Санкт-Петербургский государственный университет
Прикладная математика и информатика

Отчет по научно-исследовательской работе

Замена непрерывных распределений на дискретные для применения на практике

(семестр 8)

Выполнила:
Нагуманова Карина Ильнуровна,
группа 19.Б04-мм

Научный руководитель:
к.ф.-м.н., доцент
Голяндина Нина Эдуардовна.
Кафедра статистического моделирования

Санкт-Петербург

2023

Содержание

1 Введение

В практических задачах нередко требуется заменить непрерывное распределение на дискретное с сохранением математического ожидания и дисперсии. Одним из методов нахождения такого распределения для аппроксимации нормального распределения является метод Свонсона [1]. Однако в ряде областей, например, в нефтяной промышленности распределением, описывающим запасы нефти, общепринятым является логнормальное распределение. Метод Свонсона используется в этих областях, хотя распределение и логнормальное. Соответственно, реальной задачей является аппроксимация логнормального распределения.

С аппроксимируемыми случайными величинами производят сложение и умножение. Например, используем площадь дренирования пласта, среднюю чистую толщину и коэффициент извлечения углеводородов. При перемножении этих параметров получаем количество резервов нефти. Или, зная запасы нефти в разных скважинах, нужно оценить суммарные запасы. Соответственно, возникает задача находить аппроксимацию суммы и произведения по аппроксимациям исходных случайных величин.

Часто бывает на практике, что вместо настоящего распределения известны три его квантили, стандартно это 10-, 50- и 90-процентили. Задачей является нахождение по ним математического ожидания и дисперсии. Обычно задача решается построением весов для квантилей так, чтобы у полученного дискретного распределения были такие же математическое ожидание и дисперсия, как у исходного. Вообще говоря, иногда нужно, чтобы и более старшие моменты также аппроксимировались моментами построенного дискретного распределения с целью, чтобы для функций от распределений равенство математических ожиданий и дисперсий оставалось хотя бы приближенными.

В данной работе мы рассмотрим следующие вопросы:
Общий подход к трехточечной аппроксимации.
Трехточечная аппроксимация нормального распределения, в целом метод Свонсона и вывод правила 30-40-30.
Трехточечная аппроксимация логнормального распределения и её свойства.
Алгоритм аппроксимации произведения двух логнормальных распределений.
Алгоритм аппроксимации суммы двух логнормальных распределений.

Работа этого семестра заключена в разделах 3.1, 3.2, 3.5, 5.

В статье «Discretization, Simulation, and Swanson's (Inaccurate) Mean» [3] одна из частей исследования – сравнение различных методов дискретизации непрерывных распределений, например таких, как Extended Person-Tukey (EPT), McNamee-Celona Shortcut (MCS), Extended Swanson-Megill (ESM). Но нам это не подходит, потому что мы рассматриваем трехточечную симметричную аппроксимацию.

В статье «Discretization, Simulation, and the Value of Information» [4] замечено, что метод Свонсона значительно недооценивает среднее значение, дисперсию и асимметрию большинства распределений, особенно логнормального. Поэтому мы рассматриваем аппроксимацию конкретно для логнормального распределения.

2 Общий подход к трехточечной аппроксимации, аппроксимация нормального распределения

Пусть дана непрерывная случайная величина ξ с функцией распределения $\mathsf{F}(x)$. Обозначим

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

Для неё заданы квантили x_{π_1}, x_{π_2}, x_{π_3}. Также есть случайная дискретная величина $\tilde{\xi}$, которая задана следующим образом

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

для неё обозначим

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

Мы хотим аппроксимировать распределение случайной величины ξ дискретным распределением $\tilde{\xi}$ с сохранением первых двух моментов. Для этого нужно найти p_1, p_2, p_3 так, чтобы следующие равенства были верными.

$$p_1 + p_2 + p_3 = 1, \tag{1}$$

$$\tilde{m} = p_1 x_{\pi_1} + p_2 x_{\pi_2} + p_3 x_{\pi_3} = m, \tag{2}$$

$$\tilde{s^2} = p_1 x_{\pi_1}^2 + p_2 x_{\pi_2}^2 + p_3 x_{\pi_3}^2 - m^2 = s^2. \tag{3}$$

Запишем уравнения (1)–(3) в матричной форме следующим образом

$$\begin{pmatrix} 1 & 1 & 1 \\ x_{\pi_1} & x_{\pi_2} & x_{\pi_3} \\ x_{\pi_1}^2 & x_{\pi_2}^2 & x_{\pi_3}^2 \end{pmatrix} \begin{pmatrix} p_1 \\ p_2 \\ p_3 \end{pmatrix} = \begin{pmatrix} 1 \\ m \\ m^2 + s^2 \end{pmatrix}.$$

Теперь введём более изящную форму, которая подчёркивает связь вероятностей с формой распределения путём стандартизации.

Предложение 1 (Swanson, 2000 год). *Пусть верно*

$$\begin{pmatrix} 1 & 1 & 1 \\ \hat{x}_{\pi_1} & \hat{x}_{\pi_2} & \hat{x}_{\pi_3} \\ \hat{x}_{\pi_1}^2 & \hat{x}_{\pi_2}^2 & \hat{x}_{\pi_3}^2 \end{pmatrix} \begin{pmatrix} p_1 \\ p_2 \\ p_3 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}, \tag{4}$$

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* $s^2 = \tilde{s}^2$.

Замечание 1. *Предложение 1 дает требуемую аппроксимацию дискретным распределением, если найденные вероятности* p_i *являются неотрицательными.*

Аппроксимация нормального распределения. Если $\xi \sim N(\mu, \sigma)$ имеет нормальное распределение, то $\hat{\xi}$ имеет стандартное нормальное распределение, поэтому $\hat{\xi} \sim N(0,1)$ в Предложении 1.

Предложение 2 (Swanson, 2000 год). *Пусть* $\xi \sim N(\mu, \sigma)$, $\pi_3 = 1 - \pi_1$, $\pi_2 = 0.5$ *и пусть верно*

$$\begin{cases} p_1 = \dfrac{\delta}{2}, \\ p_2 = 1 - \delta, \\ p_3 = \dfrac{\delta}{2}, \end{cases} \tag{5}$$

где $\delta = \dfrac{1}{\Phi^{-1}(\pi_1)^2}$. *Тогда* $m = \tilde{m}$ *и* $s^2 = \tilde{s}^2$.

Доказательство.

Обозначим $\Phi(y) = \mathsf{P}\left(\eta = \dfrac{\xi - m}{s} \le y\right)$ — функция распределения стандартного нормального распределения, тогда система (4) записывается как

$$\begin{pmatrix} 1 & 1 & 1 \\ \Phi^{-1}(\pi_1) & \Phi^{-1}(\pi_2) & \Phi^{-1}(\pi_3) \\ \Phi^{-1}(\pi_1)^2 & \Phi^{-1}(\pi_2)^2 & \Phi^{-1}(\pi_3)^2 \end{pmatrix} \begin{pmatrix} p_1 \\ p_2 \\ p_3 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}. \tag{6}$$

В частном случае симметричных квантилей вида $\pi_1 = \pi$, $\pi_2 = 0.5$, $\pi_3 = 1 - \pi$ получаем $\Phi^{-1}(\pi) = -\Phi^{-1}(1 - \pi)$, $\Phi^{-1}(0.5) = 0$, тогда система (6) упрощается до

$$\begin{pmatrix} 1 & 1 & 1 \\ \Phi^{-1}(\pi) & 0 & -\Phi^{-1}(\pi) \\ \Phi^{-1}(\pi)^2 & 0 & \Phi^{-1}(\pi)^2 \end{pmatrix} \begin{pmatrix} p_1 \\ p_2 \\ p_3 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}.$$

Запишем следующим образом

$$\begin{cases} p_1 + p_2 + p_3 = 1, \\ (p_1 - p_3)\Phi^{-1}(\pi) = 0, \\ (p_1 + p_3)\Phi^{-1}(\pi)^2 = 1. \end{cases} \tag{7}$$

Обозначим $\delta = \dfrac{1}{\Phi^{-1}(\pi)^2}$, тогда из системы (7) получим утверждение Предложения 2. ∎

Рассмотрим случай $\pi = 0.1$, имеем $\Phi^{-1}(0.1) = -\Phi^{-1}(0.9) \approx -1.28$, $\Phi^{-1}(0.5) = 0$, из уравнений системы (7) находим значения p_1, p_2, p_3.

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

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

3 Аппроксимация логнормального распределения

3.1 Свойства логнормального распределения

Пусть случайная величина η имеет логнормальное распределение, тогда случайная величина $\xi = \ln(\eta)$ имеет нормальное распределение, $\xi \sim N(\mu, \sigma)$. И поэтому для нее можно использовать формулы, полученные в предыдущих разделах.

Параметры $m = \mathbf{E}(\eta)$, $s^2 = \mathbf{D}(\eta)$ логнормального распределения можно найти через параметры μ и σ^2 соответствующего нормального распределения по следующим формулам

$$m = \exp\left(\mu + \frac{\sigma^2}{2}\right), \tag{8}$$

$$s^2 = m^2(\exp(\sigma^2) - 1). \tag{9}$$

Заметим, что математическое ожидание логнормально распределенной случайной величины всегда положительное.

Коэффициент асимметрии выражается [3] следующей формулой

$$\gamma_3 = \sqrt{\exp(\sigma^2) - 1}(\exp(\sigma^2) + 2). \tag{10}$$

Коэффициент эксцесса находится [3] как

$$\gamma_4 = \exp(4\sigma^2) + 2\exp(3\sigma^2) + 3\exp(2\sigma^2) - 6. \tag{11}$$

Обратная функция распределения имеет вид

$$F_\eta^{-1}(p) = \exp(\mu + \sigma\sqrt{2}\mathrm{erf}^{-1}(2p - 1)). \tag{12}$$

3.2 Связь параметров с квантилями

Предложение 3. *Параметр σ выражается через любые два квантиля как*

$$\sigma = \frac{\log\left(\dfrac{x_{\pi_2}}{x_{\pi_1}}\right)}{\Phi^{-1}(\pi_2) - \Phi^{-1}(\pi_1)}, \qquad \pi_1 \neq \pi_2. \tag{13}$$

Доказательство. Покажем, что дисперсию логнормального распределения можно вычислить из отношения двух квантилей. Распишем вероятность

$$\mathsf{P}(\xi \leq x_\pi) = \pi,$$

$$\mathsf{P}\left(\frac{\log(\xi) - \mu}{\sigma} \leq \frac{\log(x_\pi) - \mu}{\sigma}\right) = \pi.$$

Следовательно,

$$\Phi\left(\frac{\log(x_\pi) - \mu}{\sigma}\right) = \pi,$$

и тогда

$$\log(x_\pi) = \mu + \sigma\Phi^{-1}(\pi).$$

С помощью двух квантилей мы можем исключить μ из соответствующих уравнений. Запишем

$$\log\left(\frac{x_{\pi_3}}{x_{\pi_1}}\right) = \sigma(\Phi^{-1}(\pi_3) - \Phi^{-1}(\pi_1)).$$

И в итоге получаем

$$\sigma = \frac{\log\left(\frac{x_{\pi_2}}{x_{\pi_1}}\right)}{\Phi^{-1}(\pi_2) - \Phi^{-1}(\pi_1)}.$$

■

Параметр μ выражается как

$$\mu = \log(x_{\pi_i}) - \sigma\Phi^{-1}(\pi_i) \tag{14}$$

и результат не зависит от i.

