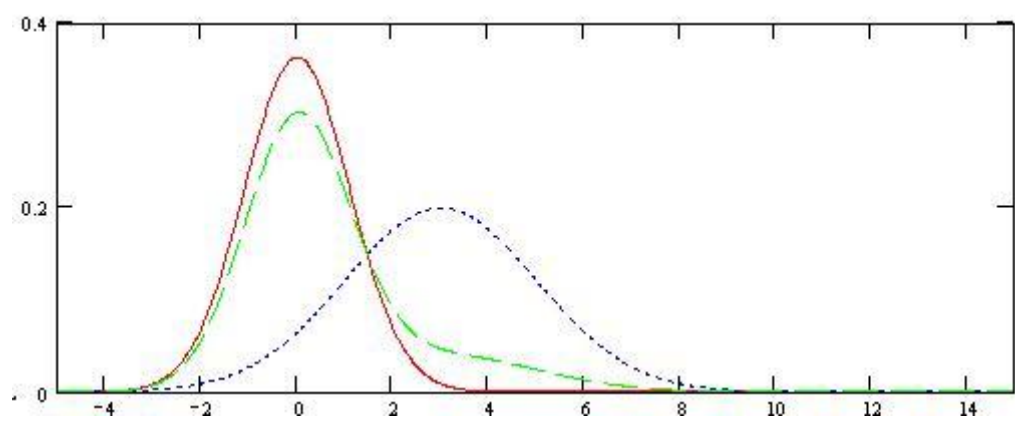




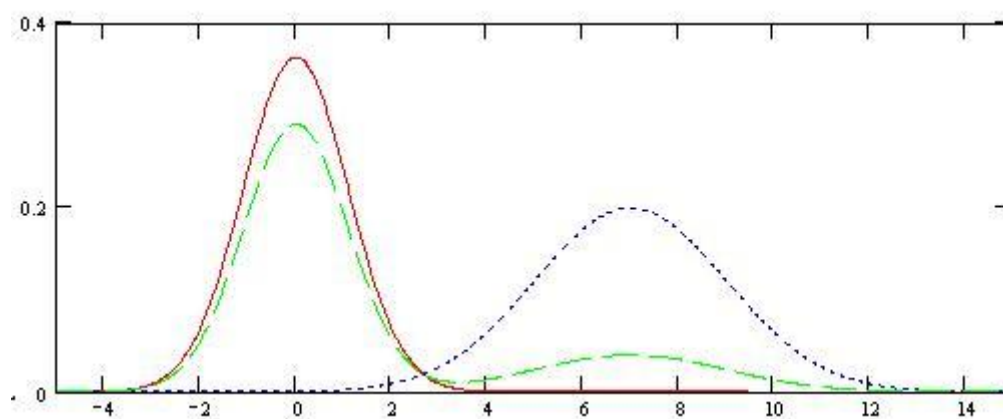
Случай симметричного засорения (красная – чистое, синяя – засоряющее, зеленая – засоренное распределения)



