Замена непрерывного распределения на дискретное для применения на практике

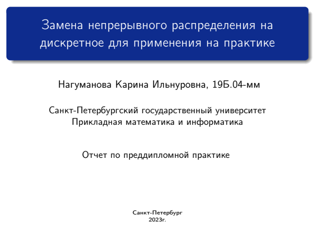
Нагуманова Карина Ильнуровна, 19Б.04-мм

Санкт-Петербургский государственный университет
Прикладная математика и информатика

Отчет по преддипломной практике

Санкт-Петербург
2023г.

Замена непрерывного распределения



Научный руководитель д.ф.-м.н., доцент Голяндина Н. Э., кафедра статистического моделирования

В практических задачах нередко требуется заменить непрерывное распределение на дискретное с сохранением математического ожидания и дисперсии. Одним из методов нахождения такого распределения для трехточечной аппроксимации нормального распределения является **метод Свонсона**.

Аппроксимируемые случайные величины складывают и умножают.

Задача: находить аппроксимацию суммы и произведения логнормальных случайных величин по аппроксимациям исходных случайных величин.

Введение

В практических задачах нередко требуется заменить непрерывное распределение на дискретное с сохранением математического ожидания и дисперсии. Одним из методов нахождения такого распределения для трехточечной аппроксимации нормального распределения является метод Свонсона.

Аппроксимируемые случайные величины складывают и умножают.

Задача: находить аппроксимацию суммы и произведения логнормальных случайных величин по аппроксимациям исходных случайных величин.

В практических задачах нередко требуется заменить непрерывное распределение на дискретное с сохранением математического ожидания и дисперсии. Одним из методов нахождения такого распределения для трехточечной аппроксимации нормального распределения является метод Свонсона. Однако в ряде областей, например, в нефтяной промышленности, общепринятым распределением, описывающим запасы нефти, является логнормальное распределение. Соответственно, реальной задачей является аппроксимация логнормального распределения. При этом здесь тоже применим метод Свонсона, потому что логнормальное распределение можно свести к нормальному. Ставится задача нахождения аппроксимации суммы и произведения логнормальных случайных величин по аппроксимациям исходных случайных величин.

Введение

План работы.

1. Рассмотреть общий подход к трехточечной аппроксимации.
2. Рассмотреть трехточечную аппроксимацию нормального распределения, в целом метод Свонсона и вывод правила 30-40-30.
3. Рассмотреть трехточечную аппроксимацию логнормального распределения и её свойства.
4. Построить алгоритм аппроксимации произведения двух логнормальных распределений.
5. Построить алгоритм аппроксимации суммы двух логнормальных распределений.

Введение

План работы.

1. Рассмотреть общий подход к трехточечной аппроксимации.
2. Рассмотреть трехточечную аппроксимацию нормального распределения, в целом метод Свонсона и вывод правила 30-40-30.
3. Рассмотреть трехточечную аппроксимацию логнормального распределения и её свойства.
4. Построить алгоритм аппроксимации произведения двух логнормальных распределений.
5. Построить алгоритм аппроксимации суммы двух логнормальных распределений.

Имеем следующий план работы. Рассмотреть общий подход к трехточечной аппроксимации, рассмотреть трехточечную аппроксимацию нормального распределения, в целом метод Свонсона и вывод правила 30-40-30, рассмотреть трехточечную аппроксимацию логнормального распределения и её свойства, построить алгоритм аппроксимации произведения двух логнормальных распределений, построить алгоритм аппроксимации суммы двух логнормальных распределений.

Часть 1: Общий подход к трехточечной аппроксимации

Пусть ξ — непрерывная случайная величина, обозначим

$$m = \mathbf{E}(\xi), \qquad s^2 = \mathbf{D}(\xi),$$

$F(x)$ — функция распределения, x_{π_1}, x_{π_2}, x_{π_3} — квантили ξ, $\tilde{\xi}$ — случайная дискретная величина

$$\tilde{\xi}: \quad \begin{pmatrix} x_{\pi_1} & x_{\pi_2} & x_{\pi_3} \\ p_1 & p_2 & p_3 \end{pmatrix}$$

$$\tilde{m} = \mathbf{E}(\tilde{\xi}), \qquad \tilde{s}^2 = \mathbf{D}(\tilde{\xi}).$$

Задача: аппроксимировать распределение случайной величины ξ дискретным распределением $\tilde{\xi}$, то есть найти p_1, p_2, p_3 такие, что

$$p_1 + p_2 + p_3 = 1,$$

$$\tilde{m} = p_1 x_{\pi_1} + p_2 x_{\pi_2} + p_3 x_{\pi_3} = m,$$

$$\tilde{s}^2 = p_1 x_{\pi_1}^2 + p_2 x_{\pi_2}^2 + p_3 x_{\pi_3}^2 - m^2 = s^2.$$

Часть 1: Общий подход к трехточечной аппроксимации
Пусть ξ — непрерывная случайная величина, обозначим
$m = \mathbf{E}(\xi), \quad s^2 = \mathbf{D}(\xi),$
$F(x)$ — функция распределения, x_{π_1}, x_{π_2}, x_{π_3} — квантили ξ, $\tilde{\xi}$ — случайная дискретная величина
$\tilde{\xi}: \begin{pmatrix} x_{\pi_1} & x_{\pi_2} & x_{\pi_3} \\ p_1 & p_2 & p_3 \end{pmatrix}$
$\tilde{m} = \mathbf{E}(\tilde{\xi}), \quad \tilde{s}^2 = \mathbf{D}(\tilde{\xi}).$
Задача: аппроксимировать распределение случайной величины ξ дискретным распределением $\tilde{\xi}$, то есть найти p_1, p_2, p_3 такие, что
$p_1 + p_2 + p_3 = 1,$
$\tilde{m} = p_1 x_{\pi_1} + p_2 x_{\pi_2} + p_3 x_{\pi_3} = m,$
$\tilde{s}^2 = p_1 x_{\pi_1}^2 + p_2 x_{\pi_2}^2 + p_3 x_{\pi_3}^2 - m^2 = s^2.$

Пусть ξ — непрерывная случайная величина, вводим основные обозначения для ξ, так же вводим дискретную случайную величину, которой будем аппроксимировать ξ. Ставится задача аппроксимации распределения ξ дискретным распределением $\tilde{\xi}$, то есть нужно вычислить p_1, p_2, p_3 такие, что верны следующие равенства.