Предложение 4. *В терминах Предложения 1 функция* $\hat{\mathsf{F}}^{-1}(\pi)$ *выражается через* σ *как*

$$\hat{\mathsf{F}}^{-1}(\pi) = y = \frac{\exp\left(\sigma\Phi^{-1}(\pi) - \frac{\sigma^2}{2}\right) - 1}{\sqrt{\exp(\sigma^2) - 1}}. \tag{15}$$

Доказательство. Выразим $\hat{\mathsf{F}}(y)$ через функцию стандартного нормального распределения

$$\hat{\mathsf{F}}(y) = \Phi\left(\frac{\log(m + sy) - \mu}{\sigma}\right),$$

так как $\xi = \ln(\eta) \sim N(\mu, \sigma)$. Выразим $\log(m + sy)$ через μ и σ, используя формулы (8) и (9). Получаем

$$m + sy = e^{\mu+\frac{\sigma^2}{2}} + ye^{\mu+\frac{\sigma^2}{2}}\sqrt{e^{\sigma^2} - 1} = e^{\mu+\frac{\sigma^2}{2}}(1 + y\sqrt{\exp(\sigma^2) - 1}),$$

возьмем натуральный логарифм от обеих частей, получаем

$$\begin{aligned}\log(m + sy) &= \log(e^{\mu+\frac{\sigma^2}{2}}(1 + y\sqrt{\exp(\sigma^2) - 1})) = \\ &= \mu + \frac{\sigma^2}{2} + \log(1 + y\sqrt{\exp(\sigma^2) - 1}),\end{aligned}$$

тогда

$$\frac{\log(m + sy) - \mu}{\sigma} = \frac{\sigma}{2} + \frac{\log(1 + y\sqrt{\exp(\sigma^2) - 1})}{\sigma}.$$

То есть можно выразить

$$\hat{\mathsf{F}}(y) = \Phi\left(\frac{\log(m + sy) - \mu}{\sigma}\right) = \Phi\left(\frac{\sigma}{2} + \frac{\log(1 + y\sqrt{\exp(\sigma^2) - 1})}{\sigma}\right).$$

Далее находим $\Phi^{-1}(\pi)$. Получаем

$$\Phi\left(\frac{\sigma}{2}+\frac{\log(1+y\sqrt{\exp(\sigma^2)-1})}{\sigma}\right)=\pi,$$

$$\Phi^{-1}(\pi)=\frac{\sigma}{2}+\frac{\log(1+y\sqrt{\exp(\sigma^2)-1})}{\sigma}.$$

Теперь можно выразить $\log(1+y\sqrt{\exp(\sigma^2)-1})$ как

$$\log(1+y\sqrt{\exp(\sigma^2)-1})=\sigma\Phi^{-1}(\pi)-\frac{\sigma^2}{2},$$

$$1+y\sqrt{\exp(\sigma^2)-1}=\exp\left(\sigma\Phi^{-1}(\pi)-\frac{\sigma^2}{2}\right).$$

В итоге получаем

$$\hat{\mathsf{F}}^{-1}(\pi)=y=\frac{\exp(\sigma\Phi^{-1}(\pi)-\frac{\sigma^2}{2})-1}{\sqrt{\exp(\sigma^2)-1}}.$$

■

3.3 Варианты постановки задачи

Задача: имеются квантили x_π, $x_{0.5}$, $x_{1-\pi}$ логнормальной случайной величины η. Нужно уметь считать её математическое ожидание и дисперсию.

Варианты решения задачи:

1. Не переходить к аппроксимации дискретной случайной величиной, а сразу же из двух уравнений вида (14), записанных для двух квантилей, найти значения параметров μ и σ нормальной случайной величины $\ln(\eta)\sim N(\mu,\sigma)$. Далее по формулам (8) и (9) вычислить значения мат. ожидания m и дисперсии s^2 случайной величины η.

2. Перейти к трехточечной аппроксимации дискретной случайной величиной $\tilde{\xi}$, у которой $\tilde{m}=m$, $\tilde{s}^2=s^2$ и считать значения m и s через квантили x_π, $x_{0.5}$, $x_{1-\pi}$ и вероятности p_1, p_2, p_3. Если условие для положительных вероятностей не выполняется, можно воспринимать задачу не как поиск вероятностей для $\tilde{\xi}$, а как поиск весов для линейной комбинации x_π, $x_{0.5}$, $x_{1-\pi}$ таких, чтобы параметры, полученные по формулам (2) и (3), были равны мат. ожиданию и дисперсии η.

В реальных задачах в нефтяной промышленности используются следующие диапазоны параметров:

$$\mu\le 12,\qquad \sigma\le 1.5.$$

Поэтому мы будем обращать на них особое внимание.

3.4 Способ нахождения весов для x_π, $x_{0.5}$, $x_{1-\pi}$ через математическое ожидание и дисперсию нормального распределения и непосредственная аппроксимация логнормального распределения

Заметим, что если $x_{\pi_1}, x_{\pi_2}, x_{\pi_3}$ — квантили логнормального распределения, то $\ln(x_{\pi_1})$, $\ln(x_{\pi_2})$, $\ln(x_{\pi_3})$ — квантили нормального распределения соответствующие тем же вероятностям. Можно взять эти квантили и использовать в способе нахождения вероятностей для нормального распределения, пользуясь Предложением 2.

Имеем следующий алгоритм.

Алгоритм 1. Дано: *квантили $x_{\pi_1}, x_{\pi_2}, x_{\pi_3}$ логнормальной случайной величины η, $\ln(\eta) \sim N(\mu, \sigma)$.*

Шаги:

1. *Выражаем параметры μ и σ мат. ожидание и дисперсию соответствующего нормального распределения через известные $x_{\pi_1}, x_{\pi_2}, x_{\pi_3}$ по формулам* (13) *и* (14)*.*
2. *Вычисляем значения мат. ожидания m и дисперсии s^2 случайной величины η, используя μ и σ по формулам* (8) *и* (9)*.*
3. *С помощью* (1)–(3) *находим значения весов p_1, p_2, p_3, используя вычисленные m и s^2.*

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

Есть другой способ нахождения этого результата. Можно не переходить к нормальному распределению, а сразу вычислять вероятности для квантилей логнормального распределения.

Алгоритм 2. Дано: *квантили $x_{\pi_1}, x_{\pi_2}, x_{\pi_3}$ логнормальной случайной величины η, $\ln(\eta) \sim N(\mu, \sigma)$.*

Шаги:

1. *Выражаем параметр σ из отношения x_{π_3} к x_{π_1}, используя формулу* (13)*.*
2. *Вычисляем значения $\hat{\mathsf{F}}^{-1}(\pi)$ для случайной величины η по формуле* (15)*.*
3. *С помощью системы* (4) *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 $\tilde{\xi}$.*

Замечание 2. *Результаты Алгоритмов 1 и 2 совпадают, так как веса для аппроксимации единственны.*

3.5 Условие на параметр σ для существования трехточечной аппроксимации логнормального распределения

Мы рассмотрели способы 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. Докажем следующее предложение.

Предложение 5. *Неотрицательные вероятности p_1, p_2, p_3 для аппроксимации логнормальной случайной величины η с квантилями вида x_π, $x_{0.5}$, $x_{1-\pi}$ существуют только при условии*

$$\exp(\sigma^2) + \exp(-\sigma^2) - \exp\left(-\frac{\sigma^2}{2}\right)(\exp(c\sigma) + \exp(-c\sigma)) \leq 0, \tag{16}$$

где $c = \Phi^{-1}(\pi)$.

Доказательство. Рассматриваем $\ln(\eta) \sim N(\mu, \sigma^2)$ и случай симметричных квантилей $\pi_1 = \pi$, $\pi_2 = 0.5$, $\pi_3 = 1 - \pi$.

С помощью формулы (15) найдем $\tilde{\mathsf{F}}^{-1}(\pi_i)$, делаем следующие обозначения

$$\tilde{\mathsf{F}}^{-1}(\pi) = t_1, \qquad \tilde{\mathsf{F}}^{-1}(0.5) = t_2, \qquad \tilde{\mathsf{F}}^{-1}(1-\pi) = t_3.$$

Теперь рассмотрим систему (4), запишем ее через t_1, t_2, t_3 и выразим вероятности p_1, p_2, p_3. Имеем

$$p_2(t_2 - t_3) = p_1(t_3 - t_1) - t_3,$$
$$p_1(t_1^2 - t_3^2) + p_2(t_2^2 - t_3^2) = 1 - t_3^2.$$

Тогда получаем

$$p_1(t_1^2 - t_3^2) + (t_2 + t_3)(p_1(t_3 - t_1) - t_3) = 1 - t_3^2,$$
$$p_1(t_1 - t_3)(t_1 - t_2) = 1 + t_2 t_3.$$

В итоге вероятности записываются следующим образом

$$p_1 = \frac{1 + t_2 t_3}{(t_1 - t_3)(t_1 - t_2)}, \tag{17}$$

$$p_2 = \frac{p_1(t_3 - t_1) - t_3}{t_2 - t_3} = \frac{1 + t_1 t_3}{(t_2 - t_1)(t_2 - t_3)}, \tag{18}$$

$$p_3 = 1 - p_1 - p_2. \tag{19}$$

Все вероятности должны быть положительными, подставим в формулы для вероятностей значения переменных t_1, t_2, t_3, где $\tilde{\mathsf{F}}^{-1}(\pi_i)$ ищутся по формуле (15). Вероятность p_1 выражается как

$$p_1 = \frac{1 + \dfrac{\left(\exp\left(-\dfrac{\sigma^2}{2}\right) - 1\right)\left(\exp\left(-c\sigma - \dfrac{\sigma^2}{2}\right) - 1\right)}{\exp(\sigma^2) - 1}}{\dfrac{\exp\left(c\sigma - \dfrac{\sigma^2}{2}\right) - \exp\left(-c\sigma - \dfrac{\sigma^2}{2}\right)}{\sqrt{\exp(\sigma^2) - 1}} \dfrac{\exp\left(c\sigma - \dfrac{\sigma^2}{2}\right) - \exp\left(-\dfrac{\sigma^2}{2}\right)}{\sqrt{\exp(\sigma^2) - 1}}} =$$