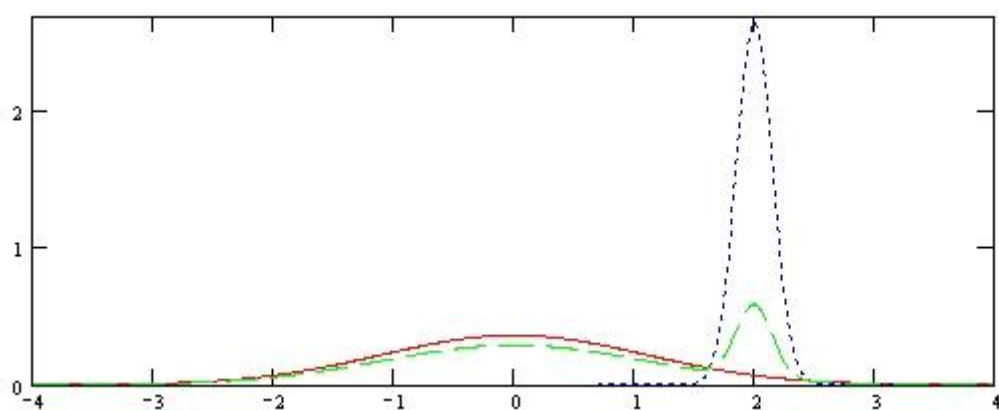
Случай асимметричного засорения



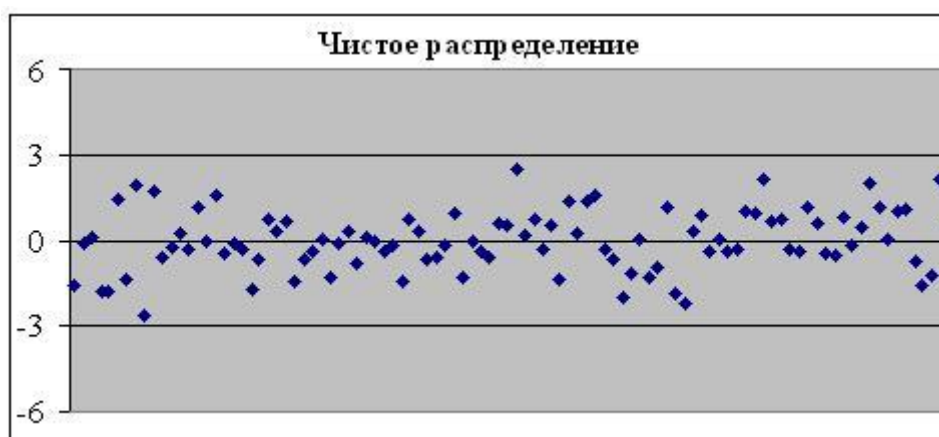
Случай асимметричного засорения



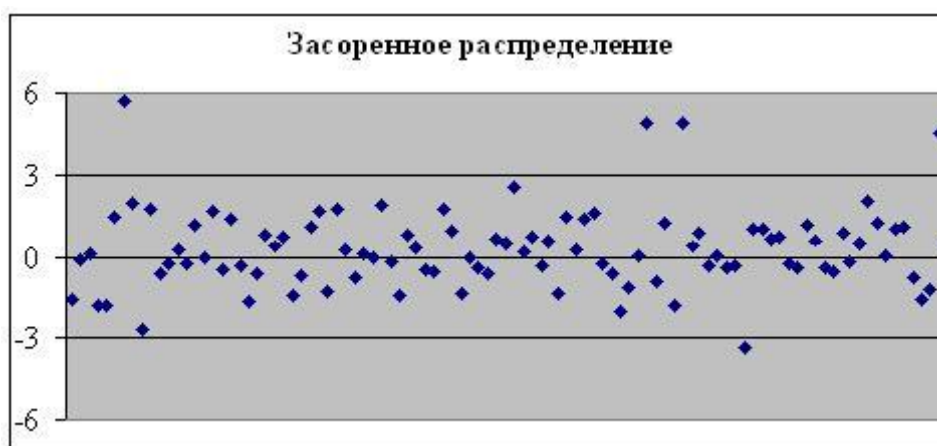
Случай асимметричного засорения (не рекомендуется)



Случай асимметричного засорения (не рекомендуется)



Выборка из чистого распределения



Выборка из засоренного распределения

### 3.4. Использование метода Монте-Карло для исследования методов оценивания

Общая схема метода Монте-Карло, применительно к исследованию методов оценивания параметров состоит в следующем.

1. Задается число испытаний  $M$  в пределах от нескольких сотен до нескольких десятков тысяч. Необходимо выбрать значение  $M$  настолько большим, насколько позволяют возможности вычислительной техники, но не меньше 100 (желательно – не меньше 1000).

2. Моделируется  $M$  выборок с одинаковыми характеристиками (объемом выборки, значениями параметров и т.д.). Для каждой выборки оценивается параметр (требуемыми способами).

3. По  $M$  значениям каждой оценки вычисляются значения критериев качества.

В нашей работе будем использовать следующие критерии:

- дисперсия оценки  $\mathbf{D}\hat{\theta} = \mathbf{E}(\hat{\theta} - \mathbf{E}\hat{\theta})^2$ ;
- квадрат смещения оценки  $b^2(\hat{\theta}) = (\mathbf{E}\hat{\theta} - \theta_{\text{ист}})^2$ ;
- квадратичная ошибка оценки  $\tau = \mathbf{D}\hat{\theta} + b^2(\hat{\theta})$ .

Выборочные значения критериев качества, вычисляемые методом Монте-Карло, имеют вид:

- дисперсия оценки  $\bar{\mathbf{D}}\hat{\theta} = \frac{1}{M} \sum_{m=1}^M (\hat{\theta}_m - \bar{\hat{\theta}})^2$ ;
- квадрат смещения оценки  $\bar{b}^2(\hat{\theta}) = (\bar{\hat{\theta}} - \theta_{\text{ист}})^2$ ;
- квадратичная ошибка оценки  $\bar{\tau} = \bar{\mathbf{D}}\hat{\theta} + \bar{b}^2(\hat{\theta})$ .

Здесь  $\bar{\hat{\theta}} = \frac{1}{M} \sum_{m=1}^M \hat{\theta}_m$  – оценка  $\mathbf{E}\hat{\theta}$ ,  $\hat{\theta}_m$  – значение оценки в  $m$ -м испытании.

Особенность метода Монте-Карло состоит в том, что погрешность вычислений, как правило, пропорциональна  $C/\sqrt{M}$ , где  $C$  – некоторая постоянная. Отсюда видно, что для того, чтобы уменьшить погрешность в 10 раз (иначе говоря, чтобы получить в ответе еще один верный десятичный знак), нужно увеличить  $M$  (т.е. объем работы) в 100 раз. Добиться высокой точности таким путем невозможно, например, при  $M = 100$  погрешность будет иметь порядок 0.1.

### 3.5. Случайные величины и способы их моделирования

Ниже приведены сведения о случайных величинах с распределениями, используемыми в вариантах. При этом рассматриваются **стандартные распределения**, т.е. распределения с нулевым параметром сдвига и единичным параметром масштаба.

Плотность случайной величины с произвольными параметрами сдвига и масштаба задается **формулой (1.1)**. При моделировании перейти от случайной величины со стандартным распределением к случайной величине с произвольными параметрами сдвига и масштаба можно согласно **примеру 3**.

Математическое ожидание распределения с плотностью (1.1), если оно существует, равно математическому ожиданию стандартного распределения, умноженному на  $\lambda$  и увеличенному на  $\theta$ . Дисперсия распределения с плотностью (1.1), если она существует, равна дисперсии стандартного распределения, увеличенной в  $\lambda^2$  раз. Коэффициенты асимметрии и эксцесса при сдвиг-масштабном преобразовании не изменяются.

Для всех распределений представлены генераторы, которые рассчитаны на низкоуровневую реализацию в том смысле, что они требуют только значений случайной величины, равномерно распределенной на интервале (0, 1).

**Замечания.** 1. Низкоуровневые средства нередко содержат генераторы неотрицательных целочисленных (а не непрерывных) равномерных псевдослучайных чисел. В последнем случае можно получать равномерные на  $[0, 1]$  псевдослучайные числа путем деления выдаваемого генератором целого числа на максимальное значение, которое генератор может выдавать. Однако некоторые алгоритмы могут быть некорректными при значениях 0 и 1, поэтому их рекомендуется отбрасывать.

2. В помощь программирующим на C/C++ имеется библиотечка для вычисления специальных функций `spec_func`.