Часть 1: Общий подход к трехточечной аппроксимации

Предложение (Swanson)

Пусть верно

$$\begin{pmatrix} 1 & 1 & 1 \\ \hat{x}_{\pi 1} & \hat{x}_{\pi 2} & \hat{x}_{\pi 3} \\ \hat{x}^2_{\pi 1} & \hat{x}^2_{\pi 2} & \hat{x}^2_{\pi 3} \end{pmatrix} \begin{pmatrix} p_1 \\ p_2 \\ p_3 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix},$$

где $\hat{x}_{\pi_i} = \hat{\mathsf{F}}^{-1}(\pi_i)$, $\hat{\mathsf{F}}(y)$ *— функция распределения* $\hat{\xi} = \dfrac{\xi - m}{s}$.
Тогда $m = \tilde{m}$ *и* $s^2 = \tilde{s}^2$.

Часть 1: Общий подход к трехточечной аппроксимации
Предложение (Swanson)
Пусть верно
где $\hat{x}_{\pi_i} = \hat{\mathsf{F}}^{-1}(\pi_i)$, $\hat{\mathsf{F}}(y)$ — функция распределения $\hat{\xi} = \frac{\xi - m}{s}$.
Тогда $m = \tilde{m}$ и $s^2 = \tilde{s}^2$.

Пусть верна следующая система, тогда $m = \tilde{m}$ и $s^2 = \tilde{s}^2$. Это Предложение дает требуемую аппроксимацию дискретным распределением, если найденные вероятности p_i являются неотрицательными.

Часть 2: Аппроксимация нормального распределения

$$\begin{pmatrix} 1 & 1 & 1 \\ \Phi^{-1}(\pi) & 0 & -\Phi^{-1}(\pi) \\ \Phi^{-1}(\pi)^2 & 0 & \Phi^{-1}(\pi)^2 \end{pmatrix} \begin{pmatrix} p_\pi \\ p_{0.5} \\ p_{1-\pi} \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}.$$

Предложение (Swanson)

$\xi \sim N(\mu, \sigma)$, *пусть верно*

$$\begin{cases} p_\pi = p_{1-\pi} = \dfrac{\delta}{2}, \\ p_{0.5} = 1 - \delta, \end{cases}$$

где $\delta = \dfrac{1}{\Phi^{-1}(\pi)^2}$. *Тогда* $m = \tilde{m}$ *и* $s^2 = \tilde{s}^2$.

Часть 2: Аппроксимация нормального распределения

$$\begin{pmatrix} 1 & 1 & 1 \\ \Phi^{-1}(\pi) & 0 & -\Phi^{-1}(\pi) \\ \Phi^{-1}(\pi)^2 & 0 & \Phi^{-1}(\pi)^2 \end{pmatrix} \begin{pmatrix} p_\pi \\ p_{0.5} \\ p_{1-\pi} \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}.$$

Предложение (Swanson)

$\xi \sim N(\mu, \sigma)$, пусть верно

$$\begin{cases} p_\pi = p_{1-\pi} = \dfrac{\delta}{2}, \\ p_{0.5} = 1 - \delta, \end{cases}$$

где $\delta = \dfrac{1}{\Phi^{-1}(\pi)^2}$. Тогда $m = \tilde{m}$ и $s^2 = \tilde{s}^2$.

Если $\xi \sim N(\mu, \sigma)$ имеет нормальное распределение, то $\hat{\xi}$ имеет стандартное нормальное распределение, поэтому $\hat{\xi} \sim N(0, 1)$ в предыдущем Предложении. Поэтому значения вероятностей можно выразить через функцию распределения стандартного нормального распределения.

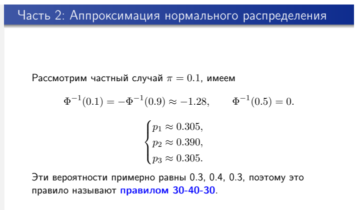
Часть 2: Аппроксимация нормального распределения

Рассмотрим частный случай $\pi = 0.1$, имеем

$$\Phi^{-1}(0.1) = -\Phi^{-1}(0.9) \approx -1.28, \qquad \Phi^{-1}(0.5) = 0.$$

$$\begin{cases} p_1 \approx 0.305, \\ p_2 \approx 0.390, \\ p_3 \approx 0.305. \end{cases}$$

Эти вероятности примерно равны 0.3, 0.4, 0.3, поэтому это правило называют **правилом 30-40-30**.



Рассматриваем частный случай $\pi = 0.1$. Для него вероятности примерно равны 0.3, 0.4, 0.3, поэтому это правило называют правилом 30-40-30.

Часть 3: Аппроксимация логнормального распределения

Дано: квантили $x_{\pi_1}, x_{\pi_2}, x_{\pi_3}$ логнормальной случайной величины η, $\ln(\eta) \sim N(\mu, \sigma)$.

1. Вычисляем значения мат. ожидания m и дисперсии s^2 случайной величины η, используя известные $x_{\pi_1}, x_{\pi_2}, x_{\pi_3}$.
2. Выражаем параметры μ и σ мат. ожидание и дисперсию соответствующего нормального распределения через параметры m и s^2 логнормального распределения
3. С помощью системы уравнений из метода для нормального распределения находим значения весов p_1, p_2, p_3.

Результат: веса p_1, p_2, p_3 для $x_{\pi_1}, x_{\pi_2}, x_{\pi_3}$ случайной величины $\tilde{\xi}$.

Построим алгоритм для нахождения трехточечной симметричной аппроксимации логнормального распределения. Заданы три квантиля $x_{\pi_1}, x_{\pi_2}, x_{\pi_3}$, вычисляем по ним мат. ожидание m и дисперсию s^2 случайной величины η. Затем выражаем параметры μ и σ и наконец находим значения весов p_1, p_2, p_3.

Часть 3: Условие на параметр σ

Мною доказаны следующие предложения.

Предложение

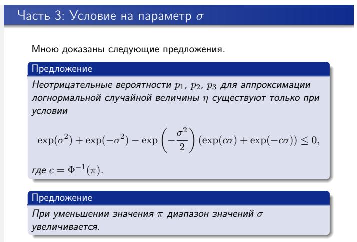
Неотрицательные вероятности p_1, p_2, p_3 для аппроксимации логнормальной случайной величины η существуют только при условии

$$\exp(\sigma^2) + \exp(-\sigma^2) - \exp\left(-\frac{\sigma^2}{2}\right)(\exp(c\sigma) + \exp(-c\sigma)) \leq 0,$$

где $c = \Phi^{-1}(\pi)$.