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

Вероятность p_2 выражается как

$$p_2 = \frac{1 + \frac{\left(\exp\left(\mathrm{c}\sigma - \frac{\sigma^2}{2}\right) - 1\right)\left(\exp\left(-\mathrm{c}\sigma - \frac{\sigma^2}{2}\right) - 1\right)}{\exp(\sigma^2) - 1}}{\frac{\exp\left(-\frac{\sigma^2}{2}\right) - \exp\left(\mathrm{c}\sigma - \frac{\sigma^2}{2}\right)}{\sqrt{\exp(\sigma^2) - 1}} \frac{\exp\left(-\frac{\sigma^2}{2}\right) - \exp\left(-\mathrm{c}\sigma - \frac{\sigma^2}{2}\right)}{\sqrt{\exp(\sigma^2) - 1}}} =$$

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

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

Докажем, что вероятности p_1 и p_3 положительные при любом параметре σ. Сначала распишем знаменатель p_1.

$$\exp(2c\sigma - \sigma^2) - \exp(c\sigma - \sigma^2) - \exp(-\sigma^2) + \exp(-c\sigma - \sigma^2) =$$
$$= \exp(-\sigma^2)(\exp(2c\sigma) - \exp(c\sigma) - 1 + \exp(-c\sigma)) =$$
$$= \exp(-\sigma^2)(\exp(2c\sigma) - 2\exp(c\sigma) + 1 + \exp(c\sigma) - 2 + \exp(-c\sigma)) =$$
$$= \exp(-\sigma^2)\left((\exp(\mathrm{c}\sigma) - 1)^2 + \frac{(\exp(\mathrm{c}\sigma) - 1)^2}{\exp(\mathrm{c}\sigma)}\right) \geq 0.$$

Теперь анализируем числитель p_1, так как

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

то числитель p_1 тоже всегда неотрицательный. Рассмотрим знаменатель p_2. Имеем:

$$2\exp(-\sigma^2) - \exp(-\sigma^2)(\exp(-\mathrm{c}\sigma) + \exp(\mathrm{c}\sigma)) =$$
$$= \exp(-\sigma^2)(2 - \exp(-\mathrm{c}\sigma) - \exp(\mathrm{c}\sigma)) =$$
$$= -\frac{\exp(-\sigma^2)(\exp(\mathrm{c}\sigma) - 1)^2}{\exp(\mathrm{c}\sigma)} \leq 0.$$

Вероятность p_3 выражается как

$$p_3 = 1 - \frac{1 + t_2 t_3}{(t_1 - t_3)(t_1 - t_2)} - \frac{1 + t_1 t_3}{(t_2 - t_1)(t_2 - t_3)}.$$

Выяснили, что $(t_1-t_3)(t_1-t_2)\geq 0$, а $(t_2-t_1)(t_2-t_3)\leq 0$, значит, знаменатель p_3 всегда отрицательный. Осталось показать, что числитель p_3 тоже всегда отрицательный. Для этого распишем его как

$$(t_1-t_3)(t_1-t_2)(t_2-t_1)(t_2-t_3)-(1+t_2t_3)(t_2-t_1)(t_2-t_3)-(1+t_1t_3)(t_1-t_3)(t_1-t_2)=$$

$$=(t_1^2-t_1t_2-t_1t_3+t_2t_3)(t_2^2-t_2t_3-t_1t_2+t_1t_3)-$$
$$-(1+t_2t_3)(t_2^2-t_2t_3-t_1t_2+t_1t_3)-(1+t_1t_3)(t_1^2-t_1t_2-t_1t_3+t_2t_3)=$$
$$=2t_1^2t_2^2-t_1^2t_2t_3-t_1^3t_2+t_1^3t_3-t_1t_2^3+t_1t_2t_3^2-t_1^2t_3^2+t_2^3t_3-t_2^2t_3^2-t_1t_2^2t_3+t_1t_2t_3^3-$$
$$-t_2^2+2t_1t_2-t_1^2-t_2^3t_3+t_2^2t_3^2+t_1t_2^2t_3-2t_1t_2t_3^2-t_1^3t_3+t_1^2t_2t_3=$$
$$=(2t_1^2t_2^2-t_1^3t_2-t_1t_2^3)-(t_2^2-2t_1t_2+t_1^2)=$$
$$=-(t_1^{\frac{1}{2}}t_2^{\frac{3}{2}}-t_1^{\frac{3}{2}}t_2^{\frac{1}{2}})^2-(t_1-t_2)^2\leq 0.$$

Из условия отрицательности числителя p_2 получаем ограничение (16). ∎

Предложение 6. *Пусть* $\Gamma(\pi)=\{\sigma:(16)\quad \text{выполнено}\}$. *Тогда из* $\pi\leq\tilde{\pi}$ *следует, что* $\Gamma(\pi)\supset\Gamma(\tilde{\pi})$.

Доказательство. Имеем неравенство (16), рассмотрим $\exp(\mathrm{c}\sigma)+\exp(-\mathrm{c}\sigma)$. Эта сумма увеличивается при уменьшении π и фиксированной σ, так как увеличивается по модулю значение $\mathrm{c}=\Phi^{-1}(\pi)$. Значит, вычитаемое

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

становится все больше, и неравенство (16) выполняется для большего множества значений σ.

∎

Например, для $\pi=0.1$ получаем ограничение $\sigma\leq 0.6913$, $\sigma^2\leq 0.4779$. Посмотрим, какому коэффициенту асимметрии соответствует это значение σ. По формуле (10) находим $\gamma_3=2.82778$.

Рассмотрим $\pi=0.05$, получаем ограничение $\sigma\leq 1.04585$ и значение коэффициента асимметрии $\gamma_3=7.02529$.

Вычислим, при каком значении π получается ограничение $\sigma\leq 1.5$, имеем

$$\exp(1.5^2)+\exp(-1.5^2)-\exp\left(-\frac{1.5^2}{2}\right)(\exp(1.5c)+\exp(-1.5c))=0,$$

$$9.4877+0.1054-0.3247(\exp(1.5c)+\exp(-1.5c))=0,$$
$$(\exp(1.5c)+\exp(-1.5c))=29.5491,$$
$$(\exp(3c)-29.5491\exp(1.5c))+1=0,$$
$$\exp(1.5c)=0.0678,\qquad c=\frac{\ln(0.0678)}{1.5}=-1.794522,$$
$$\pi=\Phi(-1.794522)\approx 3.636\%.$$

3.6 Точность неправильной аппроксимации на основе дискретной аппроксимации нормального распределения

Предлагаемые методы аппроксимации трехточечным дискретным распределением логнормального распределения не работают при $\sigma \leq 0.6913$. На практике часто используют правило 30-40-30 выведенное для аппроксимации нормального распределения, значения весов вычисляются с помощью системы (5). Посмотрим на точность правила 30-40-30, особенно это важно при $\sigma \geq 0.6913$.

Предложение 7. *Пусть $\pi_1 = \pi$, $\pi_2 = 0.5$, $\pi_3 = 1 - \pi$ и значения вероятностей аппроксимации равны $p_1 = \delta/2$, $p_2 = 1 - \delta$, $p_3 = \delta/2$, тогда*

1. *Ошибка аппроксимации мат. ожидания равна*

$$\frac{\mid m - \widetilde{m} \mid}{m} = \frac{\left|\exp\left(\dfrac{\sigma^2}{2}\right) - \dfrac{1}{2c^2}\bigg(\exp(\sigma c) - 1 + \exp(-\sigma c)\bigg) + 1\right|}{\exp\left(\dfrac{\sigma^2}{2}\right)},$$

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

2. *Ошибка аппроксимации дисперсии 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.

Доказательство. 1. Выразим ошибку аппроксимации мат. ожидания логнормального распределения через параметры μ и σ, используя формулы (8) и (9). Значения вероятностей p_1, p_2, p_3 находятся из системы (5) Предложения 2. Тогда математическое ожидание аппроксимации равно

$$\tilde{m} = \frac{1}{2\mathrm{c}^2}\exp(\mu + \sigma\mathrm{c}) + \left(1 - \frac{1}{\mathrm{c}^2}\right)\exp(\mu) + \frac{1}{2\mathrm{c}^2}\exp(\mu - \sigma\mathrm{c}) =$$

$$= \frac{1}{2\mathrm{c}^2}\exp(\mu)(\exp(\sigma\mathrm{c}) - 1 + \exp(-\sigma\mathrm{c})) + \exp(\mu).$$

Получили ошибку

$$\frac{\mid m - \widetilde{m} \mid}{m} =$$

$$= \frac{\left|\exp\left(\mu + \dfrac{\sigma^2}{2}\right) - \dfrac{1}{2\mathrm{c}^2}\exp(\mu)(\exp(\sigma\mathrm{c}) - 1 + \exp(-\sigma\mathrm{c})) + \exp(\mu)\right|}{\exp\left(\mu + \dfrac{\sigma^2}{2}\right)} =$$

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

2. Выразим аппроксимации дисперсии через параметры распределения.

$$s^2 = \exp(2\mu + \sigma^2)(\exp(\sigma^2 - 1)),$$

$$\tilde{s}^2 = \frac{1}{2\text{c}^2}\exp(2\mu + 2\sigma\text{c}) + \left(1 - \frac{1}{\text{c}^2}\right)\exp(2\mu) + \frac{1}{2\text{c}^2}\exp(2\mu - 2\sigma\text{c}) - \tilde{m}^2.$$

Получили ошибку

$$\frac{|\ s^2 - \widetilde{s}^2\ |}{s^2} = \left|\exp(2\mu + \sigma^2)(\exp(\sigma^2 - 1)) - \frac{1}{2c^2}\exp(2\mu + 2\sigma c) - \left(1 - \frac{1}{c^2}\right)\exp(2\mu) - \right.$$

$$\left. - \frac{1}{2c^2}\exp(2\mu - 2\sigma c) + \tilde{m}^2\right| / \exp(2\mu + \sigma^2)(\exp(\sigma^2 - 1)) =$$

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

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

■

Замечание 3. *При уменьшении значения π ошибки аппроксимации мат. ожидания и дисперсии становятся меньше.*

Замечание 4. *В предложении 7 можно убрать знак модуля и оно останется верным.*

Построим график зависимости от σ. Видим, что при $\sigma \leq 1.5$, взятых из нашего диапазона, ошибка аппроксимации мат.ожидания меньше 12%, а ошибка аппроксимации дисперсии может достигать 80%. Для $\sigma \geq 0.69$, когда условие (16) не выполнено, ошибка мат. ожидания может быть как маленькой, так и очень большой. Ошибка дисперсии при этом точно больше 25%.