#### 1. Нормальное распределение.

Нормальное распределение имеет плотность

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

единичную дисперсию и коэффициент эксцесса 0.

Моделирование случайной величины можно осуществить методом функциональных преобразований по одной из двух формул (метод Бокса–Мюллера)

$$x = \sqrt{-2 \ln r_1} \cos 2\pi r_2, \quad x = \sqrt{-2 \ln r_1} \sin 2\pi r_2,$$

где  $r_1, r_2$  – реализации случайной величины, равномерно распределенной на интервале  $(0, 1)$ .

## 2. Обобщенное гауссовское распределение.

Обобщенное гауссовское распределение (распределение Субботина, двустороннее экспоненциальное распределение, экспоненциально-степенное распределение, обобщённое распределение ошибок) имеет плотность

$$f(x, \nu) = \frac{\nu}{2\Gamma(1/\nu)} \exp[-|x|^\nu],$$

где  $\nu > 0$ ,  $\Gamma$  – гамма-функция.

При  $\nu = 1$  распределение совпадает с лапласовским, при  $\nu = 2$  – с нормальным, при  $\nu \rightarrow \infty$  стремится к равномерному. Дисперсия определяется формулой  $\sigma^2 = \Gamma(3/\nu)/\Gamma(1/\nu)$ , где  $\Gamma$  – гамма-функция. Коэффициент эксцесса определяется

$$\text{формулой } \gamma_2 = \frac{\Gamma(5/\nu)\Gamma(1/\nu)}{[\Gamma(3/\nu)]^2} - 3.$$

В таблице приведены значения дисперсии и коэффициента эксцесса для некоторых значений параметра формы.

$\nu$	1.0	1.1	1.2	1.3	1.4	1.5	1.6	1.7	1.8	1.9
$\sigma^2$	2	1.49	1.18	0.98	0.84	0.74	0.67	0.61	0.56	0.53
$\gamma_2$	3	2.28	1.74	1.34	1.02	0.76	0.55	0.38	0.23	0.11
$\nu$	2	3	4	5	6	7	8	9	10	$\infty$
$\sigma^2$	0.5	0.37	0.34	0.32	0.32	0.32	0.32	0.31	0.31	1/3
$\gamma_2$	0	-0.58	-0.81	-0.93	-1	-1.04	-1.08	-1.1	-1.12	-1.2

Для моделирования случайной величины, имеющей распределение с параметром формы  $1 \leq \nu < 2$ , можно использовать следующий алгоритм, реализующий метод исключения.

Шаг 0. Вычислить константы  $a = \frac{1}{\nu} - 1$ ,  $b = \frac{1}{\nu^{1/\nu}}$  (один раз для каждого значения  $\nu$ ).

Шаг 1. Получить методом обратной функции реализацию  $x$  лапласовской с масштабом  $b$  случайной величины. Для этого получить реализацию  $r_1$  случайной величины, равномерно распределенной на интервале  $(0, 1)$ , и вычислить

$$x = \begin{cases} b \ln(2r_1), & r_1 \leq 1/2 \\ -b \ln(2(1-r_1)), & r_1 > 1/2 \end{cases}.$$

Шаг 2. Получить реализацию  $r_2$  случайной величины, равномерно распределенной на интервале  $(0, 1)$ .

Шаг 3. Если  $\ln(r_2) \leq -|x|^\nu + |x|/b + a$ , то  $x$  – реализация целевой случайной величины, иначе перейти на шаг 1.

Для моделирования случайной величины, имеющей распределение с параметром формы  $v \geq 2$ , можно использовать следующий алгоритм, реализующий метод исключения.

Шаг 0. Вычислить константы  $a = \frac{1}{v} - \frac{1}{2}$ ,  $b = \frac{1}{v^{1/v}}$ ,  $c = 2b^2$  (один раз для каждого значения  $v$ ).

Шаг 1. Получить реализацию  $x$  нормальной с масштабом  $b$  случайной величины. Для этого получить реализации  $r_1$ ,  $r_2$  случайной величины, равномерно распределенной на интервале  $(0, 1)$ , и вычислить  $x$  по одной из двух формул

$$x = b\sqrt{-2 \ln r_1} \cos 2\pi r_2, \quad x = b\sqrt{-2 \ln r_1} \sin 2\pi r_2.$$

Шаг 2. Получить реализацию  $r_3$  случайной величины, равномерно распределенной на интервале  $(0, 1)$ .

Шаг 3. Если  $\ln(r_3) \leq -|x|^v + x^2/c + a$ , то  $x$  – реализация целевой случайной величины, иначе перейти на шаг 1.

### 3. Распределение Стьюдента.

Распределение Стьюдента имеет плотность

$$f(x, v) = \frac{\Gamma((v+1)/2)}{\sqrt{v\pi}\Gamma(v/2)} \left[ 1 + \frac{x^2}{v} \right]^{-\frac{v+1}{2}}$$

где  $v > 0$ ,  $\Gamma$  – гамма-функция. Дисперсия существует при  $v > 2$  и равна  $\frac{v}{v-2}$ , ко-

эффициент эксцесса существует при  $v > 4$  и равен  $\frac{6}{v-4}$ .

В таблице приведены значения дисперсии, коэффициента эксцесса для некоторых значений параметра формы.

$v$	5	7	9	15	25
$\sigma^2$	1.67	1.4	1.29	1.15	1.09
$\gamma_2$	6	2	1.2	0.55	0.29

Моделирование случайной величины можно осуществить методом функциональных преобразований по одной из двух формул (обобщенный метод Бокса–Мюллера)

$$x = \sqrt{v(r_1^{-2/v} - 1)} \cos 2\pi r_2, \quad x = \sqrt{v(r_1^{-2/v} - 1)} \sin 2\pi r_2,$$

где  $r_1, r_2$  – реализации случайной величины, равномерно распределенной на интервале  $(0, 1)$ .

