Замена непрерывного распределения на дискретное для применения на практике

Нагуманова Карина Ильнуровна, 19Б.04-мм

Санкт-Петербургский государственный университет
Прикладная математика и информатика
Вычислительная стохастика и статистические модели

Санкт-Петербург
2023г.

Введение

В практических задачах часто 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**.

Однако в ряде областей, например, в нефтяной промышленности распределением, описывающим запасы нефти, общепринятым является логнормальное распределение. Соответственно, реальной задачей является аппроксимация логнормального распределения.

Введение

Аппроксимируемые случайные величины складывают и умножают.

Пример перемножения: используем площадь дренирования пласта, среднюю чистую толщину и коэффициент извлечения углеводородов. При перемножении этих параметров получаем количество резервов нефти.

Пример сложения: зная запасы, параметры нефти и породы для всех залежей можно оценить профиль добычи нефти с каждой залежи и суммарный профиль, оценить экономическую эффективность проекта, которая учитывает выручку, налоги, капитальные затраты, операционные затраты, оптимальные решения по проекту.

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

Структура работы.

- В **разделе 2** рассмотрен общий подход к трехточечной аппроксимации.
- В **разделе 3** аппроксимация нормального распределения, вывод правила 30-40-30.
- В **разделе 4** рассматривается аппроксимация логнормального распределения, два метода, условие аппроксимации и что делать, если это условие не выполняется. А также точность аппроксимации при применении правила 30-40-30 к логнормальному распределению.
- В **разделе 5** алгоритм аппроксимации произведения двух логнормальных распределений.
- В **разделе 6** алгоритм аппроксимации суммы двух логнормальных распределений.

ξ — непрерывная случайная величина с математическим ожиданием m, дисперсией s^2 и функцией распределения $F(x)$. Для неё заданы квантили x_{π_1}, x_{π_2}, x_{π_3}. Также есть случайная дискретная величина ξ_n с математическим ожиданием m_n и дисперсией s_n^2.

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

Нужно аппроксимировать распределение случайной величины ξ дискретным распределением ξ_n.

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

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

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

Условия аппроксимации в общем случае

Предложение

Пусть верно

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

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

Доказательство:

$$\mathsf{P}\left(\frac{\xi - m}{s} \le \frac{x_{\pi_i} - m}{s}\right) = \tilde{\mathsf{F}}\left(\frac{x_{\pi_i} - m}{s}\right) = \pi_i,$$

$$\tilde{x}_{\pi_i} = \frac{x_{\pi_i} - m}{s} = \tilde{\mathsf{F}}^{-1}(\pi_i).$$

Условия аппроксимации в общем случае

Предположим, что $m = m_n$ и $s^2 = s_n^2$, и получим систему для вероятностей. Для этого подставим $\tilde{x}_{\pi_i}$ в формулу мат.ожидания, получаем

$$m(p_1 + p_2 + p_3) + s(p_1\tilde{x}_{\pi_1} + p_2\tilde{x}_{\pi_2} + p_3\tilde{x}_{\pi_3}) = m.$$

Получаем

$$s(p_1\tilde{x}_{\pi_1} + p_2\tilde{x}_{\pi_2} + p_3\tilde{x}_{\pi_3}) = 0.$$

$$p_1\tilde{x}_{\pi_1} + p_2\tilde{x}_{\pi_2} + p_3\tilde{x}_{\pi_3} = 0.$$

Подставим в формулу дисперсии

$$p_1(m + s\tilde{x}_{\pi_1})^2 + p_2(m + s\tilde{x}_{\pi_2})^2 + p_3(m + s\tilde{x}_{\pi_3})^2 - m^2 = s^2,$$

$$p_1\tilde{x}_{\pi_1}^2 + p_2\tilde{x}_{\pi_2}^2 + p_3\tilde{x}_{\pi_3}^2 = 1.$$

Получившиеся уравнения в матричной форме

$$\begin{pmatrix} 1 & 1 & 1 \\ \tilde{x}_{\pi_1} & \tilde{x}_{\pi_2} & \tilde{x}_{\pi_3} \\ \tilde{x}_{\pi_1}^2 & \tilde{x}_{\pi_2}^2 & \tilde{x}_{\pi_3}^2 \end{pmatrix} \begin{pmatrix} p_1 \\ p_2 \\ p_3 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}.$$

Аппроксимация нормального распределения

В общем случае вероятности p_1, p_2, p_3 будут зависеть от математического ожидания и дисперсии.