4 Произведение двух логнормальных распределений

Рассмотрим произведение логнормально распределенных случайных величин. Эта процедура применяется в нефтяной промышленности, например, используем площадь дренирования пласта, среднюю чистую толщину и коэффициент извлечения углеводородов. При перемножении этих параметров получаем количество резервов нефти. Идеи доказательств Предложений этого раздела взяты из статьи «Uncertainties impacting reserves, revenue, and costs» [2].

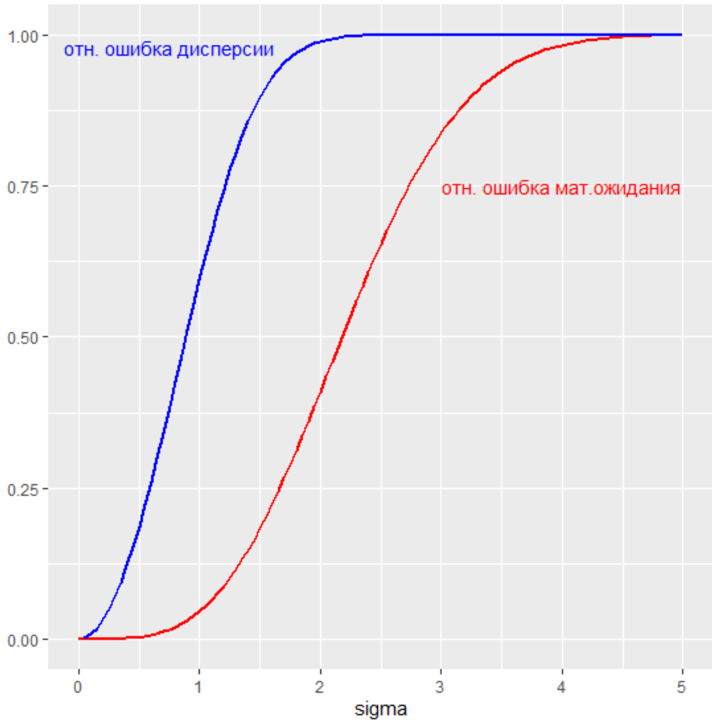

Рис. 1: Ошибка аппроксимации мат. ожидания и дисперсии

Мы рассмотрим произведение двух логнормально распределенных случайных величин

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

Введем следующие обозначения:
x_π, $x_{0.5}$, $x_{1-\pi}$ — квантили случайной величины ξ_1,
y_π, $y_{0.5}$, $y_{1-\pi}$ — квантили случайной величины ξ_2.

4.1 Произведение квантилей

Предложение 8. *Величина $x_\pi y_\pi$ является q-квантилью случайной величины $\xi_1\xi_2$, где*

$$q = \mathsf{P}(\xi_1\xi_2 < x_\pi y_\pi) = \Phi\left(\frac{\Phi^{-1}(\pi)(\ln(x_{0.5}) + \ln(y_{0.5}) - \ln(x_\pi) - \ln(y_\pi))}{\sqrt{(\ln(x_{0.5}) - \ln(x_\pi))^2 + (\ln(y_{0.5}) - \ln(y_\pi))^2}}\right). \tag{20}$$

Доказательство. Выразим параметры распределений μ_1, μ_2, σ_1, σ_2 через квантили, используя формулу (14). Теперь рассмотрим случайную величину $\eta = \xi_1\xi_2$. Мы хотим вычислить, каким квантилем для η является произведение квантилей x_π и y_π. Для этого надо найти, чему равна вероятность $\mathsf{P}(\xi_1\xi_2 < x_\pi y_\pi)$. Получаем

$$\mathsf{P}(\xi_1\xi_2 < x_\pi y_\pi) = \mathsf{P}(\ln(\xi_1) + \ln(\xi_2) < \ln(x_\pi) + \ln(y_\pi)) =$$

$$= \mathsf{P}\left(\frac{\ln(\xi_1)+\ln(\xi_2)-(\mu_1+\mu_2)}{\sqrt{\sigma_1^2+\sigma_2^2}} < \frac{\ln(x_\pi)+\ln(y_\pi)-(\mu_1+\mu_2)}{\sqrt{\sigma_1^2+\sigma_2^2}}\right).$$

Так как ξ_1 распределена логнормально с параметрами μ_1 и σ_1^2, а ξ_2 распределена логнормально с параметрами μ_2 и σ_2^2, то

$$\ln(\xi_1)+\ln(\xi_2) \sim N(\mu_1+\mu_2, \sigma_1^2+\sigma_2^2),$$
$$\frac{\ln(\xi_1)+\ln(\xi_2)-(\mu_1+\mu_2)}{\sqrt{\sigma_1^2+\sigma_2^2}} \sim N(0,1).$$

Тогда можно записать

$$\mathsf{P}(\xi_1\xi_2 < x_\pi y_\pi) =$$
$$= \mathsf{P}\left(\frac{\ln(\xi_1)+\ln(\xi_2)-(\mu_1+\mu_2)}{\sqrt{\sigma_1^2+\sigma_2^2}} < \frac{(\mu_1+\Phi^{-1}(\pi)\sigma_1)+(\mu_2+\Phi^{-1}(\pi)\sigma_2)-(\mu_1+\mu_2)}{\sqrt{\sigma_1^2+\sigma_2^2}}\right) =$$

$$= \mathsf{P}\left(\frac{\ln(\xi_1)+\ln(\xi_2)-(\mu_1+\mu_2)}{\sqrt{\sigma_1^2+\sigma_2^2}} < \frac{\Phi^{-1}(\pi)(\sigma_1+\sigma_2)}{\sqrt{\sigma_1^2+\sigma_2^2}}\right) =$$
$$= \Phi\left(\frac{\Phi^{-1}(\pi)(\sigma_1+\sigma_2)}{\sqrt{\sigma_1^2+\sigma_2^2}}\right).$$

Перепишем эту дробь через значения квантилей, получаем

$$\frac{\Phi^{-1}(\pi)(\sigma_1+\sigma_2)}{\sqrt{\sigma_1^2+\sigma_2^2}} = \frac{\Phi^{-1}(\pi)\left(\dfrac{(\ln(x_{0.5})-\ln(x_\pi))+(\ln(y_{0.5})-\ln(y_\pi))}{-\Phi^{-1}(\pi)}\right)}{\sqrt{\dfrac{(\ln(x_{0.5})-\ln(x_\pi))^2+(\ln(y_{0.5})-\ln(y_\pi))^2}{(\Phi^{-1}(\pi))^2}}} =$$

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

Тогда получаем следующую формулу

$$\mathsf{P}(\xi_1\xi_2 < x_\pi y_\pi) = \Phi\left(\frac{\Phi^{-1}(\pi)(\ln(x_{0.5})+\ln(y_{0.5})-\ln(x_\pi)-\ln(y_\pi))}{\sqrt{(\ln(x_{0.5})-\ln(x_\pi))^2+(\ln(y_{0.5})-\ln(y_\pi))^2}}\right).$$

■

Следствие 1. *При перемножение квантилей $x_{0.5}$ и $y_{0.5}$ получается снова 0.5-ый квантиль.*

Доказательство. Запишем вероятность $\mathsf{P}(\xi_1\xi_2 < x_{0.5}y_{0.5})$ следующим образом:

$$\mathsf{P}(\xi_1\xi_2 < x_{0.5}y_{0.5}) = \Phi\left(\frac{\ln(x_{0.5})+\ln(y_{0.5})-(\mu_1+\mu_2)}{\sqrt{\sigma_1^2+\sigma_2^2}}\right).$$

Но в числителе получается 0, значит,

$$\mathsf{P}(\xi_1\xi_2 < x_{0.5}y_{0.5}) = \Phi(0) = 0.5.$$

■

4.2 q-Квантили произведения логнормальных случайных величин для $q = \pi$, $q = 0.5$, $q = 1 - \pi$

Как по каким-то произвольным получившимся квантилям, полученным при перемножении данных квантилей для двух логнормальных случайных величин, найти нужные нам, такие же, как исходные π, 0.5, $1-\pi$ квантили произведения этих двух случайных величин? Сначала нужно понять, на какой прямой лежат точки вида $(x_\pi; \Phi^{-1}(\pi))$.

Нужно выяснить, как связаны параметры нормального распределения, квантили которого откладываются по оси X, и параметры прямой, на которой лежат точки QQ плота.

Предложение 9. *Точки QQ-плота: $\left\{x_i, \mathsf{F}_\eta^{-1}(\mathsf{F}_\xi(x_i))\right\}_{i=1}^n$, где ось X: $\xi \sim N(a, b^2)$, ось Y: $\eta \sim N(0,1)$ лежат на прямой $y = \dfrac{x-a}{b}$.*

Доказательство. Возьмем две точки и построим по ним уравнение прямой. Например, точки

$$(\mathsf{F}_\xi^{-1}(0.1), \mathsf{F}_\eta^{-1}(0.1)),$$
$$(\mathsf{F}_\xi^{-1}(0.5), \mathsf{F}_\eta^{-1}(0.5)).$$

Имеем

$$\Phi\left(\frac{x_p - a}{b}\right) = p \qquad \Rightarrow \qquad \frac{x_p - a}{b} = \Phi^{-1}(p).$$

Получаем, что

$$x_p = a + b\Phi^{-1}(p).$$

Для первой точки возьмем $p = 0.1$, тогда

$$(a + b\Phi^{-1}(0.1); \Phi^{-1}(0.1)).$$

Для второй точки возьмем $p = 0.5$, тогда

$$(a + b\Phi^{-1}(0.5); \Phi^{-1}(0.5)) \qquad \Rightarrow \qquad (a; 0).$$

Составим уравнение прямой:

$$\frac{x-a}{(a+\Phi^{-1}(0.1)b) - a} = \frac{y}{\Phi^{-1}(0.1)}, \qquad \frac{x-a}{\Phi^{-1}(0.1)b} = \frac{y}{\Phi^{-1}(0.1)}.$$

Следовательно,

$$by = x - a,$$

Получили уравнение прямой на которой лежат точки данного QQ-плота:

$$y = \frac{x-a}{b}.$$

■

Предложение 10 (Swanson, 2000 год). *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

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

Доказательство. С помощью формулы (19) можно посчитать, какой получается квантиль для случайной величины $\xi_1\xi_2$, если перемножить квантили x_π и y_π исходных случайных величин. Обозначим z_π, $z_{0.5}$, $z_{1-\pi}$ — квантили случайной величины η. Тогда по Следствию 1 имеем $x_{0.5}y_{0.5} = z_{0.5}$.