### 4. Косинусно-экспоненциальное распределение.

Косинусно-экспоненциальное распределение имеет плотность

$$f(x, v) = \begin{cases} b_2 \cos^2(b_1 x), & |x| \leq 1 \\ b_3 \exp(-b_4 |x|), & |x| > 1 \end{cases},$$

где  $0 \leq v < 1$ , константы  $b_1, \dots, b_4$  – неотрицательные числа, которые зависят от  $v$  и определяются из уравнений

$$1 + b_1 \operatorname{tg} b_1 - \frac{\cos^2 b_1}{v} = 0, \quad b_4 = 2b_1 \operatorname{tg} b_1, \quad b_3 = \frac{vb_4}{2} \exp b_4, \quad b_2 = \frac{b_4}{2 + b_4}.$$

Плотность распределения в центре соответствует усеченному косинусному распределению (вероятность попадания в центр равна  $1 - v$ ), а ее хвосты являются лапласовскими. При  $v = 0$  плотность не имеет экспоненциальной части (значение плотности вне интервала  $[-1, 1]$  является нулевым), распределение совпадает с косинусным. При приближении параметра формы к 1 распределение будет приближаться к лапласовскому, но дисперсия при этом будет неограниченно возрастать. Распределение является наименее благоприятным в классе приближенных финитных моделей (см. стр. 37 учебного пособия).

Дисперсия определяется формулой

$$\sigma^2 = b_2 \left[ \frac{1}{3} + \left( \frac{1}{2b_1} - \frac{1}{4b_1^3} \right) \sin(2b_1) + \frac{1}{2b_1^2} \cos(2b_1) \right] + 2b_3 e^{-b_4} \left( \frac{2}{b_4^3} + \frac{2}{b_4^2} + \frac{1}{b_4} \right),$$

коэффициент эксцесса определяется формулой

$$\gamma_2 = \frac{b_2}{\sigma^4} \left[ \frac{1}{5} + 3 \left( \frac{1}{6b_1} - \frac{1}{2b_1^3} + \frac{1}{4b_1^5} \right) \sin(2b_1) + 3 \left( \frac{1}{3b_1^2} - \frac{1}{2b_1^4} \right) \cos(2b_1) \right] + \frac{2b_3 e^{-b_4}}{\sigma^4} \left( \frac{24}{b_4^5} + \frac{24}{b_4^4} + \frac{12}{b_4^3} + \frac{4}{b_4^2} + \frac{1}{b_4} \right) - 3.$$

В таблице приведены значения параметра  $b_1$ , дисперсии и коэффициента эксцесса для некоторых значений параметра формы  $v$ .

$v$	0	0.05	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
$b_1$	$\pi/2$	1.14	1.03	0.88	0.77	0.67	0.59	0.5	0.42	0.33	0.23
$\sigma^2$	0.13	0.28	0.39	0.69	1.18	2.03	3.59	6.77	14.27	37.6	174.3
$\gamma_2$	-0.59	-0.12	0.36	1.18	1.8	2.24	2.54	2.74	2.87	2.95	2.99

Для моделирования случайной величины можно использовать следующий алгоритм, включающий в себя применение метода суперпозиции для составной плотности, метода исключения для моделирования центра распределения, метода обратной функции для моделирования правого хвоста (экспоненциальное распределение со сдвигом 1 и масштабом  $1/b_4$ ) и метода функциональных преобразований (зеркальное отражение) для моделирования левого хвоста. Все величины  $r_i$  являются реализациями случайной величины, равномерно распределенной на интервале  $(0, 1)$ .

Шаг 1. Получить реализацию  $r_1$ . Если  $r_1 \geq v$ , перейти на шаг 2, иначе перейти на шаг 4.

Шаг 2. Получить реализации  $r_2, r_3$  и вычислить  $x_1 = 2r_2 - 1$ .

Шаг 3. Если  $r_3 \leq \cos^2(b_1 x_1)$ , то  $x_1$  – реализация целевой случайной величины, иначе перейти на шаг 2.

Шаг 4. Получить реализацию  $r_4$ , вычислить  $x_2 = 1 - \frac{\ln r_4}{b_4}$ .

Шаг 5. Если  $r_1 < v/2$ , то  $x_2$  – реализация целевой случайной величины, иначе реализацией целевой случайной величины является  $-x_2$ .

5. *SN-распределение*. SN-распределение (sh-нормальное) имеет плотность

$$f(x, v) = \phi\left(\frac{2}{v} \operatorname{sh} x\right) \frac{2}{v} \operatorname{ch} x,$$

где  $0 < v < 2$ ,  $\phi(x)$  – плотность стандартного нормального распределения,  $\operatorname{sh} x$  – синус гиперболический,  $\operatorname{ch} x$  – косинус гиперболический.

Числовые характеристики распределения не имеют аналитического представления. В таблице приведены значения дисперсии и коэффициента эксцесса для некоторых значений параметра формы  $v$ .

$v$	0.1	0.3	0.5	0.8	1
$\sigma^2$	0.00249	0.022	0.0591	0.141	0.209
$\gamma_2$	-0.01	-0.08	-0.19	-0.37	-0.48

Для моделирования случайной величины можно использовать следующую формулу, основанную на методе функциональных преобразований:

$$x = \operatorname{Arsh}(vz/2),$$

где  $\operatorname{Arsh}(y) = \ln\left(y + \sqrt{y^2 + 1}\right)$  – функция, обратная к  $\operatorname{sh}$ ,  $z$  – реализация нормальной случайной величины (способ ее моделирования см. в пп. 1).

6.  $S_U^S$ -распределение Джонсона (симметричное  $S_U$ -распределение Джонсона).