Если $\xi \sim N(\mu, \sigma)$ имеет нормальное распределение, то η имеет нормальное стандартное распределение, которое не зависит ни от μ, ни от σ.

Предложение

$\xi \sim N(\mu, \sigma)$, *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* $m = m_n$ *и* $s^2 = s_n^2$.

Аппроксимация нормального распределения. Случай симметричных квантилей

Доказательство: Предположим, что $m = m_n$ и $s^2 = s_n^2$, и получим значения вероятностей.

$\Phi(y) = \mathsf{P}(\eta = \dfrac{\xi - m}{s} \le y)$ — функция распределения стандартного нормального распределения, тогда

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

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

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

Аппроксимация нормального распределения. Случай симметричных квантилей

$$\begin{cases} p_\pi + p_{0.5} + p_{1-\pi} = 1, \\ (p_\pi - p_{1-\pi})\Phi^{-1}(\pi) = 0, \\ (p_\pi + p_{1-\pi})\Phi^{-1}(\pi)^2 = 1. \end{cases}$$

Обозначим $\delta = \dfrac{1}{\Phi^{-1}(\pi)^2}$, тогда из этой системы получим исходную. Рассмотрим 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**.

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 d 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 d логнормального распределения

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

$$s^2 = m^2[\exp(\sigma^2) - 1].$$

3. С помощью системы метода для нормального распределения находим значения вероятностей p_1, p_2, p_3.

Результат: вероятности p_1, p_2, p_3 для $x_{\pi_1}, x_{\pi_2}, x_{\pi_3}$ случайной величины ξ_n.

Аппроксимация логнормального распределения

Предложение

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

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

Доказательство:

$$\begin{aligned}
\tilde{\mathsf{F}}(y) = \mathsf{P}\left(\eta \leq y\right) = \mathsf{P}\left(\frac{\xi - m}{s} \leq y\right) = \\
= \mathsf{P}(\log(\xi) \leq \log(m + sy)) = \\
= \mathsf{P}\left(\frac{\log(\xi) - \mu}{\sigma} \leq \frac{\log(m + sy) - \mu}{\sigma}\right) = \\
= \Phi\left(\frac{\log(m + sy) - \mu}{\sigma}\right).
\end{aligned}$$

Аппроксимация логнормального распределения

Найдём $\log(m+sy)$, используя $m = e^{\mu+\frac{\sigma^2}{2}}$ и $s = m\sqrt{e^{\sigma^2}-1}$.

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

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

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

Аппроксимация логнормального распределения

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

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

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

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

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

В итоге получаем

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

Аппроксимация логнормального распределения

Предложение

Параметр σ логнормального распределения выражается как

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

Доказательство:

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

Следовательно,

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

и тогда

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

Аппроксимация логнормального распределения

С помощью двух квантилей можем исключить μ из соответствующих уравнений. Пусть есть π_1-ый и π_3-ый квантили со значениями x_{π_1} и x_{π_3}.

$$\log(x_{\pi_1}) = \mu + \sigma\Phi^{-1}(\pi_1),$$

$$\log(x_{\pi_3}) = \mu + \sigma\Phi^{-1}(\pi_3).$$

Вычтем из второго уравнения первое, получаем

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

И в итоге получаем

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

Аппроксимация логнормального распределения

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}

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

.

2. Вычисляем значения $\tilde{\mathsf{F}}^{-1}(\pi)$ для случайной величины η

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

3. Находим значения вероятностей p_1, p_2, p_3.

Результат: вероятности p_1, p_2, p_3 для квантилей $x_{\pi_1}, x_{\pi_2}, x_{\pi_3}$ случайной величины ξ_n.

Условие на параметр σ

Предложение

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

$$\exp(\sigma^2)+\exp(-\sigma^2)-\exp(\Phi(0.1)\sigma-\frac{\sigma^2}{2})-\exp(\Phi^{-1}(0.9)\sigma-\frac{\sigma^2}{2}) \leq 0,$$

(при $\sigma \leq 0.6913$, $\gamma_3 \leq 2.8278$ примерно).