Нужно вычислить значения z_π и $z_{1-\pi}$. Введем обозначение:

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

Тогда по Предложению 9 с помощью точек $(\ln(x_\pi y_\pi), t)$ и $(\ln(x_{0.5}y_{0.5}), 0)$ можно найти параметры a и b прямой, на которой они лежат.

$$\frac{\ln(x_{0.5}y_{0.5}) - a}{b} = 0 \qquad \Rightarrow \qquad a = \ln(x_{0.5}y_{0.5}),$$

$$\frac{\ln(x_\pi y_\pi) - a}{b} = t,$$

$$b = \frac{\ln(x_\pi y_\pi) - a}{t} = \frac{\ln(x_\pi y_\pi) - \ln(x_{0.5}y_{0.5})}{t}.$$

Так как точки $(\ln(z_\pi), \Phi^{-1}(\pi))$ и $(\ln(z_{1-\pi}), \Phi^{-1}(1-\pi))$ тоже лежат на этой прямой, то мы можем вычислить значения $\ln(z_\pi)$ и $\ln(z_{0.5})$, зная уравнение прямой, следующим образом:

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

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

И, наконец, находим z_π и $z_{1-\pi}$.

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

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

$\blacksquare$

По Алгоритму 1, используя найденные z_π, $z_{0.5}$, $z_{1-\pi}$, можно вычислить значения весов p_1, p_2, p_3 дискретной аппроксимации.

5 Сумма двух логнормальных распределений

Рассмотрим сумму двух логнормальных случайных величин.

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

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

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

Дано: 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.

Поставим задачу аппроксимации суммы логнормальным распределением $\ln(\eta) \sim N(\mu, \sigma^2)$, так как нужно рассматривать сумму не обязательно двух, а произвольного числа случайных величин.

Нужно найти квантили z_π, $z_{0.5}$, $z_{1-\pi}$ случайной величины η. По известным квантилям уже знаем, как вычислять вероятности p_1, p_2, p_3 такие, что $m = \tilde{m}$ и $s^2 = \tilde{s}^2$.

У нас есть следующие ограничения на параметры: $\mu_1, \mu_2 < 12$, $\sigma_1, \sigma_2 < 1.5$. Пусть мы нашли аппроксимацию суммы двух логнормальных величин, тогда с учетом этих ограничений её значения μ и σ тоже будут иметь свои ограничения. При этом, чтобы найти значения вероятностей p_1, p_2, p_3 нужно, чтобы выполнялось то же условие, что в разделе 4.3. А именно, $\sigma < 0.6913$.

Имеем следующий алгоритм для решения задачи.

Алгоритм 3. 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

   По набору квантилей ξ_1 *находим параметры* μ_1, σ_1 *нормального распределения по формулам* (13) *и* (14).

2. y_π, $y_{0.5}$, $y_{1-\pi} \to \mu_2$, σ_2

   По набору квантилей ξ_2 *находим параметры* μ_2, σ_2 *нормального распределения по формулам* (13) *и* (14).

3. $\mu_i,\ \sigma_i \to m_i,\ s_i^2$

   С помощью формул (8) *и* (9) *находим мат. ожидания и дисперсии* ξ_1 *и* ξ_2.

4. $m = m_1 + m_2$

   Вычисляем мат.ожидание $\xi_1 + \xi_2$.

5. $s^2 = s_1^2 + s_2^2$

   Вычисляем дисперсию $\xi_1 + \xi_2$.

6. $m,\ s^2 \to \mu,\ \sigma$

   С помощью формул (8) *и* (9) *находим параметры нормального распределения.*

7. $\mu,\ \sigma \to z_\pi,\ z_{0.5},\ z_{1-\pi}$

   С помощью формулы (14) *находим значения квантилей через* μ *и* σ.

8. $z_\pi,\ z_{0.5},\ z_{1-\pi} \to p_1,\ p_2,\ p_3$

   По Алгоритму 1 находим значения вероятностей $p_1,\ p_2,\ p_3$.

Результат: *вероятности* $p_1,\ p_2,\ p_3$ *для квантилей* $z_{\pi_1}, z_{\pi_2}, z_{\pi_3}$ *случайной величины* η, *которая является дискретной аппроксимацией аппроксимации суммы логнормальным распределением.*

Точность аппроксимации. Выразим ошибки аппроксимации квантилей $q_\pi, q_{0.5}, q_{1-\pi}$ случайной величины ξ через параметры $\mu_1,\ \mu_2,\ \sigma_1^2,\ \sigma_2^2$.

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

$$z_\pi = F_\eta^{-1}(\pi), \qquad z_{0.5} = \exp(\mu), \qquad z_{1-\pi} = F_\eta^{-1}(1-\pi), \quad \text{где}$$
$$F_\eta^{-1}(p) = \exp(\mu + \sigma\sqrt{2}\mathrm{erf}^{-1}(2p-1)).$$

Параметры μ, σ можно найти через параметры случайных величин ξ_1, ξ_2, используя формулы (8), (9) и вычисленные значения

$$m = \exp\left(\mu_1 + \frac{\sigma_1^2}{2}\right) + \exp\left(\mu_2 + \frac{\sigma_2^2}{2}\right),$$

$$s^2 = m_1^2(\exp(\sigma_1^2) - 1) + m_2^2(\exp(\sigma_2^2) - 1).$$

Квантили η выражаются как

$$q_\pi = F_\xi^{-1}(\pi), \qquad q_{0.5} = F_\xi^{-1}(0.5), \qquad q_{1-\pi} = F_\xi^{-1}(1-\pi), \quad \text{где}$$

$$F_\xi(x) = \int_0^x \left(\frac{1}{2} + \frac{1}{2}\mathrm{erf}\left(\frac{\ln(x-y) - \mu_1}{\sigma_1\sqrt{2}}\right)\right)\left(\frac{1}{\sqrt{2\pi}y\sigma_2}\exp\left(-\left(\frac{\ln(y) - \mu_2}{\sqrt{2}\sigma_2}\right)^2\right)\right)dy.$$

Здесь $F_\xi(x)$ — функция распределения $\xi = \xi_1 + \xi_2$, найденная с помощью формулы свертки.

В таблицах 1, 2 и 3 представлены ошибки для $\ln(\xi_1) \sim N(4, \sigma_1^2)$, $\ln(\xi_2) \sim N(4, \sigma_2^2)$ при $\pi = 0.1$, полученные с помощью моделирования, объемы выборок равны 10^6. 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%, ошибка квантиля q_{10} достигает 67%, ошибка квантиля q_{90} достигает 20%.

Таблица 1: Ошибка аппроксимации медианы (%) в зависимости от σ_1^2 (строка) и σ_2^2 (столбец) при $\mu_1 = \mu_2 = 4$.

	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

Таблица 2: Ошибка аппроксимации q_{10} (%) в зависимости от σ_1^2 (строка) и σ_2^2 (столбец) при $\mu_1 = \mu_2 = 4$.

	0.25	**0.75**	**1.25**	**1.75**	**2.25**
0.25	0.72	12.75	33.79	52.58	66.98
0.75	12.30	3.81	15.13	35.85	53.91
1.25	33.42	14.81	10.81	22.58	40.49
1.75	52.84	35.57	19.95	18.15	27.68
2.25	66.40	53.63	41.42	26.75	24.57

Таблица 3: Ошибка аппроксимации q_{90} (%) в зависимости от σ_1^2 (строка) и σ_2^2 (столбец) при $\mu_1 = \mu_2 = 4$.

	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

Построим графики 2, 3 и 4 зависимости ошибки аппроксимации квантилей от σ_2^2 при фиксированной $\sigma_1^2 = 0.45$. При моделировании объемы выборок равны 10^6.

Теперь 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 таблицах 4, 5 и 6.

Построим оценки плотности для ξ и η, когда ошибки имеют очень маленькие значения и когда достаточно большие. Они представлены на рисунках 5 и 6.

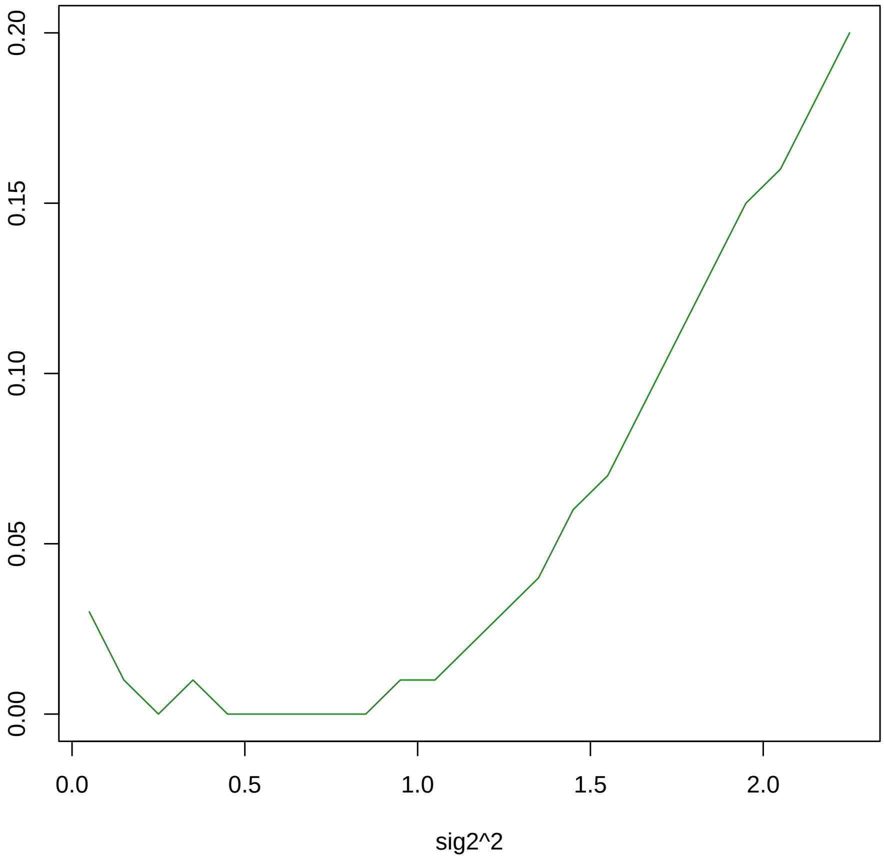

Рис. 2: Ошибка аппроксимации медианы при $\sigma_1^2 = 0.75$.

Таблица 4: $F_\eta(z_{50})$ (%) в зависимости от σ_1^2 (строка) и σ_2^2 (столбец) при $\mu_1 = \mu_2 = 4$.

	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

Таблица 5: $F_\eta(z_{10})$ (%) в зависимости от σ_1^2 (строка) и σ_2^2 (столбец) при $\mu_1 = \mu_2 = 4$.