Предложение

При уменьшении значения π диапазон значений σ увеличивается.



Мы рассмотрели способ вычисления весов для квантилей при аппроксимации логнормального распределения. Но найденные веса являются вероятностями не при любом σ. Выясним, какое должно быть ограничение на этот параметр. В следующем предложении получено условие на σ для существования трехточечной симметричной аппроксимации логнормального распределения. При уменьшении значения π диапазон значений σ увеличивается.

Часть 3: Варианты постановки задачи

Задача: имеются квантили x_π, $x_{0.5}$, $x_{1-\pi}$ логнормальной случайной величины η. Нужно уметь считать m и s^2.

1. Используя значения двух квантилей, найти значения параметров μ и σ нормальной случайной величины $\ln(\eta) \sim N(\mu, \sigma)$. Через них вычислить значения m и s^2.
2. Найти значения p_1, p_2, p_3 такие, что

$$p_1 x_{\pi_1} + p_2 x_{\pi_2} + p_3 x_{\pi_3} = m,$$

$$p_1 x_{\pi_1}^2 + p_2 x_{\pi_2}^2 + p_3 x_{\pi_3}^2 - m^2 = s^2.$$

Часть 3: Варианты постановки задачи

Задача: имеются квантили x_π, $x_{0.5}$, $x_{1-\pi}$ логнормальной случайной величины η. Нужно уметь считать m и s^2.

1. Используя значения двух квантилей, найти значения параметров μ и σ нормальной случайной величины $\ln(\eta) \sim N(\mu, \sigma)$. Через них вычислить значения m и s^2.
2. Найти значения p_1, p_2, p_3 такие, что

$$p_1 x_{\pi_1} + p_2 x_{\pi_2} + p_3 x_{\pi_3} = m,$$

$$p_1 x_{\pi_1}^2 + p_2 x_{\pi_2}^2 + p_3 x_{\pi_3}^2 - m^2 = s^2.$$

Рассмотрим варианты постановки задачи поиска m и s^2. Можно не переходить к аппроксимации дискретным распределением, а сразу вычислить моменты, используя значения квантилей. Второй вариант искать значения p_1, p_2, p_3. Если они положительные, то рассматривается аппроксимация дискретной $\tilde{\xi}$, у которой $\tilde{m} = m$ и $\tilde{s}^2 = s^2$. Если не все положительные, то можно воспринимать задачу формально, как поиск коэффициентов линейной комбинации x_π, $x_{0.5}$, $x_{1-\pi}$.

Часть 3: Точность неправильной аппроксимации

Проблема: метод Свонсона выведенный для аппроксимации нормального распределения используют для логнормального.
Вопрос: какова точность аппроксимации m и s^2?

Предложение

Ошибка аппроксимации мат.ожидания логнормального распределения по методу Свонсона, выведенному для аппроксимации нормального распределения, равна

$$\frac{|\, m - \widetilde{m} \,|}{m} = \left| \exp\left(\frac{\sigma^2}{2}\right) - \frac{1}{2c^2} \times \right.$$

$$\left. \times (\exp(c\sigma) - 1 + \exp(-c\sigma)) + 1 \right| / \exp\left(\frac{\sigma^2}{2}\right),$$

где $c = \Phi^{-1}(\pi)$, и не зависит от параметра μ.

Часть 3: Точность неправильной аппроксимации
Проблема: метод Свонсона выведенный для аппроксимации нормального распределения используют для логнормального.
Вопрос: какова точность аппроксимации m и s^2?
Предложение
Ошибка аппроксимации мат.ожидания логнормального распределения по методу Свонсона, выведенному для аппроксимации нормального распределения, равна
$$\frac{|\, m - \widetilde{m} \,|}{m} = \left| \exp\left(\frac{\sigma^2}{2}\right) - \frac{1}{2c^2} \times (\exp(c\sigma) - 1 + \exp(-c\sigma)) + 1 \right| / \exp\left(\frac{\sigma^2}{2}\right),$$
где $c = \Phi^{-1}(\pi)$, и не зависит от параметра μ.

Предлагаемые методы аппроксимации трехточечным дискретным распределением логнормального распределения не работают при $\sigma \leq 0.6913$. На практике часто используют метод Свонсона выведенный для аппроксимации нормального распределения. Мною доказано это предложение. Ошибка аппроксимации мат. ожидания выражается следующим образом и не зависит от параметра μ.

Часть 3: Точность метода Свонсона для логнормального распределения

Предложение

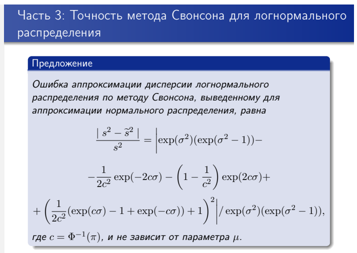
Ошибка аппроксимации дисперсии логнормального распределения по методу Свонсона, выведенному для аппроксимации нормального распределения, равна

$$\frac{\mid s^2 - \widetilde{s}^2 \mid}{s^2} = \Bigg| \exp(\sigma^2)(\exp(\sigma^2 - 1)) -$$

$$- \frac{1}{2c^2} \exp(-2c\sigma) - \left(1 - \frac{1}{c^2}\right) \exp(2c\sigma) +$$

$$+ \left(\frac{1}{2c^2} (\exp(c\sigma) - 1 + \exp(-c\sigma)) + 1 \right)^2 \Bigg| / \exp(\sigma^2)(\exp(\sigma^2 - 1)),$$

где $c = \Phi^{-1}(\pi)$, и не зависит от параметра μ.



Ошибка аппроксимации дисперсии выражается следующим образом и тоже не зависит от параметра μ.