Доказательство:

$$\ln(\eta) \sim N(\mu, \sigma^2), \qquad \tilde{\mathsf{F}}(y) = \mathsf{P}\left(\eta \leq y\right).$$

С помощью формулы (12) найдем $\tilde{\mathsf{F}}^{-1}(\pi_i)$ и сделаем следующие обозначения

$$\tilde{\mathsf{F}}^{-1}(0.1) = t_1, \qquad \tilde{\mathsf{F}}^{-1}(0.5) = t_2, \qquad \tilde{\mathsf{F}}^{-1}(0.9) = t_3.$$

Рассмотрим систему (7), запишем ее через t_1, t_2, t_3.

Условие на параметр σ

Выразим вероятности p_1, p_2, p_3.

$$p_2(t_2 - t_3) = p_1(t_3 - t_1) - t_3,$$

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

Тогда получаем

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

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

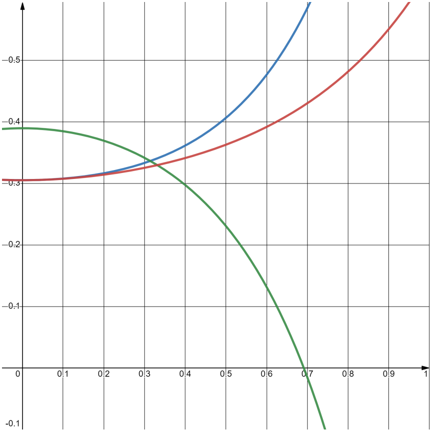
$$p_1 = \frac{1 + t_2 t_3}{(t_1 - t_3)(t_1 - t_2)},$$

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

$$p_3 = 1 - p_1 - p_2.$$

Условие на параметр σ

Построим график зависимости p_1, p_2, p_3 от σ.



$p_2 \geq$ при $\sigma \leq 0.6913$. Вероятности p_1 и p_3 положительные при любом параметре σ.

Получили условие $\sigma \leq 0.6913$.

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. Перейти к трехточечной аппроксимации дискретной случайной величиной ξ_n, у которой $m_n = m$ и $s^2 = s^2$. И считать значения m и s через квантили x_π, $x_{0.5}$, $x_{1-\pi}$ и вероятности p_1, p_2, p_3.
3. Если условие для положительных вероятностей не выполняется, можно воспринимать задачу не как поиск вероятностей для ξ_n, а как поиск коэффициентов для x_π, $x_{0.5}$, $x_{1-\pi}$ таких, чтобы параметры, посчитанные как мат.ожидание и дисперсия, были равны m и s^2.

Точность аппроксимации мат.ожидания

Проблема: правило 30-40-30 используют для логнормального распределения.
Вопрос: какова точность аппроксимации m и s^2?

Предложение

Ошибка аппроксимации мат.ожидания логнормального распределения с помощью правила 30-40-30 равна

$$\frac{\mid m-\widetilde{m}\mid}{m}=\mid\exp(\frac{\sigma^2}{2})-\frac{1}{2(\Phi^{-1}(0.1))^2}\times$$

$$\times(\exp(\sigma\Phi^{-1}(0.1))-1+\exp(\sigma\Phi^{-1}(0.9)))+1\mid/\exp\left(\frac{\sigma^2}{2}\right)$$

и не зависит от параметра μ.

Доказательство: Выразим ошибку аппроксимации через параметры μ и σ.

Точность аппроксимации мат.ожидания

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

Имеем следующие квантили

$$x_\pi = \exp(\mu + \sigma\Phi^{-1}(0.1)),$$

Точные значения вероятностей

$$p_1 = p_3 = \frac{1}{2(\Phi^{-1}(0.1))^2},$$

$$p_2 = 1 - \frac{1}{(\Phi^{-1}(0.1))^2}.$$

Получили ошибку

$$\frac{\mid m - \widetilde{m} \mid}{m} = \mid \exp(\frac{\sigma^2}{2}) - \frac{1}{2(\Phi^{-1}(0.1))^2} \times$$

$$\times(\exp(\sigma\Phi^{-1}(0.1)) - 1 + \exp(\sigma\Phi^{-1}(0.9))) + 1 \mid / \exp\left(\frac{\sigma^2}{2}\right)$$

Построим график зависимости от σ.