$S_U^S$ -распределение Джонсона имеет плотность

$$f(x, v) = \frac{v}{\sqrt{2\pi}} \frac{1}{\sqrt{x^2 + 1}} \exp\left\{-\frac{v^2}{2} \left[\ln\left(x + \sqrt{x^2 + 1}\right)\right]^2\right\},$$

где  $v > 0$ . Распределение при  $v \rightarrow \infty$  стремится к нормальному, однако, при одновременном стремлении дисперсии к нулю.

Дисперсия определяется формулой  $\sigma^2 = (w-1)/2$ , коэффициент эксцесса  $\gamma_2 = \frac{w^2 + 2w - 3}{2}$ , где  $w = \exp(2/v^2)$ .

В таблице приведены значения дисперсии и коэффициента эксцесса для некоторых значений параметра формы  $v$ .

$v$	1.5	2	2.5	3	3.5	4	4.5	5	6
$\sigma^2$	0.716	0.324	0.189	0.124	0.0887	0.0666	0.0519	0.0416	0.0286



$\gamma_2$	3.696	1.508	0.825	0.529	0.370	0.275	0.213	0.170	0.116
------------	-------	-------	-------	-------	-------	-------	-------	-------	-------

Для моделирования случайной величины можно использовать следующую формулу, основанную на методе функциональных преобразований:

$$x = \text{sh}(z/v),$$

где  $\text{sh } x$  – синус гиперболический,  $z$  – реализация нормальной случайной величины (способ ее моделирования см. в пп. 1).

### 7. Распределение Хьюбера.

Распределение Хьюбера имеет плотность

$$f(x, v) = \frac{1-v}{\sqrt{2\pi}} \begin{cases} \exp\left(-\frac{x^2}{2}\right), & |x| \leq k \\ \exp\left(\frac{1}{2}k^2 - k|x|\right), & |x| > k \end{cases},$$

где  $0 < v < 1$  – параметр формы, вспомогательный параметр  $k > 0$  связан с  $v$  условием нормировки плотности

$$\frac{2}{k} \phi(k) + 2[\Phi(k) - 1] = \frac{v}{1-v},$$

$\Phi(x)$  и  $\phi(x)$  – функция и плотность стандартного нормального распределения.

Плотность распределения нормальная в центре, а ее хвосты являются лапласовскими. В предельном случае при  $v \rightarrow 0$  ( $k \rightarrow \infty$ ) получаем нормальное распределение, при приближении параметра  $v$  к 1 ( $k$  приближается к 0) распределение приближается к лапласовскому (однако, при неограниченном росте дисперсии).

Дисперсия определяется формулой  $\sigma^2 = 1 + \frac{2\phi(k)(k^2 + 2)}{2k^2\phi(k) + k^3[2\Phi(k) - 1]}$ , коэф-

фициент эксцесса определяется формулой

$$\gamma_2 = \frac{1}{\sigma^4 [2\phi(k)/k + 2\Phi(k) - 1]} \left[ 3(2\Phi(k) - 1) + 2\phi(k) \left( \frac{24}{k^5} + \frac{24}{k^3} + \frac{12}{k} + k \right) \right] - 3.$$

В таблице приведены значения параметра  $k$ , дисперсии, коэффициента эксцесса и вероятности попадания в центральный интервал  $P = \frac{2(1-v)}{k} f(k, v)$  для некоторых значений параметра  $v$ .

$v$	0.001	0.002	0.005	0.01	0.02	0.05	0.1	0.2	0.3
$k$	2.63	2.435	2.16	1.945	1.717	1.398	1.140	0.862	0.684
$\sigma^2$	1.01	1.02	1.05	1.09	1.18	1.41	1.84	2.89	4.40
$\gamma_2$	0.11	0.19	0.36	0.57	0.90	1.52	2.09	2.61	2.82
$P$	0.990	0.983	0.964	0.939	0.896	0.796	0.671	0.489	0.354

Для моделирования случайной величины можно использовать следующий алгоритм, включающий в себя применение метода суперпозиции для составной плотности, метода исключения для моделирования центра распределения, метода обратной функции для моделирования правого хвоста (экспоненциальное распределение

со сдвигом  $k$  и масштабом  $1/k$ ) и метода функциональных преобразований (зеркальное отражение) для моделирования левого хвоста.

Шаг 1. Получить реализацию  $r_1$  случайной величины, равномерно распределенной на интервале  $(0, 1)$ . Если  $r_1 \leq P$ , перейти на шаг 2, иначе перейти на шаг 4.

Шаг 2. Получить реализацию  $x_1$  нормальной случайной величины (способ ее моделирования см. в пп. 1).

Шаг 3. Если  $x_1$  принадлежит интервалу  $[-k, k]$ , то это реализация целевой случайной величины, иначе перейти на шаг 2.

Шаг 4. Получить реализацию  $r_2$  случайной величины, равномерно распределенной на интервале  $(0, 1)$ . Вычислить

$$x_2 = k - \frac{\ln r_2}{k}.$$

Шаг 5. Если  $r_1 < \frac{1+P}{2}$ , то  $x_2$  – реализация целевой случайной величины, иначе реализацией целевой случайной величины является  $-x_2$ .

#### 8. Парето-нормальное распределение.

Парето-нормальное распределение имеет плотность

$$f(x, v) = \frac{1}{K} \begin{cases} \phi(x), & |x| \leq v \\ \phi(v) \left( \frac{v}{|x|} \right)^{v^2}, & |x| > v \end{cases},$$

где  $K = 2\Phi(v) - 1 + \frac{2v}{v^2 - 1} \phi(v)$ ,  $\Phi(x)$  и  $\phi(x)$  – функция и плотность стандартного нормального распределения,  $v > 1$  – параметр формы.