Часть 3: Точность метода Свонсона для логнормального распределения

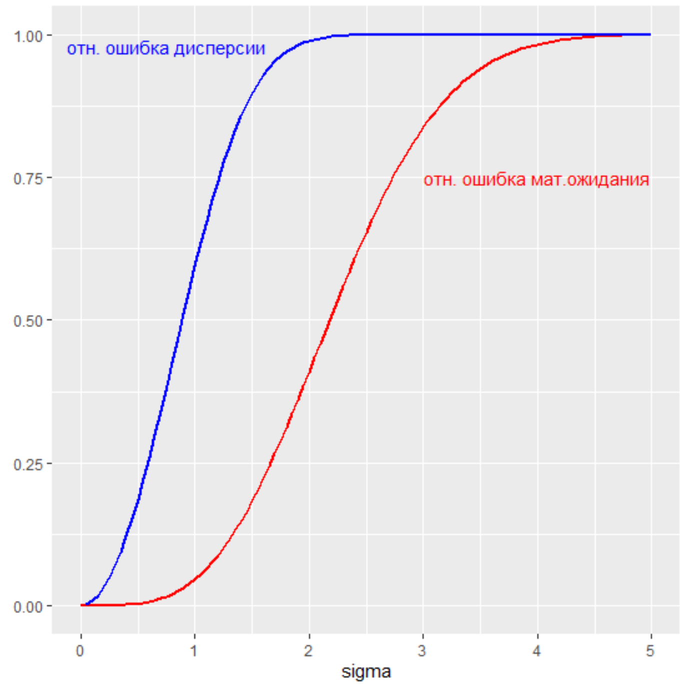


Рис.: Относительная ошибка аппроксимации мат.ож. и дисперсии

Теперь посмотрим на график зависимости ошибок аппроксимации мат.ожидания и дисперсии от параметра σ. Видим, что при $\sigma \leq 1.5$ ошибка аппроксимации мат.ожидания меньше 12%, а ошибка аппроксимации дисперсии может достигать 80%. Для интересующих нас значений $\sigma \geq 0.69$ ошибка мат. ожидания может быть как маленькой, так и очень большой. Ошибка дисперсии при этом точно больше 25%.

Часть 4: Произведение двух логнормальных распределений

Рассмотрим две логнормальные случайные величины

- $\ln(\xi_1) \sim N(\mu_1, \sigma_1^2)$,
- $\ln(\xi_2) \sim N(\mu_2, \sigma_2^2)$,

которые заданы своими квантилями

- x_π, $x_{0.5}$, $x_{1-\pi}$ — симметричные квантили ξ_1,
- y_π, $y_{0.5}$, $y_{1-\pi}$ — симметричные квантили ξ_2.

Задача: аппроксимировать непрерывную случайную величину $\eta = \xi_1 \xi_2$ дискретной, то есть найти квантили вида z_π, $z_{0.5}$, $z_{1-\pi}$.

Часть 4: Произведение двух логнормальных распределений
Рассмотрим две логнормальные случайные величины
- $\ln(\xi_1) \sim N(\mu_1, \sigma_1^2)$,
- $\ln(\xi_2) \sim N(\mu_2, \sigma_2^2)$,
которые заданы своими квантилями
- x_π, $x_{0.5}$, $x_{1-\pi}$ — симметричные квантили ξ_1,
- y_π, $y_{0.5}$, $y_{1-\pi}$ — симметричные квантили ξ_2.
Задача: аппроксимировать непрерывную случайную величину $\eta = \xi_1 \xi_2$ дискретной, то есть найти квантили вида z_π, $z_{0.5}$, $z_{1-\pi}$.

Рассмотрим произведение логнормально распределенных случайных величин. Эта процедура применяется в нефтяной промышленности. Для случайных величин ξ_1 и ξ_2 заданы наборы квантилей. Ставим задачу аппроксимировать случайную величину $\eta = \xi_1 \xi_2$ дискретной, то есть найти квантили вида z_π, $z_{0.5}$, $z_{1-\pi}$.

Часть 4: Произведение двух логнормальных распределений

Предложение (Swanson)

Зная квантили x_π, $x_{0.5}$, $x_{1-\pi}$ случайной величины ξ_1 и квантили y_π, $y_{0.5}$, $y_{1-\pi}$ случайной величины ξ_2 можно найти квантили z_π, $z_{0.5}$, $z_{1-\pi}$ случайной величины $\xi_1\xi_2$, как

$$z_\pi = \exp(b\Phi^{-1}(\pi) + a),$$

$$z_{0.5} = x_{0.5}y_{0.5},$$

$$z_{1-\pi} = \exp(b\Phi^{-1}(1-\pi) + a),$$

где a и b такие, что прямая $y = \dfrac{x-a}{b}$, проходит через точки $(\ln(x_\pi y_\pi), t)$ и $(\ln(x_{0.5}y_{0.5}), 0)$ при

$$t = \frac{\Phi^{-1}(\pi)((\ln(x_{0.5}) + \ln(y_{0.5})) - (\ln(x_\pi) + \ln(y_\pi)))}{\sqrt{(\ln(x_{0.5}) - \ln(x_\pi))^2 + (\ln(y_{0.5}) - \ln(y_\pi))^2}}.$$

Часть 4: Произведение двух логнормальных распределений

Предложение (Swanson)

Зная квантили x_π, $x_{0.5}$, $x_{1-\pi}$ случайной величины ξ_1 и квантили y_π, $y_{0.5}$, $y_{1-\pi}$ случайной величины ξ_2 можно найти квантили z_π, $z_{0.5}$, $z_{1-\pi}$ случайной величины $\xi_1\xi_2$, как

$$z_\pi = \exp(b\Phi^{-1}(\pi) + a),$$

$$z_{0.5} = x_{0.5}y_{0.5},$$

$$z_{1-\pi} = \exp(b\Phi^{-1}(1-\pi) + a),$$

где a и b такие, что прямая $y = \dfrac{x-a}{b}$, проходит через точки $(\ln(x_\pi y_\pi), t)$ и $(\ln(x_{0.5}y_{0.5}), 0)$ при

$$t = \frac{\Phi^{-1}(\pi)((\ln(x_{0.5}) + \ln(y_{0.5})) - (\ln(x_\pi) + \ln(y_\pi)))}{\sqrt{(\ln(x_{0.5}) - \ln(x_\pi))^2 + (\ln(y_{0.5}) - \ln(y_\pi))^2}}.$$

Следующее предложение было мною подробно доказано, но сама идея доказательства взята из статьи. Зная квантили x_π, $x_{0.5}$, $x_{1-\pi}$ случайной величины ξ_1 и квантили y_π, $y_{0.5}$, $y_{1-\pi}$ случайной величины ξ_2 можно найти квантили z_π, $z_{0.5}$, $z_{1-\pi}$ случайной величины $\xi_1\xi_2$.

