1.1. Распределения, использующиеся в настоящей работе

1.1.1. Распределение Парето

Распределение Парето – это двухпараметрическое семейство абсолютно непрерывных распределений, являющихся степенными. Называется по имени Вилфредо Парето. Встречается при исследовании различных явлений, в частности, социальных, экономических, физических и других. Вне области экономики иногда называется также распределением Брэдфорда.

Пусть случайная X величина такова, что её распределение задаётся равенством

$$F_X(x) = P(X < x) = 1 - \left(\frac{x_{\mathrm{m}}}{x}\right)^k, \ \forall x \geqslant x_{\mathrm{m}},$$

где x_m, $k > 0$. Тогда говорят, что X имеет распределение Парето с параметрами x_m и k. Плотность распределения Парето имеет вид

$$f_X(x) = \begin{cases} \frac{k x_m^k}{x^{k+1}}, & x \geqslant x_m, \\ 0, & x < x_m. \end{cases}$$

Моменты случайной величины, имеющей распределение Парето, задаются формулой

$$\mathbb{E}\left[X^n\right] = \frac{k x_m^n}{k-n},$$

откуда, в частности,

$$\mathbb{E}[X] = \frac{k x_m}{k-1},$$
$$\mathrm{D}[X] = \left(\frac{x_m}{k-1}\right)^2 \frac{k}{k-2}.$$

1.1.2. Распределение Пуассона

Распределение Пуассона является предельной формой для биномиального распределения, когда число испытаний $N \to \infty$, а вероятность успеха в одном испытании p_N зависит от N и $p_N \to 0$ так, что $\lambda = N p_N$ является конечным числом. Поскольку Np представляет собой среднее значение числа успехов для биномиального распределения, то его предельное значение также имеет смысл среднего значения, или математического ожидания для распределения Пуассона

$$P(\xi = M) = \frac{\lambda^M}{M!} e^{-\lambda}.$$

Если биномиальное распределение зависит от двух параметров N и p, то распределение Пуассона определяется одним параметром λ, так как N и p_N связаны приведенным выше соотношением.

Как уже упоминалось, математическое ожидание для распределения Пуассона равно λ. Дисперсия СВ, подчиненной распределению Пуассона, также равна λ. Асимметрия γ_1 и коэффициент эксцесса γ равны соответственно $\gamma_1 = 1/\sqrt{\lambda}$, $\gamma = 1/(\lambda + 3)$.

Сопоставляя, как и для биномиального распределения, отношение вероятностей соседних значений

$$\frac{P(\xi = M)}{P(\xi = M-1)} = \frac{\dfrac{\lambda^M}{M!} e^{-\lambda}}{\dfrac{\lambda^{M-1}}{(M-1)!} e^{-\lambda}} = \frac{\lambda}{M},$$

приходим к выводу, что при целочисленных значениях параметра λ наивероятнейшими будут значения $\lambda - 1$ и λ.

1.1.3. Логнормальное распределение

Случайная величина, имеющая в качестве ПВ своего логарифма выражение (2), называется СВ, распределенной по логнормальному закону.

$$W(x) = \begin{cases} \dfrac{1}{x\sigma\sqrt{2\pi}} \exp\left[-\dfrac{(\ln x - a)^2}{2\sigma^2}\right], & x > 0, \\ 0, \quad x \le 0. \end{cases}$$

Логнормальное распределение получается при преобразовании ПВ гауссовской случайной величины η с помощью функции $x = y^{\eta}$. Поскольку функция преобразования определяет величину x как строго положительную, то данное ограничение содержится и в выражении для ПВ.

Основные статистические характеристики логнормального распределения:

ФР	$\frac{1}{2}+\frac{1}{2}\mathrm{Erf}\left[\frac{\ln(x)-\mu}{\sigma\sqrt{2}}\right]$
Математическое ожидание (МО)	$e^{\mu+\sigma^2/2}$
Медиана	e^{μ}
Мода	$e^{\mu-\sigma^2}$
Дисперсия	$(e^{\sigma^2}-1)e^{\mu+\sigma^2}$

1.1.4. Экспоненциальное распределение

Экспоненциа́льное (или показа́тельное) распределе́ние – это абсолютно непрерывное распределение, моделирующее время между двумя последовательными свершениями одного и того же события.