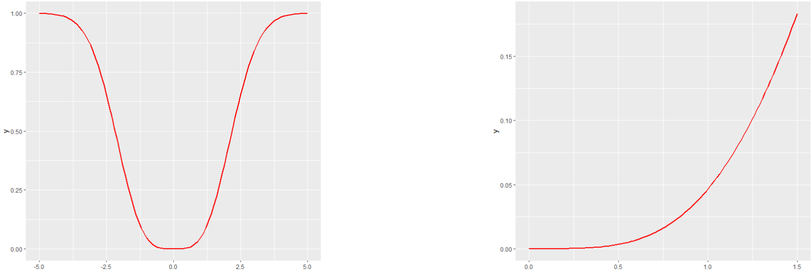


Рис.: Ошибка аппроксимации мат.ожидания

Видим, что при $\sigma \leq 1.5$ ошибка аппроксимации мат.ожидания меньше 12%.

Точность аппроксимации дисперсии

Предложение

Ошибка аппроксимации дисперсии логнормального распределения с помощью правила 30-40-30 равна

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

$$-\frac{1}{2(\Phi^{-1}(0.1))^2}\exp(2\sigma\Phi^{-1}(0.1)) -$$

$$-\left(1-\frac{1}{(\Phi^{-1}(0.1))^2}\right)\exp(2\sigma\Phi^{-1}(0.5)) -$$

$$-\frac{1}{2(\Phi^{-1}(0.1))^2}\exp(2\sigma\Phi^{-1}(0.9)) + m_2^2/2\mu \mid / \exp(\sigma^2)(\exp(\sigma^2-1))$$

и не зависит от параметра μ.

Доказательство: Имеем ошибку

$$\frac{\mid d-\widetilde{d}\mid}{d} = \mid \exp(2\mu+\sigma^2)(\exp(\sigma^2-1)) -$$

$$-\frac{1}{2(\Phi^{-1}(0.1))^2}\exp(2\mu+2\sigma\Phi^{-1}(0.1)) -$$

$$-\left(1-\frac{1}{(\Phi^{-1}(0.1))^2}\right)\exp(2\mu+2\sigma\Phi^{-1}(0.5)) -$$

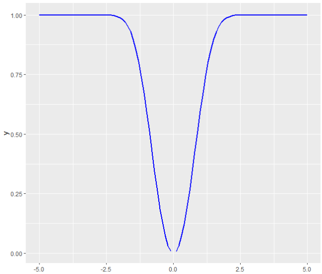
$$-\frac{1}{2(\Phi^{-1}(0.1))^2}\exp(2\mu+2\sigma\Phi^{-1}(0.9))+m_2^2 \mid / \exp(2\mu+\sigma^2)(\exp(\sigma^2-1))$$

$$= \mid \exp(\sigma^2)(\exp(\sigma^2-1)) - \frac{1}{2(\Phi^{-1}(0.1))^2}\exp(2\sigma\Phi^{-1}(0.1)) -$$

$$-\left(1-\frac{1}{(\Phi^{-1}(0.1))^2}\right)\exp(2\sigma\Phi^{-1}(0.5)) -$$

$$-\frac{1}{2(\Phi^{-1}(0.1))^2}\exp(2\sigma\Phi^{-1}(0.9))+m_2^2/2\mu \mid / \exp(\sigma^2)(\exp(\sigma^2-1)).$$

Построим график зависимости от σ.



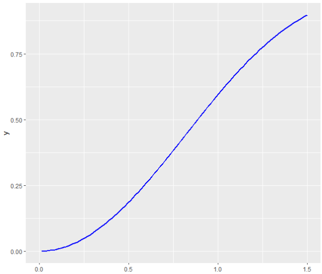


Рис.: Ошибка аппроксимации дисперсии

Видим, что при $\sigma \leq 1.5$ ошибка аппроксимации дисперсии может достигать 80%.

Рассмотрим две логнормально распределенные случайные величины.

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

Задача: замена непрерывной случайной величины, полученной с помощью произведения двух логнормально распределенных случайных величин, дискретной. То есть записать произведение в виде дискретной аппроксимации с кванилями того же вида.