Часть 5: Сумма двух логнормальных распределений

Рассмотрим сумму двух логнормальных случайных величин.

$$\ln(\xi_1) \sim N(\mu_1, \sigma_1^2),$$

$$\ln(\xi_2) \sim N(\mu_2, \sigma_2^2),$$

$$\xi = \xi_1 + \xi_2.$$

ξ_1 и ξ_2 заданы своими квантилями.

Поставим задачу аппроксимации суммы логнормальным распределением $\ln(\eta) \sim N(\mu, \sigma)$, так как нужно рассматривать сумму не обязательно двух, а произвольного числа случайных величин.

Задача: найти квантили z_π, $z_{0.5}$, $z_{1-\pi}$ случайной величины η.

Часть 5: Сумма двух логнормальных распределений
Рассмотрим сумму двух логнормальных случайных величин.
$\ln(\xi_1) \sim N(\mu_1, \sigma_1^2)$,
$\ln(\xi_2) \sim N(\mu_2, \sigma_2^2)$,
$\xi = \xi_1 + \xi_2$.
ξ_1 и ξ_2 заданы своими квантилями.
Поставим задачу аппроксимации суммы логнормальным распределением $\ln(\eta) \sim N(\mu, \sigma)$, так как нужно рассматривать сумму не обязательно двух, а произвольного числа случайных величин.
Задача: найти квантили z_π, $z_{0.5}$, $z_{1-\pi}$ случайной величины η.

Рассмотрим сумму двух логнормальных случайных величин,которые заданы наборами своих квантилей. Поставим задачу аппроксимации суммы логнормальным распределением $\ln(\eta) \sim N(\mu, \sigma)$, так как нужно рассматривать сумму не обязательно двух, а произвольного числа случайных величин. По известным квантилям уже знаем, как вычислить вероятности p_1, p_2, p_3 такие, что $m = \tilde{m}$ и $s^2 = \tilde{s}^2$.

Часть 5: Сумма двух логнормальных распределений

Дано: Квантили x_π, $x_{0.5}$, $x_{1-\pi}$ — квантили ξ_1, y_π, $y_{0.5}$, $y_{1-\pi}$ — квантили ξ_2.

1. x_π, $x_{0.5}$, $x_{1-\pi} \to \mu_1$, σ_1
2. y_π, $y_{0.5}$, $y_{1-\pi} \to \mu_2$, σ_2
3. μ_i, $\sigma_i \to m_i$, s_i^2
4. $m = m_1 + m_2$
5. $s^2 = s_1^2 + s_2^2$
6. m, $s^2 \to \mu$, σ
7. μ, $\sigma \to z_\pi$, $z_{0.5}$, $z_{1-\pi}$
8. z_π, $z_{0.5}$, $z_{1-\pi} \to p_1$, p_2, p_3

Результат: вероятности p_1, p_2, p_3 для квантилей $z_{\pi_1}, z_{\pi_2}, z_{\pi_3}$ случайной величины $\xi_1 + \xi_2$.

Часть 5: Сумма двух логнормальных распределений
Дано: Квантили x_π, $x_{0.5}$, $x_{1-\pi}$ — квантили ξ_1, y_π, $y_{0.5}$, $y_{1-\pi}$ — квантили ξ_2.
1. x_π, $x_{0.5}$, $x_{1-\pi} \to \mu_1$, σ_1
2. y_π, $y_{0.5}$, $y_{1-\pi} \to \mu_2$, σ_2
3. μ_i, $\sigma_i \to m_i$, s_i^2
4. $m = m_1 + m_2$
5. $s^2 = s_1^2 + s_2^2$
6. m, $s^2 \to \mu$, σ
7. μ, $\sigma \to z_\pi$, $z_{0.5}$, $z_{1-\pi}$
8. z_π, $z_{0.5}$, $z_{1-\pi} \to p_1$, p_2, p_3
Результат: вероятности p_1, p_2, p_3 для квантилей $z_{\pi_1}, z_{\pi_2}, z_{\pi_3}$ случайной величины $\xi_1 + \xi_2$.

Получен следующий алгоритм аппроксимации суммы двух логнормальных распределений. По набору квантилей ξ_1 находим параметры μ_1, σ_1 соответствующего нормального распределения. По набору квантилей ξ_2 находим параметры μ_2, σ_2 соответствующего нормального распределения. Потом находим мат. ожидания и дисперсии ξ_1 и ξ_2. Затем вычисляем мат.ожидание и дисперсию $\xi_1 + \xi_2$. Через них находим параметры нормального распределения и затем находим значения квантилей через μ и σ.

Часть 5: Сумма двух логнормальных распределений. Точность аппроксимации

Ошибки аппроксимации квантилей q_π, $q_{0.5}$, $q_{1-\pi}$ случайной величины ξ равны

$$\frac{|q_\pi - z_\pi|}{q_\pi}, \qquad \frac{|q_{0.5} - z_{0.5}|}{q_{0.5}}, \qquad \frac{|q_{1-\pi} - z_{1-\pi}|}{q_{1-\pi}},$$

где

$$z_{100p} = F_\eta^{-1}(p) = \exp(\mu + \sigma\sqrt{2}\mathrm{erf}^{-1}(2p-1)).$$

Значение квантилей q_i выражаются как $q_{100p} = F_\xi^{-1}(p)$, где

$$F_\xi(x) = \int_0^x \left(\frac{1}{2} + \frac{1}{2}\mathrm{erf}\left(\frac{\ln(x-y) - \mu_1}{\sigma_1\sqrt{2}}\right)\right) \times$$

$$\times \left(\frac{1}{\sqrt{2\pi}y\sigma_2}\exp\left(-\left(\frac{\ln(y) - \mu_2}{\sqrt{2}\sigma_2}\right)^2\right)\right) dy.$$

Выразим ошибки аппроксимации квантилей q_π, $q_{0.5}$, $q_{1-\pi}$ случайной величины ξ, здесь параметры μ, σ можно найти через параметры случайных величин ξ_1, ξ_2 и вычисленные значения $m = \exp\left(\mu_1 + \frac{\sigma_1^2}{2}\right) + \exp\left(\mu_2 + \frac{\sigma_2^2}{2}\right)$, $s^2 = m_1^2(\exp(\sigma_1^2) - 1) + m_2^2(\exp(\sigma_2^2) - 1)$, так как для независимых случайных величин мат.ожидание суммы равно сумме мат.ожиданий, дисперсия суммы равна сумме дисперсий. Функция $F_\xi(x)$ получена по формуле свертки.