Плотность распределения нормальная в центре ( $|x| \leq v$ ), ее правый хвост ( $x > v$ ) соответствует распределению Парето с параметром формы  $v^2 - 1$  и масштабом  $v$ , а левый хвост ( $x < -v$ ) симметричен правому.

Момент  $q$ -го порядка существует при  $q < v^2 - 1$ . В этих условиях дисперсия определяется формулой  $\sigma^2 = 1 + \frac{4v^3 \phi(v)}{K(v^2 - 1)(v^2 - 3)}$ , коэффициент эксцесса опреде-

ляется формулой  $\gamma_2 = \frac{1}{\sigma^4 K} \left[ 3(2\Phi(v) - 1) - 2v\phi(v) \left( 3 + v^2 - \frac{v^4}{v^2 - 5} \right) \right] - 3$ .

В таблице приведены значения дисперсии, коэффициента эксцесса и вероятность попадания в центральный интервал  $P = \frac{2\Phi(v) - 1}{K}$  для некоторых значений параметра формы  $v$ .

$v$	2.3	2.35	2.4	2.5	2.6	2.7	2.85	3	3.3
-----	-----	------	-----	-----	-----	-----	------	---	-----

$\sigma^2$	1.14	1.11	1.1	1.06	1.04	1.03	1.02	1.01	1.00
$\gamma_2$	8	3.34	2.57	1.3	0.77	0.48	0.26	0.15	0.05
P	0.970	0.974	0.978	0.983	0.988	0.991	0.995	0.997	0.999

Для моделирования случайной величины можно использовать следующий алгоритм, включающий в себя применение метода суперпозиции для составной плотности, метода исключения для моделирования центра распределения, метода обратной функции для моделирования правого хвоста и метода функциональных преобразований (зеркальное отражение) для моделирования левого хвоста.

Шаг 1. Получить реализацию  $r_1$  случайной величины, равномерно распределенной на интервале (0, 1). Если  $r_1 \leq P$ , перейти на шаг 2, иначе перейти на шаг 4.

Шаг 2. Получить реализацию  $x_1$  нормальной случайной величины (способ ее моделирования см. в пп. 1).

Шаг 3. Если  $x_1$  принадлежит интервалу  $[-v, v]$ , то это реализация целевой случайной величины, иначе перейти на шаг 2.

Шаг 4. Получить реализацию  $r_2$  случайной величины, равномерно распределенной на интервале (0, 1). Вычислить

$$x_2 = \frac{v}{r_2^{1/(v^2-1)}}.$$

Шаг 5. Если  $r_1 < \frac{1+P}{2}$ , то  $x_2$  – реализация целевой случайной величины, иначе реализацией целевой случайной величины является  $-x_2$ .

9. Составное экспоненциально-степенное распределение (с дополнительным параметром, равным 4) имеет плотность

$$f(x, v) = \frac{1}{K} \begin{cases} 1 - x^4, & |x| \leq v \\ (1 - v^4) \exp[-a(|x| - v)], & |x| > v \end{cases},$$

где  $0 < v < 1$  – параметр формы,  $K = \frac{(1 - v^4)^2 + 12v^4 + 4}{10v^3}$ ,  $a = \frac{4v^3}{1 - v^4}$ .

Распределение в центре – обобщенное Симпсона с параметром формы 4, а его хвосты являются лапласовскими. При приближении параметра  $v$  к 1 распределение приближается к обобщенному симпсоновскому. При приближении параметра формы к 0 распределение будет приближаться к лапласовскому, но дисперсия при этом будет неограниченно возрастать.

Дисперсия определяется формулой

$$\sigma^2 = \frac{2}{K} \left[ \frac{v^3}{3} - \frac{v^7}{7} + (1 - v^4) \left( \frac{2}{a^3} + \frac{2v}{a^2} + \frac{v^2}{a} \right) \right],$$

коэффициент эксцесса определяется формулой

$$\gamma_2 = \frac{2}{\sigma^4 K} \left[ \frac{v^5}{5} - \frac{v^9}{9} + (1-v^4) \left( \frac{24}{a^5} + \frac{24v}{a^4} + \frac{12v^2}{a^3} + \frac{4v^3}{a^2} + \frac{v^4}{a} \right) \right] - 3.$$

В таблице приведены значения дисперсии, коэффициента эксцесса и вероятности попадания в центральный интервал  $P = \frac{2(v-v^5/5)}{K}$  для некоторых значений параметра формы  $v$ .

$v$	0.5	0.6	0.7	0.75	0.8	0.85	0.9	0.95	1
$\sigma^2$	7.166	2.206	0.822	0.548	0.394	0.309	0.264	0.244	0.238
$\gamma_2$	2.887	2.530	1.579	0.856	0.104	-0.501	-0.862	-1.01	-1.04
$P$	0.219	0.400	0.613	0.717	0.812	0.891	0.951	0.988	1

Для моделирования случайной величины можно использовать следующий алгоритм, включающий в себя применение метода суперпозиции для составной плотности, метода исключения для моделирования центра распределения, метода обратной функции для моделирования правого хвоста и метода функциональных преобразований (зеркальное отражение) для моделирования левого хвоста. Обобщенное распределение Симпсона в центре моделируется методом функциональных преобразований на основе его представления в виде распределения произведения двух независимых случайных величин: равномерной на интервале  $(-1, 1)$  и степенной с плотностью  $f(t) = 5t^4$ ,  $0 \leq t \leq 1$  (для моделирования последней используется метод обратной функции). Все величины  $r_i$  являются реализациями случайной величины, равномерно распределенной на интервале  $(0, 1)$ .

Шаг 1. Получить реализацию  $r_1$ . Если  $r_1 \leq P$ , перейти на шаг 2, иначе перейти на шаг 4.

Шаг 2. Получить реализации  $r_2, r_3$  и вычислить  $x_1 = (2r_2 - 1)r_3^{1/5}$ .