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

Произведение π-ых квантилей

Предложение

При перемножении квантилей x_π и y_π двух логнормальных случайных величин ξ_1 и ξ_2 получается квантиль случайной величины $\xi_1\xi_2$ вида z_q, где

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

Доказательство: Выразим параметры распределений μ_1, μ_2, σ_1, σ_2 через квантили.

$$\mathsf{P}(\xi_1 < x_\pi) = \mathsf{P}(\ln(\xi_1) < \ln(x_\pi)) = \mathsf{P}\left(\frac{\ln(\xi_1) - \mu_1}{\sigma_1} < \frac{\ln(x_\pi) - \mu_1}{\sigma_1}\right).$$

$$\frac{\ln(\xi_1) - \mu_1}{\sigma_1} \sim N(0, 1).$$

Произведение π-ых квантилей

$$\ln(x_\pi) = \sigma_1 \Phi^{-1}(\pi) + \mu_1.$$

$$\mu_1 = \ln(x_{0.5}).$$

$$\sigma_1 = \frac{\ln(x_\pi) - \ln(x_{0.5})}{\Phi^{-1}(\pi)}.$$

$$\frac{\ln(y_\pi) - \mu_2}{\sigma_2} = \Phi^{-1}(\pi),$$

$$\mu_2 = \ln(y_{0.5}).$$

$$\sigma_2 = \frac{\ln(y_\pi) - \ln(y_{0.5})}{\Phi^{-1}(\pi)}.$$

Рассмотрим случайную величину $\eta = \xi_1 \xi_2$. Нужно вычислить, каким квантилем для η является произведение квантилей x_π и y_π.

Для этого надо найти, чему равна вероятность $\mathsf{P}(\xi_1 \xi_2 < x_\pi y_\pi)$.

Произведение π-ых квантилей

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

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

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

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

$$\ln(x_\pi) = \mu_1 + \Phi^{-1}(\pi)\sigma_1,$$

$$\ln(y_\pi) = \mu_2 + \Phi^{-1}(\pi)\sigma_2.$$

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

$$= \Phi\left(\frac{\Phi^{-1}(\pi)(\sigma_1 + \sigma_2)}{\sqrt{\sigma_1^2 + \sigma_2^2}}\right).$$

Произведение π-ых квантилей

$$\frac{\Phi^{-1}(\pi)(\sigma_1+\sigma_2)}{\sqrt{\sigma_1^2+\sigma_2^2}} =$$

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

Тогда получаем следующую формулу

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

Квантили вида π, 0.5, $1-\pi$

Предложение

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*

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

$$z_{0.5} = x_{0.5}y_5,$$

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

где a *и* σ *– параметры прямой* $\dfrac{x-a}{\sigma}$, *на которой лежат точки* $(\ln(x_\pi y_\pi), t)$ *и* $(\ln(x_{0.5}y_{0.5}), 0)$,

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

Квантили вида π, 0.5, $1-\pi$

Доказательство: С помощью формулы (18) можно посчитать, какой получается квантиль для случайной величины $\xi_1\xi_2$, если перемножить квантили x_π и y_π исходных случайных величин.

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

Получили снова 0.5-ый квантиль.
Обозначим z_π, $z_{0.5}$, $z_{1-\pi}$ — квантили случайной величины η.
Тогда $x_{0.5}y_{0.5} = z_{0.5}$.
С помощью точек $(\ln(x_\pi y_\pi), t)$ и $(\ln(x_{0.5}y_{0.5}), 0)$ можно найти параметры a и σ прямой, на которой они лежат.

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

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

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

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:

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

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

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

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

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

Сумма двух логнормальных распределений

Рассмотрим сумму двух логнормальных случайных величин.

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

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

$$\eta = \xi_1 + \xi_2.$$

Задача: найти квантили z_π, $z_{0.5}$, $z_{1-\pi}$ случайной величины η, а также вычислить вероятности p_1, p_2, p_3 такие, что $m = m_n$ и $s^2 = s_n^2$.

Чтобы найти z_{10}, z_{50}, z_{90} будем использовать аппроксимацию суммы логнормальных распределений логнормальным распределением. $\ln(\eta) \sim N(\mu, \sigma)$.

Альтернатива: Если это ограничение на σ не выполняется и мы не можем вычислить положительные вероятности, то можно не переходить к аппроксимации дискретным распределением, а вычислить значения мат. ожидания и дисперсии η с помощью квантилей z_π, $z_{0.5}$, $z_{1-\pi}$.

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. Найти параметры μ_1, σ_1, μ_2 и σ_2 через значения квантилей, используя формулы раздела 4.2.
2. Вычислить значения мат. ожидания m и дисперсии s^2 случайной величины $\xi_1 + \xi_2$, как суммы $m = m_1 + m_2$, $d = d_1 + d_2$, где m_1, d_1 – мат.ожидание и дисперсия ξ_1, а m_2, d_2 – случайной величины ξ_2. Они пересчитываются аналогично m и d.
3. Выразить параметры μ и σ нормального распределения через параметры m и d логнормального распределения.