Часть 5: Сумма двух логнормальных распределений. Точность аппроксимации

Рассмотрим $\ln(\xi_1) \sim N(4, \sigma_1^2)$, $\ln(\xi_2) \sim N(4, \sigma_2^2)$ и найдем ошибки с помощью моделирования, объемы выборок равны 10^6.

Таблица: Ошибка аппроксимации медианы (%) в зависимости от σ_1^2 (строка) и σ_2^2 (столбец)

	0.25	**0.75**	**1.25**	**1.75**	**2.25**
0.25	0.24	0.46	4.19	11.67	21.10
0.75	0.74	0.40	3.06	11.46	20.99
1.25	4.25	3.27	2.48	6.18	16.15
1.75	12.18	10.12	5.57	5.24	9.92
2.25	20.94	20.20	16.29	9.59	8.47

Часть 5: Сумма двух логнормальных распределений. Точность аппроксимации

Рассмотрим $\ln(\xi_1) \sim N(4, \sigma_1^2)$, $\ln(\xi_2) \sim N(4, \sigma_2^2)$ и найдем ошибки с помощью моделирования, объемы выборок равны 10^6.

Таблица: Ошибка аппроксимации медианы (%) в зависимости от σ_1^2 (строка) и σ_2^2 (столбец)

	0.25	**0.75**	**1.25**	**1.75**	**2.25**
0.25	0.24	0.46	4.19	11.67	21.10
0.75	0.74	0.40	3.06	11.46	20.99
1.25	4.25	3.27	2.48	6.18	16.15
1.75	12.18	10.12	5.57	5.24	9.92
2.25	20.94	20.20	16.29	9.59	8.47

Рассмотрим $\ln(\xi_1) \sim N(4, \sigma_1^2)$, $\ln(\xi_2) \sim N(4, \sigma_2^2)$ и найдем ошибки с помощью моделирования, объемы выборок равны 10^6. В данной таблице представлены ошибки для аппроксимации медианы. По построению аппроксимации суммы двух логнормальных распределений логнормальным распределением ошибки мат. ожидания и дисперсии равны 0, то есть $m = \tilde{m}$ и $s^2 = \tilde{s}^2$. Но если для каких-либо расчетов понадобятся квантили η, то ошибка медианы может достигать 21%.

Часть 5: Сумма двух логнормальных распределений. Точность аппроксимации

Таблица: Ошибка аппроксимации q_{10} (%) в зависимости от σ_1^2 (строка) и σ_2^2 (столбец)

	0.25	**0.75**	**1.25**	**1.75**	**2.5**
0.25	2.35	13.59	23.93	33.20	42.75
0.75	1.20	10.54	21.80	33.93	42.82
1.25	3.02	7.03	18.43	29.49	40.09
1.75	14.45	5.27	14.33	26.50	36.75
2.5	34.70	11.44	11.10	23.05	32.84

Часть 5: Сумма двух логнормальных распределений. Точность аппроксимации

Таблица: Ошибка аппроксимации q_{10} (%) в зависимости от σ_1^2 (строка) и σ_2^2 (столбец)

	0.25	0.75	1.25	1.75	2.5
0.25	2.35	13.59	23.93	33.20	42.75
0.75	1.20	10.54	21.80	33.93	42.82
1.25	3.02	7.03	18.43	29.49	40.09
1.75	14.45	5.27	14.33	26.50	36.75
2.5	34.70	11.44	11.10	23.05	32.84

В данной таблице представлены ошибки для аппроксимации q_{10}, они могут достигать 67%.

Часть 5: Сумма двух логнормальных распределений. Точность аппроксимации

Таблица: Ошибка аппроксимации q_{90} (%) в зависимости от σ_1^2 (строка) и σ_2^2 (столбец)

	0.25	**0.75**	**1.25**	**1.75**	**2.25**
0.25	0.16	5.03	13.69	17.74	19.37
0.75	5.63	1.89	5.77	10.70	16.03
1.25	13.55	5.75	2.52	6.00	9.79
1.75	19.95	11.88	5.77	3.50	4.89
2.25	18.47	15.44	9.42	5.50	5.27

Часть 5: Сумма двух логнормальных распределений. Точность аппроксимации

Таблица: Ошибка аппроксимации q_{90} (%) в зависимости от σ_1^2 (строка) и σ_2^2 (столбец)

	0.25	**0.75**	**1.25**	**1.75**	**2.25**
0.25	0.16	5.03	13.69	17.74	19.37
0.75	5.63	1.89	5.77	10.70	16.03
1.25	13.55	5.75	2.52	6.00	9.79
1.75	19.95	11.88	5.77	3.50	4.89
2.25	18.47	15.44	9.42	5.50	5.27

В данной таблице представлены ошибки для аппроксимации q_{90}, они могут достигать 20%.

Часть 5: Сумма двух логнормальных распределений. Точность аппроксимации

Таблица: Коэффициент асимметрии суммы (голубой) и аппроксимации (розовый) в зависимости от σ_1^2 (строка) и σ_2^2 (столбец)

	0.25	0.75	1.25	1.75	2.5
0.25	1.77	4.23	6.71	15.59	16.68
	1.53	3.75	7.48	14.76	29.70
0.75	1.66	3.86	7.39	11.43	54.43
	1.55	3.65	7.22	14.25	28.77
1.25	2.13	3.68	8.73	13.76	29.28
	1.71	3.60	6.97	13.68	27.66
1.75	5.88	4.06	7.50	31.50	24.89
	2.17	3.71	6.79	13.09	26.41
2.5	11.18	8.85	8.55	10.34	23.61
	3.30	4.29	6.90	12.66	25.13

Часть 5: Сумма двух логнормальных распределений. Точность аппроксимации

Таблица: Коэффициент асимметрии суммы (голубой) и аппроксимации (розовый) в зависимости от σ_1^2 (строка) и σ_2^2 (столбец)

	0.25	0.75	1.25	1.75	2.5
0.25	1.77	4.23	6.71	15.59	16.68
	1.53	3.75	7.48	14.76	29.70
0.75	1.66	3.86	7.39	11.43	54.43
	1.55	3.65	7.22	14.25	28.77
1.25	2.13	3.68	8.73	13.76	29.28
	1.71	3.60	6.97	13.68	27.66
1.75	5.88	4.06	7.50	31.50	24.89
	2.17	3.71	6.79	13.09	26.41
2.5	11.18	8.85	8.55	10.34	23.61
	3.30	4.29	6.90	12.66	25.13

Теперь посмотрим на таблицу с коэффициентами асимметрии для суммы $\xi_1+\xi_2$, они выделены голубым цветом и на коэффициенты асимметрии для аппроксимации, они выделены розовым цветом.