Случайная величина X имеет экспоненциальное распределение с параметром $\lambda > 0$, если её плотность вероятности имеет вид:

$$f_X(x) = \begin{cases} \lambda\, e^{-\lambda x}, & x \ge 0, \\ 0, & x < 0. \end{cases}$$

Интегрируя плотность, получаем функцию экспоненциального распределения:

$$F_X(x) = \begin{cases} 1 - e^{-\lambda x} & , x \ge 0, \\ 0 & , x < 0. \end{cases}$$

Несложным интегрированием находим, что производящая функция моментов для экспоненциального распределения имеет вид:

$$\mathrm{M}_X(t) = \left(1 - \frac{t}{\lambda}\right)^{-1},$$

откуда получаем все моменты:

$$\mathbb{E}[X^n] = \frac{n!}{\lambda^n}.$$

$$\mathbb{E}[X] = \frac{1}{\lambda},$$

$$\mathbb{E}[X^2] = \frac{2}{\lambda^2},$$

$$\mathrm{D}[X] = \frac{1}{\lambda^2}.$$

1.2. Системы массового обслуживания (СМО)

1.2.1. Параметры СМО

Система массового обслуживания считается заданной, если определены следующие характеристики:

– Входящий поток требований или, иначе говоря, моменты поступления требований в систему;

– Структура системы обслуживания;

– Время обслуживания каждым каналом;

– Дисциплина ожидания, т. е. совокупность правил, регламентирующих хранение требований, находящихся в один и тот же момент времени в системе;

– Дисциплина обслуживания, т. е. совокупность правил, в соответствии с которыми требование выбирает прибор, которым оно будет обслужено;

– Эффективность системы массового обслуживания оценивается с помощью соответствующих показателей;

Основные показатели эффективности:

– Вероятность p_n нахождения в очереди или системе определенного числа требований n;

– Коэффициент загрузки системы u, равный отношению среднего времени обслуживания к среднему промежутку времени между последовательными моментами поступления требований;

– Среднее число требований в очереди L_q или в системе L;

– Среднее время ожидания в системе W_q или среднее время пребывания в системе W .

1.2.2. Распределение входного потока и разделение по дисциплине обслуживания

Оно зависит от характера физических процессов, протекающих в моделируемом объекте. Чаще всего для их моделирования используются следующие распределения: экспоненциальное, Эрланга k-го порядка, Рэлея, нормальное и равномерное. Входящие потоки, описываемые данными распределениями, являются нерегулярными и ординарными. Нерегулярным потоком называется такой поток, в котором заявки следуют через произвольные промежутки времени, а ординарным – такой поток, при котором за бесконечно малый промежуток времени в систему может поступить не более одного требования.

По дисциплине обслуживания системы массового обслуживания делятся на СМО без ожидания (с отказами в обслуживании) и СМО с ожиданием, подробная схема приведена на рис. 1.

По числу каналов обслуживания системы подразделяются на одноканальные и многоканальные. Многоканальные системы, в свою очередь, делятся на системы с одинаковыми параметрами каналов обслуживания (с равноценными каналами) и системы с различными параметрами каналов обслуживания (с неравноценными каналами). Длительность обслуживания заявки в канале является величиной с определенным характером закона распределения. Наиболее часто применяются экспоненциальный закон распределения и распределение Эрланга k-го порядка при различных k, хотя могут применяться и иные законы распределения. Процесс обслуживания требований в многоканальной СМО показан на рис. 2. На верхней оси в виде стрелок показаны моменты поступления требований на обслуживание. Оси, находящиеся ниже, обозначают каналы обслуживания, на которых закрашенными прямоугольниками показано время, занимаемое обслуживанием требования.