Шаг 3. Если  $|x_1| \leq v$ , то  $x_1$  – реализация целевой случайной величины, иначе перейти на шаг 2.

Шаг 4. Получить реализацию  $r_4$ , вычислить  $x_2 = v - \frac{\ln r_4}{a}$ .

Шаг 5. Если  $r_1 < \frac{1+P}{2}$ , то  $x_2$  – реализация целевой случайной величины, иначе реализацией целевой случайной величины является  $-x_2$ .

10. *Обобщенное распределение Лапласа.* Обобщенное распределение Лапласа (распределение гамма-дисперсии, распределение функции Бесселя) при натуральном параметре формы  $n$  имеет плотность

$$f(x, n) = \frac{e^{-|x|}}{(n-1)! 2^n} \sum_{j=0}^{n-1} \frac{(n-1+j)!}{(n-1-j)! j!} \frac{|x|^{n-1-j}}{2^j}.$$

Такое распределение будет иметь сумма  $n$  независимых лапласовских случайных величин. При увеличении параметра  $n$  распределение приближается к нормальному

с неограниченным ростом дисперсии. Случайная величина имеет дисперсию  $\sigma^2 = 2n$ , коэффициент эксцесса  $\gamma_2 = 3/n$ .

В таблице приведены значения дисперсии, коэффициента эксцесса и выражения для плотности распределения при некоторых значениях параметра  $n$ .

$n$	$\sigma^2$	$\gamma_2$	$f(x, n)$
1	2	3	$e^{- x }/2$
2	4	1.5	$e^{- x }( x +1)/4$
3	6	1	$e^{- x }( x ^2+3 x +3)/16$
4	8	0.75	$e^{- x }( x ^3+6 x ^2+15 x +15)/96$
5	10	0.6	$e^{- x }( x ^4+10 x ^3+45 x ^2+105 x +105)/768$
6	12	0.5	$e^{- x }( x ^5+15 x ^4+105 x ^3+420 x ^2+945 x +945)/7680$

Моделирование случайной величины осуществляется по следующей формуле, основанной на методе функциональных преобразований и методе обратной функции для моделирования лапласовских случайных величин (подробности о последнем можно найти в пп. 2):

$$x = \ln \left( \prod_{i:r_i \leq 1/2} 2r_i / \prod_{i:r_i > 1/2} 2(1-r_i) \right),$$

где  $r_i$ ,  $i = 1, \dots, n$  – реализации случайной величины, равномерно распределенной на интервале  $(0, 1)$ ; если нет ни одного элемента, удовлетворяющего условию в произведении, то это произведение равно 1.

11. *Обобщенное логистическое распределение типа III.* Обобщенное логистическое распределение типа III имеет плотность

$$f(x, v) = \frac{1}{2^{v-1} B(v/2, v/2)} \frac{1}{\operatorname{ch}^v x},$$

где  $v > 0$ ,  $B$  – бета-функция,  $\operatorname{ch} x$  – косинус гиперболический. Частными случаями этого распределения являются распределение гиперболического косинуса (гиперболического секанса) при  $v = 1$  и логистическое распределение при  $v = 2$ . В общем случае данное распределение будет иметь натуральный логарифм случайной величины с бета-распределением второго рода и параметрами формы  $v/2$  (натуральный логарифм отношения двух независимых случайных величин, имеющих гамма-распределение с параметром формы  $v/2$ ). При больших значениях  $v$  распределение близко к нормальному.

Дисперсия определяется формулой  $\sigma^2 = \frac{1}{2} \psi^{(1)}\left(\frac{v}{2}\right)$ , где  $\psi^{(1)}(z)$  – тригамма-

функция. Коэффициент эксцесса определяется формулой

$$\gamma_2 = \frac{1}{2} \psi^{(3)}\left(\frac{v}{2}\right) / \left[ \psi^{(1)}\left(\frac{v}{2}\right) \right]^2, \quad \psi^{(3)}(z) \text{ – пентагамма-функция.}$$

В таблице приведены значения дисперсии и коэффициента эксцесса для некоторых значений параметра формы  $v$ .

$v$	1	1.5	2	3	4	5	7	10	15	25
$\sigma^2$	2.467	1.271	0.822	0.467	0.322	0.245	0.165	0.111	0.071	0.042
$\gamma_2$	2	1.53	1.2	0.81	0.59	0.47	0.32	0.22	0.14	0.08

Для моделирования случайной величины при  $1 \leq v < 2$  можно использовать следующий алгоритм, реализующий метод исключения.

Шаг 1. Получить реализацию  $x$  случайной величины, имеющей распределение гиперболического косинуса, методом обратной функции. Для этого получить реализацию  $r_1$  случайной величины, равномерно распределенной на интервале  $(0, 1)$ , и вычислить

$$x = \ln[\operatorname{tg}(\pi r_1 / 2)].$$

Шаг 2. Получить реализацию  $r_2$  случайной величины, равномерно распределенной на интервале  $(0, 1)$ .

Шаг 3. Если  $r_2 \leq \operatorname{ch}^{1-v} x$ , то  $x$  – реализация целевой случайной величины, иначе перейти на шаг 1.

Для моделирования случайной величины при  $v \geq 2$  можно использовать следующий алгоритм, реализующий метод исключения.

Шаг 1. Получить реализацию  $x$  логистической случайной величины методом обратной функции. Для этого получить реализацию  $r_1$  случайной величины, равномерно распределенной на интервале  $(0, 1)$ , и вычислить

$$x = \frac{\ln[r_1 / (1 - r_1)]}{2}.$$

Шаг 2. Получить реализацию  $r_2$  случайной величины, равномерно распределенной на интервале  $(0, 1)$ .

Шаг 3. Если  $r_2 \leq \operatorname{ch}^{2-v} x$ , то  $x$  – реализация целевой случайной величины, иначе перейти на шаг 1.