Часть 5: Сумма двух логнормальных распределений. Точность аппроксимации

Таблица: Коэффициент эксцесса суммы (голубой) и аппроксимации (розовый) в зависимости от σ_1^2 (строка) и σ_2^2 (столбец)

	0.25	**0.75**	**1.25**	**1.75**	**2.5**
0.25	6.54	51.70	227.68	408.58	734.47
	4.42	32.60	180.39	1088.57	7274.56
0.75	6.21	61.66	144.59	201.69	1304.88
	4.56	30.53	164.86	990.42	6666.16
1.25	11.47	27.75	179.22	193.95	546.57
	5.61	29.53	150.21	886.71	5989.44
1.75	122.65	46.01	110.03	276.24	14081.05
	9.44	31.88	140.69	788.78	5280.07
2.5	195.77	283.81	344.56	4837.85	1292.23
	24.08	44.88	146.68	720.26	4612.33

Часть 5: Сумма двух логнормальных распределений. Точность аппроксимации

Таблица: Коэффициент эксцесса суммы (голубой) и аппроксимации (розовый) в зависимости от σ_1^2 (строка) и σ_2^2 (столбец)

	0.25	0.75	1.25	1.75	2.5
0.25	6.54	51.70	227.68	408.58	734.47
	4.42	32.60	180.39	1088.57	7274.56
0.75	6.21	61.66	144.59	201.69	1304.88
	4.56	30.53	164.86	990.42	6666.16
1.25	11.47	27.75	179.22	193.95	546.57
	5.61	29.53	150.21	886.71	5989.44
1.75	122.65	46.01	110.03	276.24	14081.05
	9.44	31.88	140.69	788.78	5280.07
2.5	195.77	283.81	344.56	4837.85	1292.23
	24.08	44.88	146.68	720.26	4612.33

Также можно посмотреть на коэффициенты эксцесса суммы и аппроксимации.

Часть 5: Сумма двух логнормальных распределений. Точность аппроксимации

Посчитаем значения функции $F_\xi(x)$ от квантилей z_{10}, z_{50}, z_{90} случайной величины η. Они показывают, каким квантилем для ξ являются квантили z_i. Результаты приведены в следующих таблицах.

Таблица: $F_\eta(z_{50})$ (%) в зависимости от σ_1^2 (строка) и σ_2^2 (столбец)

	0.25	**0.75**	**1.25**	**1.75**	**2.25**
0.25	50.10	49.87	46.96	42.18	36.95
0.75	49.87	49.82	48.30	44.55	39.74
1.25	46.96	48.30	49.00	47.19	43.31
1.75	42.18	44.55	47.19	47.91	46.03
2.25	36.95	39.74	43.31	46.03	46.73

Часть 5: Сумма двух логнормальных распределений. Точность аппроксимации

Посчитаем значения функции $F_\xi(x)$ от квантилей z_{10}, z_{50}, z_{90} случайной величины η. Они показывают, каким квантилем для ξ являются квантили z_i. Результаты приведены в следующих таблицах.

Таблица: $F_\eta(z_{50})$ (%) в зависимости от σ_1^2 (строка) и σ_2^2 (столбец)

	0.25	**0.75**	**1.25**	**1.75**	**2.25**
0.25	50.10	49.87	46.96	42.18	36.95
0.75	49.87	49.82	48.30	44.55	39.74
1.25	46.96	48.30	49.00	47.19	43.31
1.75	42.18	44.55	47.19	47.91	46.03
2.25	36.95	39.74	43.31	46.03	46.73

Посчитаем значения функции $F_\xi(x)$ от квантилей z_{10}, z_{50}, z_{90} случайной величины η. Они показывают, каким квантилем для ξ являются квантили z_i. В данной таблице приведены значения $F_\eta(z_{50})$. Видим, что они не сильно далеки от нужных 50%.

Часть 5: Сумма двух логнормальных распределений. Точность аппроксимации

Таблица: $F_\eta(z_{10})$ (%) в зависимости от σ_1^2 (строка) и σ_2^2 (столбец)

	0.25	**0.75**	**1.25**	**1.75**	**2.25**
0.25	9.79	5.84	1.82	0.32	0.04
0.75	5.84	8.89	6.45	3.14	1.19
1.25	1.82	6.45	7.85	6.00	3.35
1.75	0.32	3.14	6.00	6.89	5.43
2.25	0.04	1.19	3.35	5.43	6.08

Замена непрерывного распределения

└Часть 5: Сумма двух логнормальных распределений. Точность аппроксимации

Часть 5: Сумма двух логнормальных распределений. Точность аппроксимации

Таблица: $F_\eta(z_{10})$ (%) в зависимости от σ_1^2 (строка) и σ_2^2 (столбец)

	0.25	**0.75**	**1.25**	**1.75**	**2.25**
0.25	9.79	5.84	1.82	0.32	0.04
0.75	5.84	8.89	6.45	3.14	1.19
1.25	1.82	6.45	7.85	6.00	3.35
1.75	0.32	3.14	6.00	6.89	5.43
2.25	0.04	1.19	3.35	5.43	6.08

В данной таблице приведены значения $F_\eta(z_{10})$.

Часть 5: Сумма двух логнормальных распределений. Точность аппроксимации

Таблица: $F_\eta(z_{90})$ (%) в зависимости от σ_1^2 (строка) и σ_2^2 (столбец)

	0.25	**0.75**	**1.25**	**1.75**	**2.25**
0.25	90.08	91.47	92.31	92.42	92.19
0.75	91.47	90.38	91.11	91.83	92.02
1.25	92.31	91.11	90.57	90.93	91.31
1.75	92.42	91.83	90.93	90.62	90.75
2.25	92.19	92.02	91.31	90.75	90.56

Часть 5: Сумма двух логнормальных распределений. Точность аппроксимации

Таблица: $F_\eta(z_{90})$ (%) в зависимости от σ_1^2 (строка) и σ_2^2 (столбец)

	0.25	**0.75**	**1.25**	**1.75**	**2.25**
0.25	90.08	91.47	92.31	92.42	92.19
0.75	91.47	90.38	91.11	91.83	92.02
1.25	92.31	91.11	90.57	90.93	91.31
1.75	92.42	91.83	90.93	90.62	90.75
2.25	92.19	92.02	91.31	90.75	90.56

В данной таблице приведены значения $F_\eta(z_{90})$.