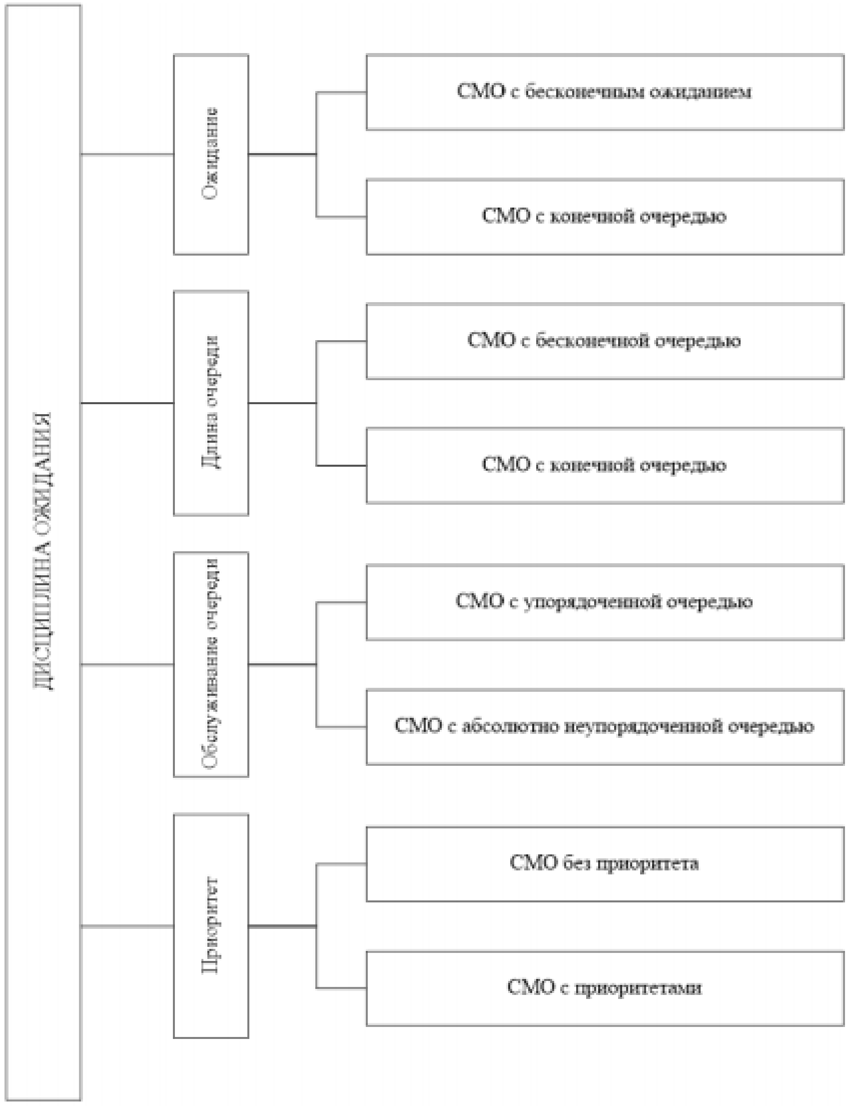

Рисунок 1 – Классификация СМО по дисциплинам ожидания

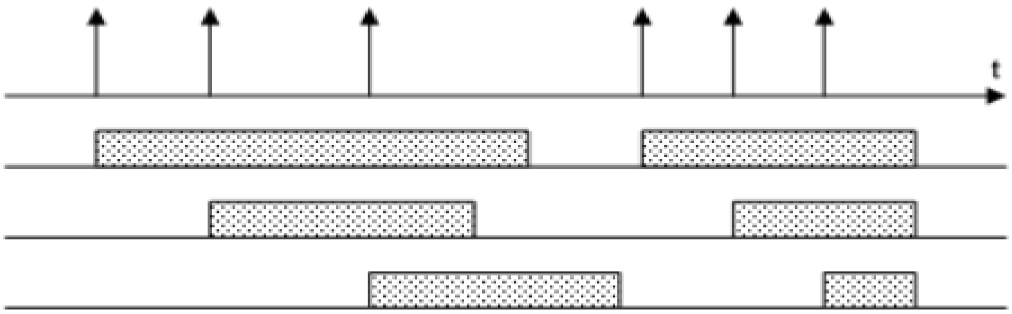

Рисунок 2 – Обслуживание требований в многоканальной СМО

В теории массового обслуживания приняты сокращенные обозначения СМО (называемые обозначениями Кендалла), в основе которых лежит обозначение вида $A/S/c$, где A и S описывают распределение временных интервалов между требованиями и распределение времени их обслуживания, а величина c – число обслуживающих устройств. В большинстве случаев A и S принимают следующие значения:

– M – показательное распределение;

– E_r – распределение Эрланга порядка r;

– D – детерминированное распределение;

– G – распределение произвольного вида.