12. Составное обобщенное гауссовское распределение (с дополнительным параметром, равным 7) имеет плотность

$$f(x, v) = \frac{1}{K} \begin{cases} \exp[-|x|^7], & |x| \leq v \\ \exp[-a(|x| - v) - v^7], & |x| > v \end{cases},$$

где  $v > 0$ ,  $a = 7v^6$ ,  $K = \frac{2}{7} \left[ \gamma(1/7, v^7) + e^{-v^7} / v^6 \right]$ ,  $\gamma(c, z) = \int_0^z x^{c-1} e^{-x} dx$  – неполная гамма-функция.

Плотность распределения в центре соответствует обобщенному гауссовскому распределению с параметром формы 7 (см. пп. 2), а хвосты распределения являются лапласовскими. При неограниченном увеличении параметра  $v$  распределение приближается к обобщенному гауссовскому. При приближении параметра формы к 0 распределение будет приближаться к лапласовскому, но дисперсия при этом будет

неограниченно возрастать. Данное распределение является наименее благоприятным в классе засоренных обобщенных гауссовских распределений с параметром формы 7 и уровнем засорения  $1 - 2\Gamma(1/7)/7K$ , где  $\Gamma$  – гамма-функция.

Дисперсия определяется формулой

$$\sigma^2 = \frac{2}{K} \left[ \frac{\gamma(3/7, v^7)}{7} + e^{-v^7} \left( \frac{2}{a^3} + \frac{2v}{a^2} + \frac{v^2}{a} \right) \right],$$

коэффициент эксцесса определяется формулой

$$\gamma_2 = \frac{2}{\sigma^4 K} \left[ \frac{\gamma(5/7, v^7)}{7} + e^{-v^7} \left( \frac{24}{a^5} + \frac{24v}{a^4} + \frac{12v^2}{a^3} + \frac{4v^3}{a^2} + \frac{v^4}{a} \right) \right] - 3.$$

В таблице приведены значения дисперсии, коэффициента эксцесса, константы  $K$  и вероятности попадания в центральный интервал  $P = \frac{2}{7K} \gamma(1/7, v^7)$  для некоторых значений параметра формы.

$v$	0.55	0.65	0.7	0.75	0.8	0.85	0.9	0.95	1	1.1	$\infty$
$\sigma^2$	53.5	7.44	3.23	1.58	0.893	0.586	0.444	0.375	0.342	0.320	0.32
$\gamma_2$	2.98	2.79	2.48	1.93	1.16	0.35	-0.28	2.33	-0.67	-1.02	-1.04
$K$	11.3	4.90	3.62	2.88	2.44	2.19	2.04	1.95	1.91	1.88	1.87
$P$	0.097	0.264	0.383	0.512	0.638	0.749	0.836	0.901	0.945	0.988	1

Для моделирования случайной величины можно использовать следующий алгоритм, включающий в себя применение метода суперпозиции для составной плотности, метода исключения для моделирования центра распределения, метода обратной функции для моделирования правого хвоста и метода функциональных преобразований (зеркальное отражение) для моделирования левого хвоста. Все величины  $r_i$  являются реализациями случайной величины, равномерно распределенной на интервале  $(0, 1)$ .

Шаг 1. Получить реализацию  $r_1$ . Если  $r_1 \leq P$ , перейти на шаг 2, иначе перейти на шаг 4.

Шаг 2. Получить реализацию  $z$  нормальной случайной величины (способ ее моделирования см. в пп. 1), вычислить  $x_1 = 7^{-1/7} z$ .

Шаг 3. Получить реализацию  $r_2$ . Если  $|x_1| \leq v$  и

$$\ln(r_2) \leq -|x_1|^7 + 7^{2/7} x_1^2 / 2 - 5/14,$$

то  $x_1$  – реализация целевой случайной величины, иначе перейти на шаг 2.

Шаг 4. Получить реализацию  $r_3$ , вычислить  $x_2 = v - \frac{\ln r_3}{a}$ .

Шаг 5. Если  $r_1 < \frac{1+P}{2}$ , то  $x_2$  – реализация целевой случайной величины, иначе реализацией целевой случайной величины является  $-x_2$ .

#### 4. Варианты задания

1. Нормальное распределение.
2. Обобщенное гауссовское распределение со следующим значением параметра  $v$  :
  - а) 1.5;
  - б) 3;
  - в) 5.
3. Распределение Стьюдента со следующим значением параметра  $v$  :
  - а) 9;
  - б) 15;
  - в) 25.
4. Косинусно-экспоненциальное распределение со следующим значением параметра  $v$  :
  - а) 0.05;
  - б) 0.1;
  - в) 0.2.
5. SN-распределение со следующим значением параметра  $v$  :
  - а) 0.1;
  - б) 0.5;
  - в) 1.
6.  $S_U^S$ -распределение Джонсона со следующим значением параметра  $v$  :
  - а) 3;
  - б) 4;
  - в) 6.
7. Распределение Хьюбера со следующим значением параметра  $v$  :
  - а) 0.002;
  - б) 0.01;
  - в) 0.05.
8. Парето-нормальное распределение со следующим значением параметра  $v$  :
  - а) 2.5;
  - б) 2.7;
  - в) 2.85.
9. Составное экспоненциально-степенное распределение со следующим значением параметра  $v$  :
  - а) 0.75;
  - б) 0.8;
  - в) 0.85.
10. Обобщенное распределение Лапласа со следующим значением параметра  $n$  :
  - а) 2;
  - б) 3;
  - в) 4.
11. Обобщенное логистическое распределение типа III со следующим значением параметра  $v$  :



а) 1.5;

б) 3;

в) 7.

12. Составное обобщенное гауссовское распределение со следующим значением параметра  $\nu$  :

а) 0.8;

б) 0.85;

в) 0.9.