Часть 5: Сумма двух логнормальных распределений. Точность аппроксимации

Построим оценки плотности для ξ и η, когда ошибки имеют очень маленькие значения и когда достаточно большие.

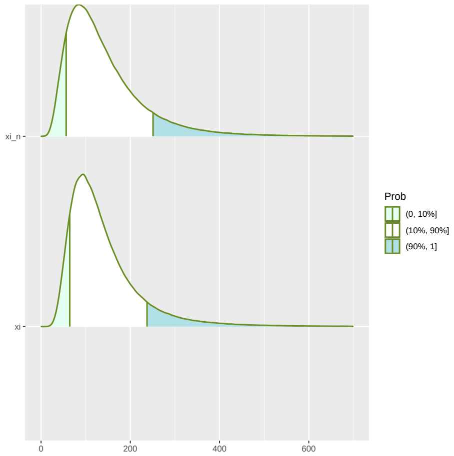
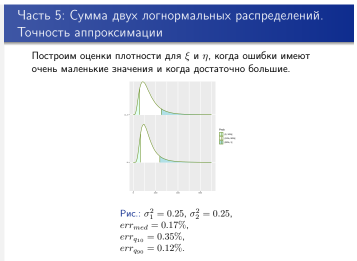


Рис.: $\sigma_1^2 = 0.25$, $\sigma_2^2 = 0.25$,
$err_{med} = 0.17\%$,
$err_{q10} = 0.35\%$,
$err_{q90} = 0.12\%$.



Построим оценки плотности для ξ и η, когда ошибки имеют очень маленькие значения и когда достаточно большие. На данном графике представлен случай $\sigma_1^2 = 0.25$, $\sigma_2^2 = 0.25$, при этом ошибки q_{10}, q_{50}, q_{90} близки к нулю.

Часть 5: Сумма двух логнормальных распределений. Точность аппроксимации

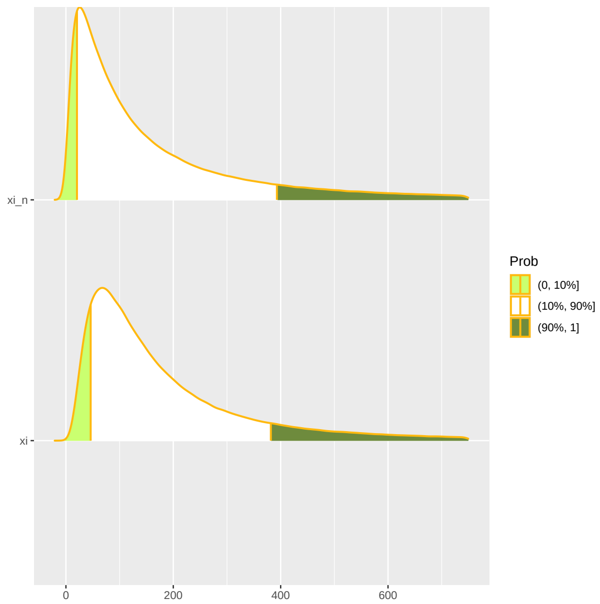
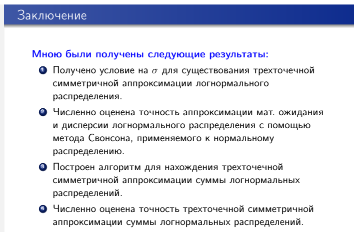


Рис.: $\sigma_1^2 = 2.25$, $\sigma_2^2 = 0.75$, $err_{med} = 20.4\%$, $err_{q_{10}} = 54.13\%$, $err_{q_{90}} = 15.54\%$.

Теперь построим оценки плотности для ξ и η, когда ошибки имеют большие значения. На данном графике представлен случай $\sigma_1^2 = 2.25$, $\sigma_2^2 = 0.75$.

Мною были получены следующие результаты:

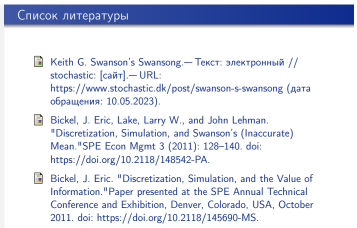
1. Получено условие на σ для существования трехточечной симметричной аппроксимации логнормального распределения.
2. Численно оценена точность аппроксимации мат. ожидания и дисперсии логнормального распределения с помощью метода Свонсона, применяемого к нормальному распределению.
3. Построен алгоритм для нахождения трехточечной симметричной аппроксимации суммы логнормальных распределений.
4. Численно оценена точность трехточечной симметричной аппроксимации суммы логнормальных распределений.



Таким образом, мною были получены следующие результаты. Получено условие на σ для существования трехточечной симметричной аппроксимации логнормального распределения. Численно оценена точность аппроксимации мат. ожидания и дисперсии логнормального распределения с помощью метода Свонсона, применяемого к нормальному распределению. Построен алгоритм для нахождения трехточечной симметричной аппроксимации суммы логнормальных распределений. Численно оценена точность трехточечной аппроксимации суммы логнормальных распределений.

Список литературы

- Keith G. Swanson's Swansong.— Текст: электронный // stochastic: [сайт].— URL: https://www.stochastic.dk/post/swanson-s-swansong (дата обращения: 10.05.2023).
- Bickel, J. Eric, Lake, Larry W., and John Lehman. "Discretization, Simulation, and Swanson's (Inaccurate) Mean."SPE Econ Mgmt 3 (2011): 128–140. doi: https://doi.org/10.2118/148542-PA.
- Bickel, J. Eric. "Discretization, Simulation, and the Value of Information."Paper presented at the SPE Annual Technical Conference and Exhibition, Denver, Colorado, USA, October 2011. doi: https://doi.org/10.2118/145690-MS.



Список литературы.