Также в некоторых случаях указывается емкость буфера системы K, общее число требований N и дисциплина ожидания D. В таком случае будет использоваться шестибуквенное обозначение $A/S/c/K/N/D$. В случае отсутствия какого-либо из последних трех параметров предполагается, что он равен значению по умолчанию ($K = \infty$, $N = \infty$, D=$FIFO$). Например, система массового обслуживания с обозначением $M/M/1/10$ есть система с одним обслуживающим устройством, пуассоновским входным потоком (интервалы между требованиями распределены по экспоненциальному закону), экспоненциальным распределением времени обслуживания и емкостью буфера, равной 10 заявкам. В теории массового обслуживания СМО $G/G/1$ представляет собой систему с одним устройством обслуживания, в которой интервалы между требования и время обслуживания обладают произвольными распределениями. Интенсивность требований равна $\lambda = 1/\langle\tau\rangle$, где $\langle\tau\rangle$ – средний интервал между требованиями, т. е. Интенсивность потока λ – это среднее число событий в единицу времени. Интенсивность потока можно рассчитать экспериментально по формуле: $\lambda = N/Tн$, где N – число событий, произошедших за время наблюдения $Tн$. Интенсивность обслуживания равна $\mu = 1/T_{обсл}$, где $T_{обсл}$ – среднее время обслуживания заявки. Коэффициент использования можно рассчитать, как $U = \lambda/\mu$.

Схема работы СМО показана на рис. 3. При моделировании предполагается, что пользовательские сессии обслуживаются непрерывно и последовательно. Соответственно, если канал свободен в момент прибытия пользовательского запроса, начинается сессия, продолжающаяся до тех пор, пока не будет передана вся запрошенная пользователем информация. Таким образом время пребывания сессии в канале равно времени обслуживания $W_i = T_i^S = v_i / c$, где c – пропускная способность канала. В другом случае, если канал занят во время прибытия запроса на создание сессии, пользователь должен ожидать освобождения канала для начала передачи. Запрос будет поставлен в очередь типа "первый пришел – первый ушел", и время пребывания в таком случае будет равно, где T_i^W – время ожидания. Основным параметром, характеризующим производительность СМО, будет являться среднее время пребывания пользовательских сессий в канале.

Пользовательские сессии характеризуются временем начала t_i и передаваемым объемом информации v_i; T_i^W , T_i^S и W – времена ожидания, обслуживания и пребывания в системе соответственно.

Для определения диапазона c определяется суммарный размер информации, переданной через исследуемый канал, как сумма всех размеров пакетов, переданных за сессию $\sum_i v_i$, и определяется нижний предел пропускной способности канала связи $c_0 = \left(\sum_i v_i \right) / T_\Sigma$, где T_Σ – длительность интервала наблюдения. Если пропускная способность канала $c < c_0$, то возникает ситуация, когда невозможно передать всю информацию в течение времени T_Σ, что приводит к непрерывному увеличению длины очереди, нестационарному времени пребывания (на это указывает коэффициент использования $U = c_0 / c$ > 1). При $c = c_0$, соответствующей $U = 1$, система работает на границе стационарности, работая только при абсолютно регулярной