	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

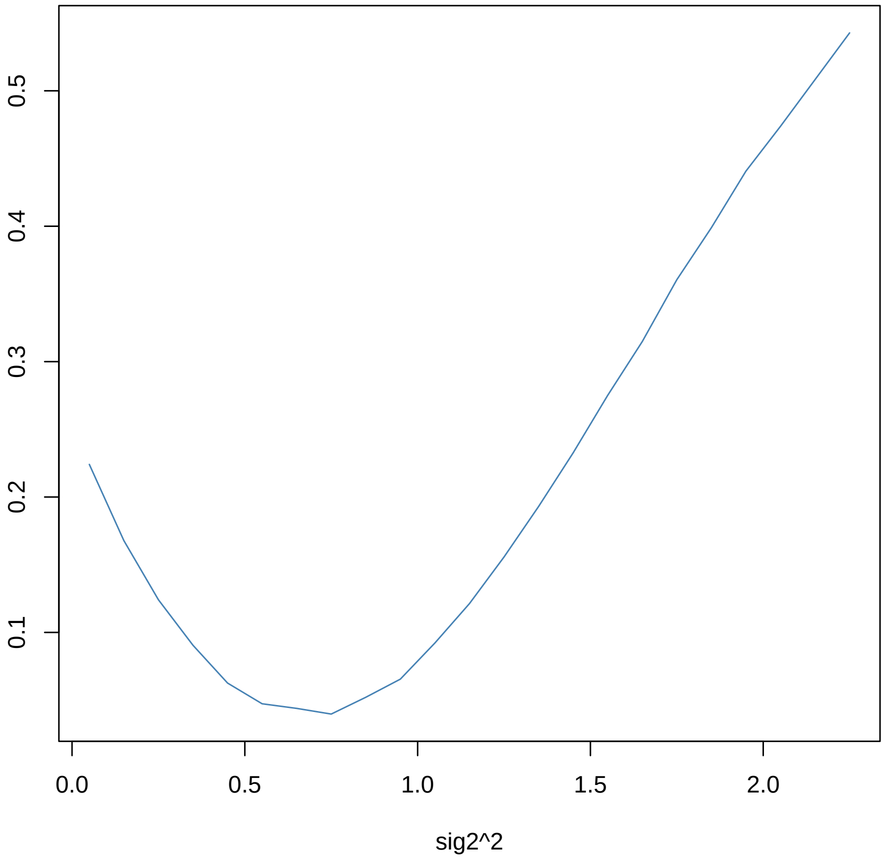

Рис. 3: Ошибка аппроксимации q_{10} при $\sigma_1^2 = 0.75$.

Таблица 6: $F_\eta(z_{90})$ (%) в зависимости от σ_1^2 (строка) и σ_2^2 (столбец) при $\mu_1 = \mu_2 = 4$.

	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

Также посмотрим на таблицы с коэффициентами асимметрии и эксцесса.

6 Трехточечная несимметричная аппроксимация логнормального распределения

Рассмотрим трехточечную аппроксимацию логнормального распределения с несимметричными квантилями. Пусть $\pi_1 = 0.1$, $\pi_2 = 0.5$. Посмотрим, как меняются вероят-

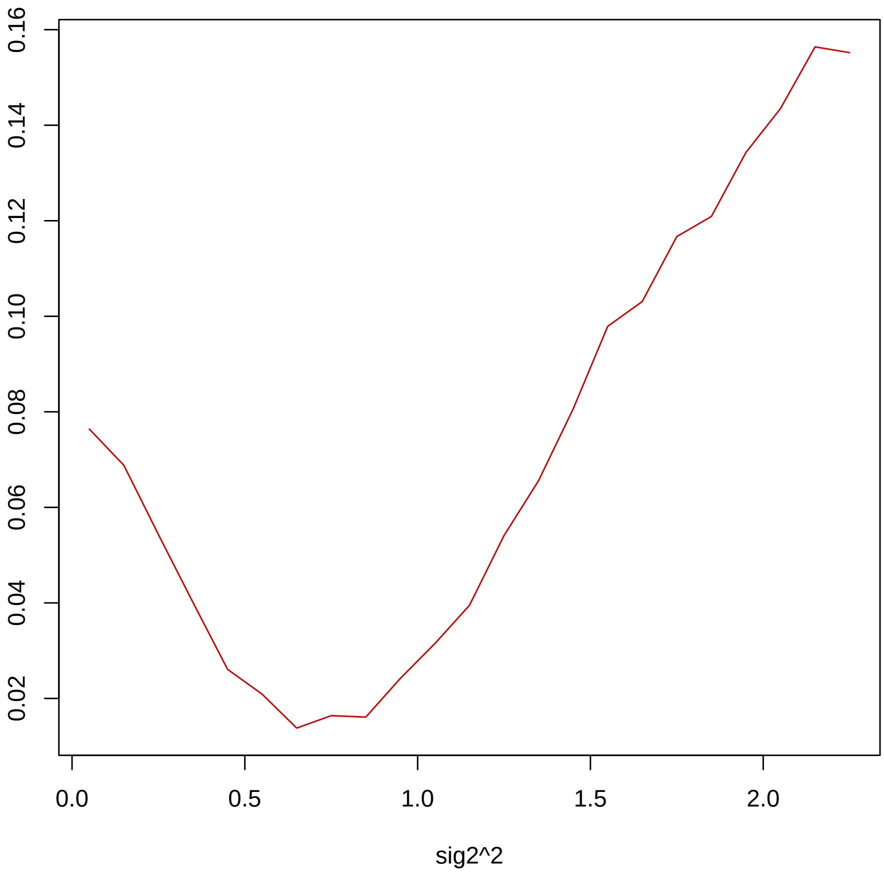

Рис. 4: Ошибка аппроксимации q_{90} при $\sigma_1^2 = 0.75$.

Таблица 7: Коэффициент асимметрии суммы (голубой) и аппроксимации (розовый) в зависимости от σ_1^2 (строка) и σ_2^2 (столбец) при $\mu_1 = \mu_2 = 4$.

	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

ности p_1, p_2, p_3 в зависимости от $\pi_3 = \pi$. На графике 7 веса, найденные при $\sigma = 0.8$. Видим, что веса являются вероятностями при $0.92 \le \pi \le 0.98$.

Теперь пусть $\pi_1 = 0.1$, $\pi_2 = \pi$, $\pi_3 = 0.9$. На рисунке 8 видим, что при любом π

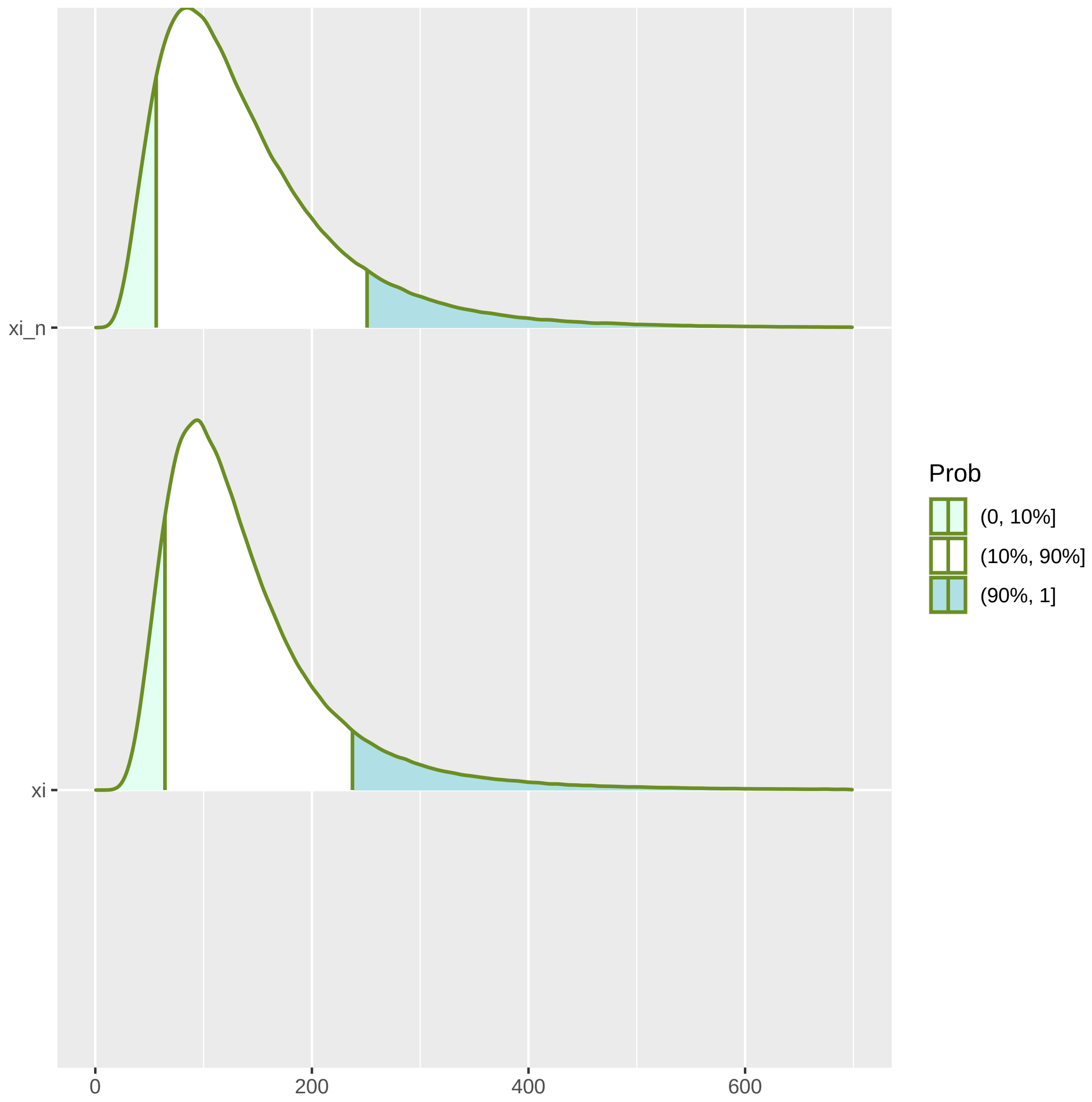

Рис. 5: $\sigma_1^2 = 0.25$, $\sigma_2^2 = 0.25$, $err_{med} = 0.17\%$, $err_{q_{10}} = 0.35\%$, $err_{q_{90}} = 0.12\%$.

значение p_2 будет отрицательным.

Теперь пусть $\pi_1 = \pi$, $\pi_2 = 0.5$, $\pi_3 = 0.9$. Видим, что веса являются вероятностями при $0 \leq \pi \leq 0.02$.