Сумма двух логнормальных распределений

1. Вычислить, какой квантиль получается при сложении x_π и y_π, используя следующую формулу

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

$$= \mathsf{P}\left(\frac{\ln(\xi_1 + \xi_2) - \mu}{\sigma} < \frac{\ln(x_\pi + y_\pi) - \mu}{\sigma}\right) =$$

$$= \Phi\left(\frac{\ln(x_\pi + y_\pi) - \mu}{\sigma}\right).$$

2. Найти значения квантилей z_{10}, z_{50}, z_{90} по Алгоритму 2.

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

Точность аппроксимации

Ошибки аппроксимации квантилей q_{10}, q_{50}, q_{90} случайной величины η равны

$$\frac{|q_{10} - z_{10}|}{q_{10}}, \qquad \frac{|q_{50} - z_{50}|}{q_{50}}, \qquad \frac{|q_{90} - z_{90}|}{q_{90}},$$

где

$$z_{p*100} = F^{-1}_{\eta_n}(p) = \exp(\mu + \sigma\sqrt{2}\mathrm{erf}^{-1}(2p - 1)).$$

Значение квантилей q_i выражаются как $q_{p*100} = F^{-1}_{\eta}(p)$, где

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

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

Точность аппроксимации

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

Ниже приведены фрагменты таблиц с ошибками, выраженными в $\%$ для $0.05 < \sigma_1^2, \sigma_2^2 < 2.25$.

Таблица: Ошибка аппроксимации медианы

	0.25	**0.75**	**1.25**	**1.75**	**2.5**
0.25	0.58	0.29	0.89	2.64	6.41
0.75	0.13	0.12	2.15	4.88	7.27
1.25	0.01	0.83	2.94	5.58	10.02
1.75	2.23	0.52	3.61	6.74	9.84
2.5	9.15	3.35	3.25	6.76	9.89

Точность аппроксимации

Таблица: Ошибка аппроксимации q_{10}

	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

Таблица: Ошибка аппроксимации q_{90}

	0.25	**0.75**	**1.25**	**1.75**	**2.5**
0.25	1.01	3.00	4.24	4.10	3.40
0.75	0.04	2.51	4.11	3.26	5.45
1.25	1.44	1.81	3.29	3.93	5.82
1.75	8.25	2.60	2.93	3.60	4.49
2.5	18.17	3.00	3.30	2.44	4.99

Точность аппроксимации

- При аппроксимации суммы двух логнормальных распределений логнормальным распределением ошибки математического ожидания и дисперсии равны 0, то есть $m = m_n$ и $s^2 = s_n^2$. Но ошибки аппроксимации квантилей η могут достигать 11% для q_{50}, ошибка квантиля q_{10} достигает 51%, ошибка квантиля q_{90} достигает 40%.

- При этом значения мат ожидания и дисперсии η, вычисленные по этим квантилям тоже имеют нулевую ошибку.

Построим графики зависимости ошибки аппроксимации квантилей от σ_2^2 при фиксированной $\sigma_1^2 = 0.45$. При моделировании объемы выборок равны 10^6, а количество рассматриваемых σ_2^2 равно 23.

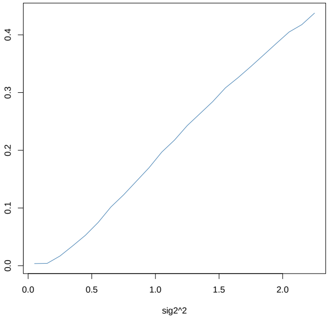


Рис.: Ошибка аппроксимации q_{10} при $\sigma_1^2 = 0.45$.

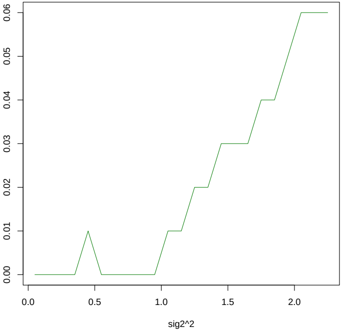


Рис.: Ошибка аппроксимации медианы при $\sigma_1^2 = 0.45$.