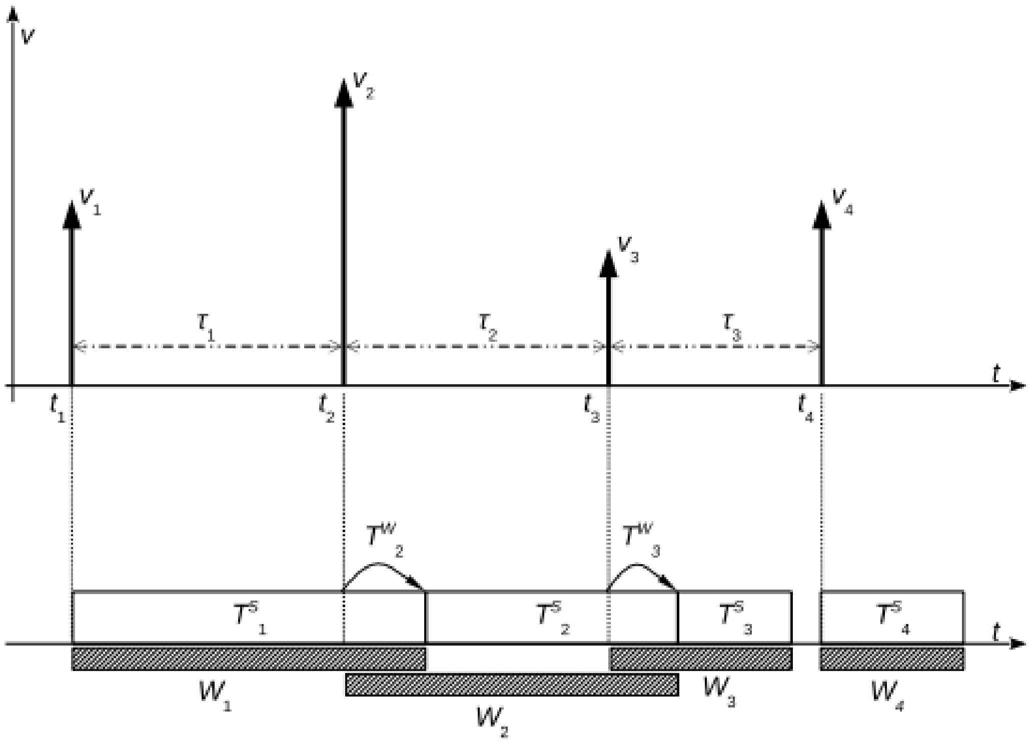

Рисунок 3 – Схема работы системы массового обслуживания

динамике входящего потока запросов, например, когда v являются постоянными. Поскольку динамика входящего потока запросов является нерегулярной, на практике $c > c_0$. При моделировании используются значения $c = 1{,}1 \cdot c_0 \dots 10 \cdot c_0$, соответствующие коэффициенту использования $U \cong 0{,}9 \dots 0{,}1$.

В силу нерегулярной активности пользователей ИКС ее трафик имеет долговременную зависимость, и таким образом, модель СМО с пуассоновским входящим потоком запросов значительно недооценивает требуемую пропускную способность канала связи. В качестве альтернативы пуассоновской модели предлагается модель СМО, в которой распределение времени между запросами и распределение времени их обслуживания определяется заданием на курсовую работу, а параметрами данных распределений выступают их оценки, полученные по эмпирическим данным.

Хорошо рассмотренная в литературе модель M/M/1 является моделью СМО, которая включает в себя один прибор обслуживания и бесконечный буфер с экспоненциально распределенными временем между пользовательскими запросами и временем их обслуживания. Среднее время ожидания, или время, проведенное в очереди, определяется как

$$E\left(W_q\right)=\frac{1}{\mu-\lambda},$$

где λ – интенсивность поступления запросов, μ – интенсивность их обслуживания. Тогда среднее время пребывания в системе равно

$$E(W)=\frac{1}{\mu-\lambda}+\frac{1}{\mu}.$$

Модель G/G/1 является одноканальной СМО с произвольными распределениями времени между запросами и времени обслуживания. Для СМО G/G/1, особенно когда она работает в режиме, при котором ее коэффициент использования близок к максимальному значению, приблизительное среднее время ожидания определяется по формуле Кингмана

$$\bar{W}=\frac{\bar{v}}{c}\left[\frac{U}{1-U}\right]\left[\frac{\rho_\tau^2+\rho_v^2}{2}\right]=\frac{\bar{v}}{c_0}\left[\frac{U^2}{1-U}\right]\left[\frac{\rho_\tau^2+\rho_v^2}{2}\right],$$

где $U=\lambda/\mu$ – коэффициент использования системы, ρ_τ – коэффициент вариации для времени между запросами τ и ρ_v – коэффициент вариации для времени обслуживания $T_i=v_i/c$. Коэффициент вариации определяется как отношение стандартного отклонения случайной величины к ее среднему значению. Интенсивность поступления запросов равная $1/\bar{\tau}$, в то время как интенсивность обслуживания $\mu=1/\bar{T}$, где $\bar{\tau}$ – среднее время между запросами, $\bar{T}$ – среднее время обслуживания.

Как было показано ранее, время пребывания заявки в системе $W_i=T_i^W+T_i^S$, где T_i^W – время ожидания, $T_i^S=v_i/c$ – время обслуживания.

Среднее число заявок в системе $n=l+u=\lambda\left(\overline{T^W}+\overline{T^S}\right)$ – формула Литтла, где $l=\lambda\overline{T^W}$ – средняя длина очереди.