Теперь построим график зависимости верхней границы для σ от π для $\pi_1 = 1 - \pi$, $\pi_2 = 0.5$, $\pi_3 = \pi$ и для $\pi_1 = 0.1$, $\pi_2 = 0.5$, $\pi_3 = \pi$. Он представлен на рисунке 10, видим, что $0 \leq \sigma \leq 1.5$ при $\pi = 0.98$. До значения $\sigma = 1.5$ быстрее доходит зеленая линия, то есть в случае $\pi_1 = 0.1$, $\pi_2 = 0.5$, $\pi_3 = \pi$.

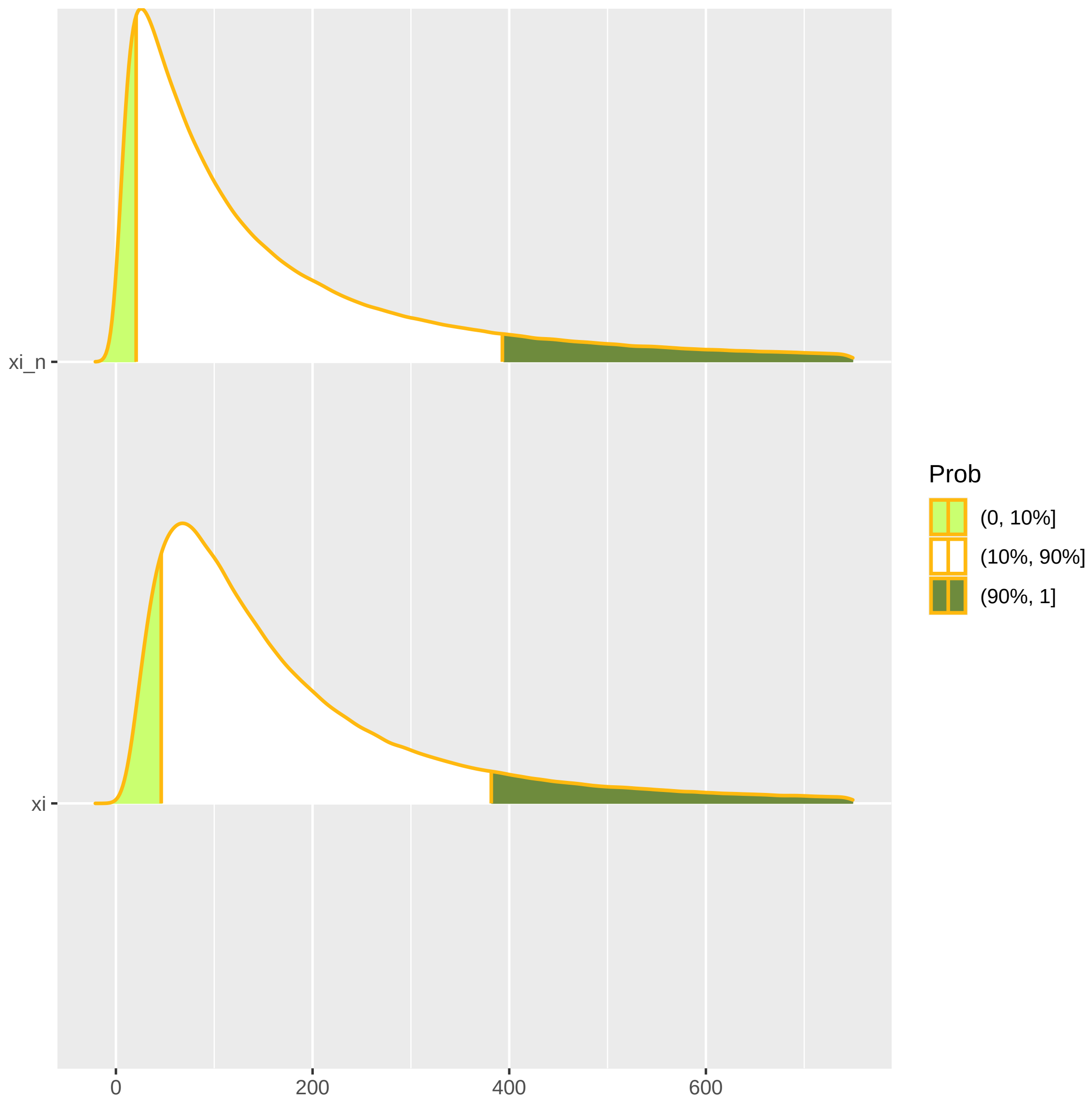

Рис. 6: $\sigma_1^2 = 2.25$, $\sigma_2^2 = 0.75$, $err_{med} = 20.4\%$, $err_{q_{10}} = 54.13\%$, $err_{q_{90}} = 15.54\%$.

Таблица 8: Коэффициент эксцесса суммы (голубой) и аппроксимации (розовый) в зависимости от σ_1^2 (строка) и σ_2^2 (столбец) при $\mu_1 = \mu_2 = 4$.

	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

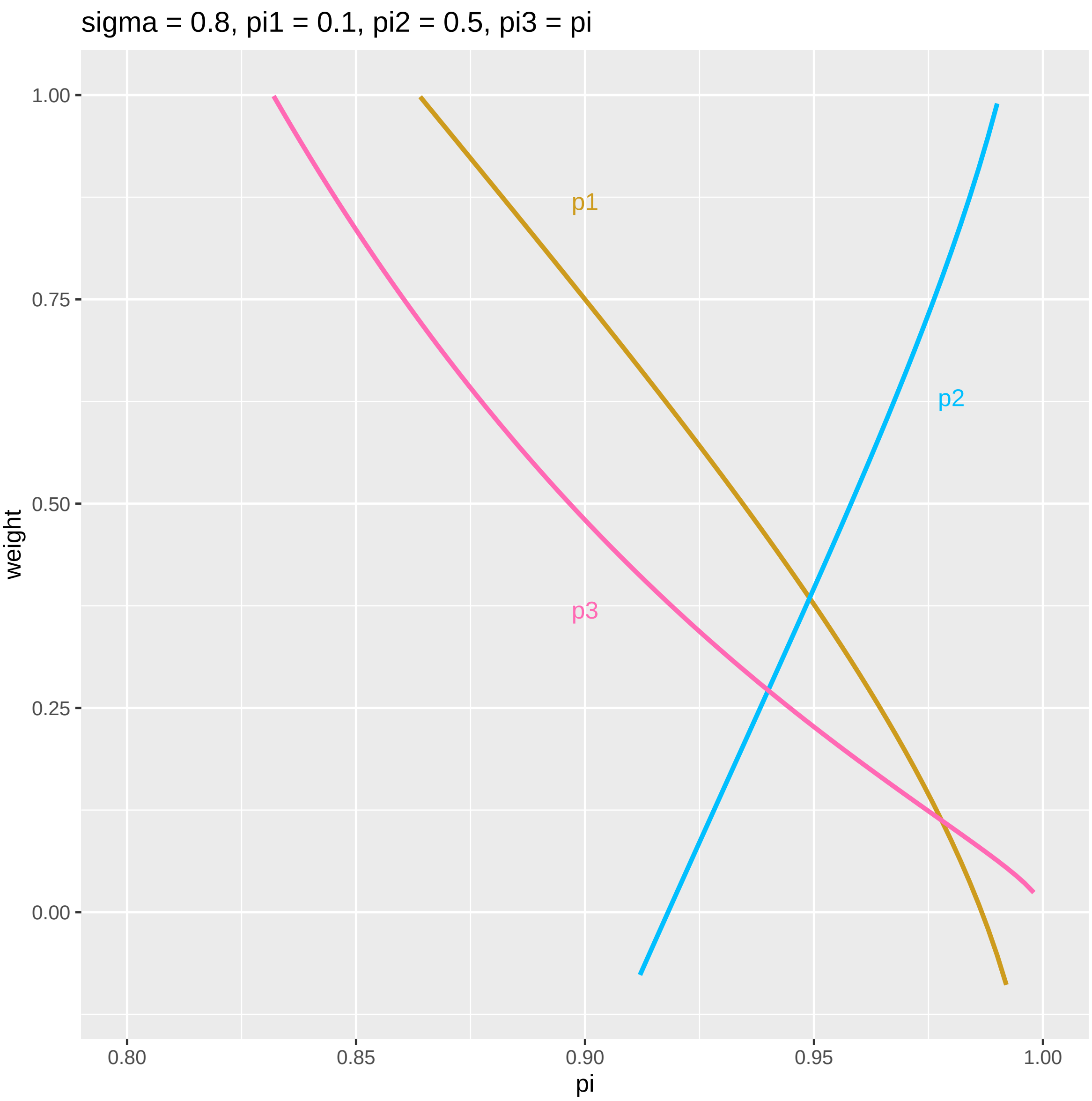

Рис. 7: Веса p_1, p_2, p_3 в зависимости от π.