1.3. Алгоритм Шрайбера-Шмитца

Алгоритм Шрайбера – Шмитца состоит из набора последовательно осуществляющихся этапов (см. рис. 4). На первом шаге выполняется синтез канонической псевдослучайной последовательности с равномерным распределением в интервале от 0 до 1. На втором этапе выполняется формирование ряда данных x(i) с требуемым видом распределения. Таковым в данном варианте настоящей работы являлось гамма-распределение. Затем осуществляется задание корреляционных свойств с использованием метода линейной фильтрации. На выходе линейной системы формируется поток данных y(i), при этом передаточная функция этого фильтра имеет вид

$$K(\omega) \sim \omega^{-\Delta\beta/2}.$$

Стоит отметить, что если требуемое распределение не является гауссовским (что характерно для данного варианта), то в силу эффекта нормализации при прохождении через линейную систему распределение искажается. С помощью операции поранговой замены отсчётов на отсчёты из требуемого распределения удаётся его восстановить. При этом наибольший отсчёт ряда y(i) заменяется наибольшим отсчётов ряда x(i); второй по рангу заменяется на второй по рангу и так далее до достижения наименьшего. Для формирования поранговой выборки использовались стандартные функции сортировки массива данных. После этого формируемый ряд z(i) имеет распределение, идентичное x(i), что в принципе соответствует поставленной задаче, однако характеризуется искаженными корреляционными свойствами. Для их сходства далее выполняется итерационное приближение. Оно заключается в том, что на каждом шаге (итерации) вычисляется передаточная функция корректирующего фильтра. Для этого выполняется оценка показателя Хёрста

$$H = \frac{2 - \gamma}{2},$$

взаимно однозначно связанного с корреляционным показателем. Поэтому для получения случайного процесса, характеризуемого спектром мощности $\omega^{-\beta'}$, где $\beta' = \beta + \Delta\beta$, β – значение, характеризующее исходный процесс, достаточно пропустить его через фильтр со степенной функцией $K(\omega) \sim \omega^{-\Delta\beta/2}$.

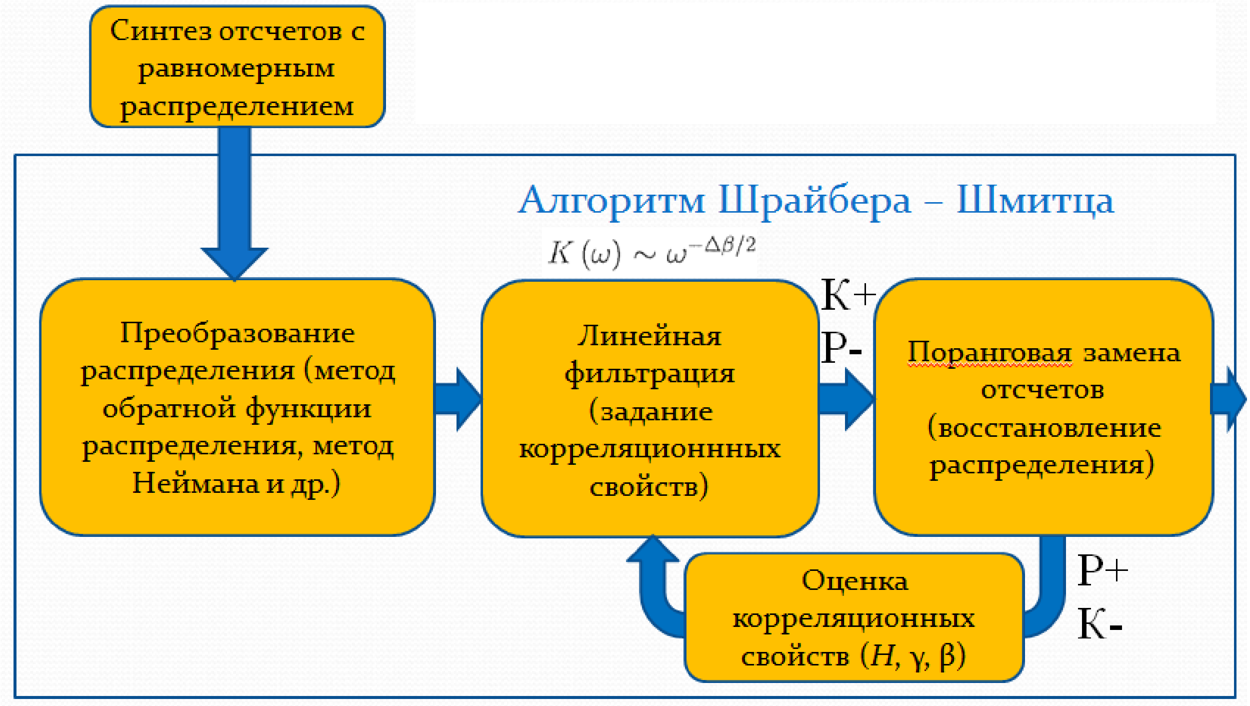

Рисунок 4 – Алгоритм Шрайбера-Шмитца