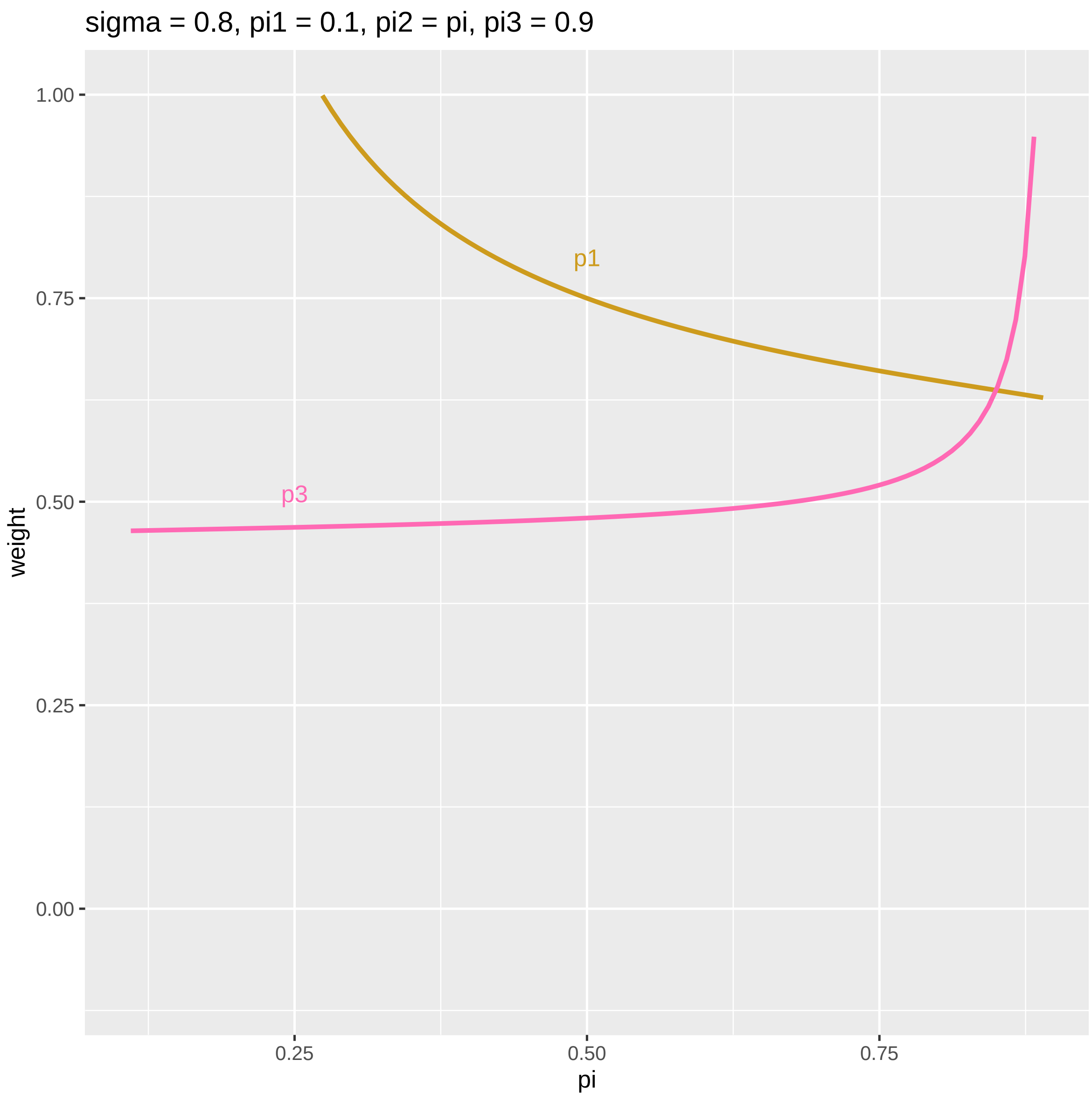

Рис. 8: Веса p_1, p_2, p_3 в зависимости от π.

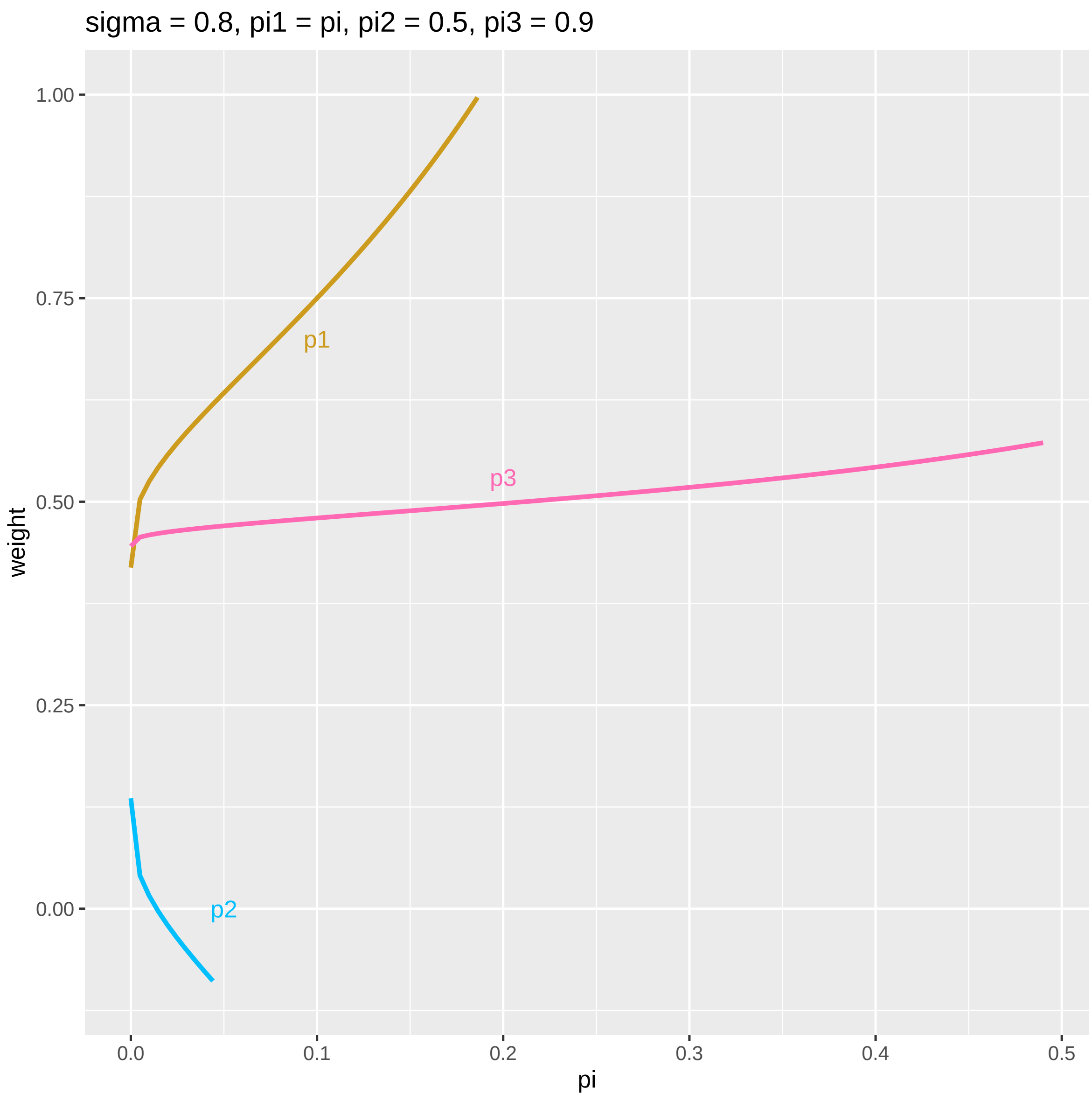

Рис. 9: Веса p_1, p_2, p_3 в зависимости от π.

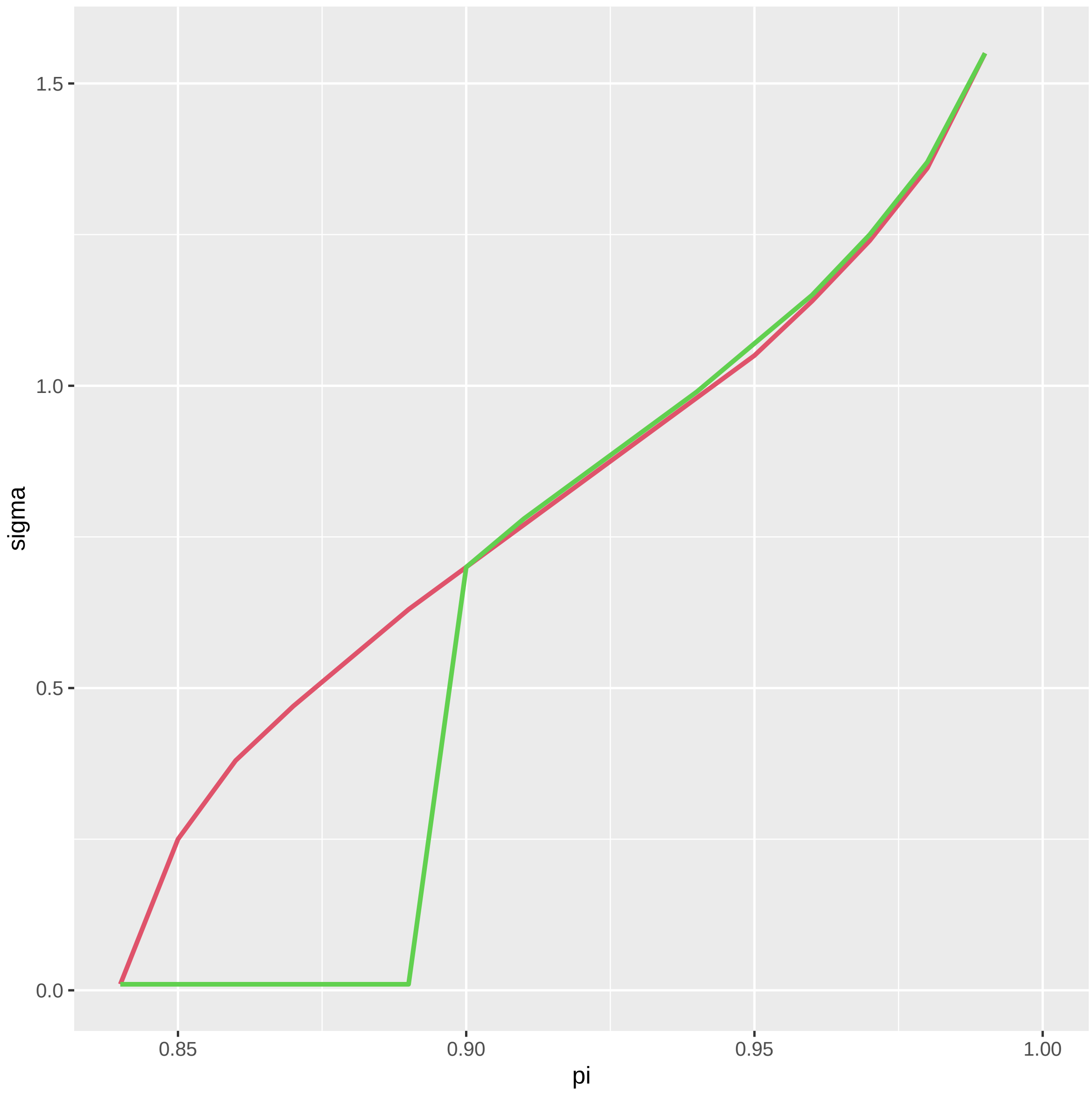

Рис. 10: Граница для σ при $\pi_1 = 1 - \pi$, $\pi_2 = 0.5$, $\pi_3 = \pi$ (красный), при $\pi_1 = 0.1$, $\pi_2 = 0.5$, $\pi_3 = \pi$ (зелёный).

7 Заключение

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

[1] Keith G. Swanson's Swansong.— Текст: электронный // stochastic: [сайт].— URL: https://www.stochastic.dk/post/swanson-s-swansong (дата обращения: 23.12.2021).

[2] Uncertainties impacting reserves, revenue, and costs— Текст: электронный // AAPG Wiki: [сайт].— URL: https://wiki.aapg.org/Uncertainties impacting reserves, revenue, and costs (дата обращения: 27.05.2022).

[3] 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.

[4] 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.

[5] Moghadasi, Maryam and Jerry L. Jensen. "Performance Evaluation of Swanson's Rule for the Case of Log-Normal Populations." (2014). DOI:10.1007/978-3-642-32408.