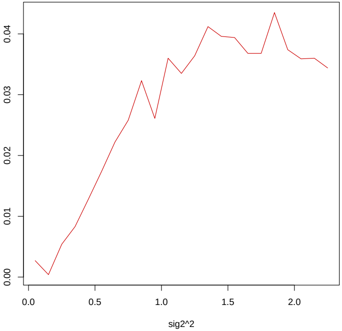


Рис.: Ошибка аппроксимации q_{90} при $\sigma_1^2 = 0.45$.

Точность аппроксимации

Посчитаем значения функции $F_\eta(x)$ от квантилей z_{10}, z_{50}, z_{90} случайной величины η_n при $0.15 < \sigma_1^2, \sigma_2^2 < 2.25$. Они показывают, каким квантилем для η являются квантили z_i. Результаты приведены в следующих таблицах.

Таблица: $F_\eta(z_{50})$ в % при $0.15 < \sigma_1^2, \sigma_2^2 < 2.25$

	0.15	**0.85**	**1.55**	**2.25**
0.15	0.51	0.50	0.49	0.48
0.85	0.50	0.50	0.48	0.47
1.55	0.49	0.50	0.48	0.46
2.25	0.40	0.49	0.48	0.46

Точность аппроксимации

Таблица: $F_\eta(z_{10})$ при $0.15 < \sigma_1^2, \sigma_2^2 < 2.25$

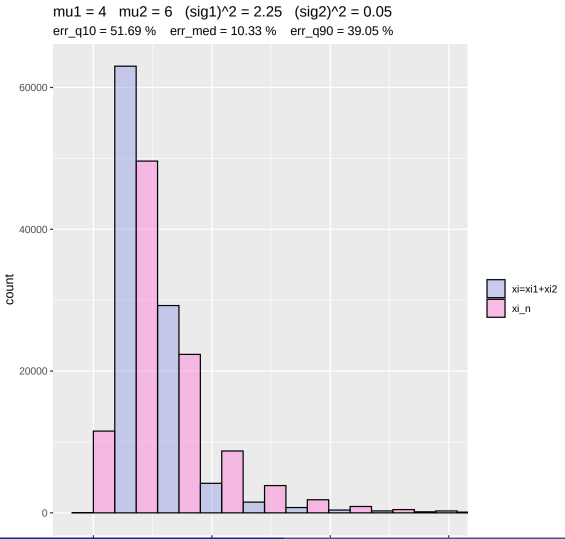
	0.15	**0.85**	**1.55**	**2.25**
0.15	0.09	0.06	0.04	0.02
0.85	0.10	0.07	0.05	0.03
1.55	0.05	0.08	0.06	0.04
2.25	0.00	0.08	0.07	0.05

Таблица: $F_\eta(z_{90})$ при $0.15 < \sigma_1^2, \sigma_2^2 < 2.25$

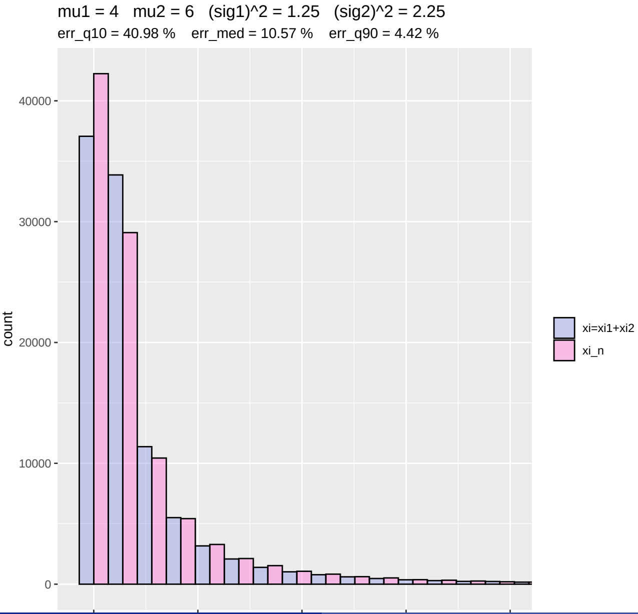
	0.15	**0.85**	**1.55**	**2.25**
0.15	0.90	0.90	0.90	0.90
0.85	0.90	0.91	0.91	0.90
1.55	0.93	0.90	0.91	0.91
2.25	0.95	0.91	0.90	0.91

Точность аппроксимации

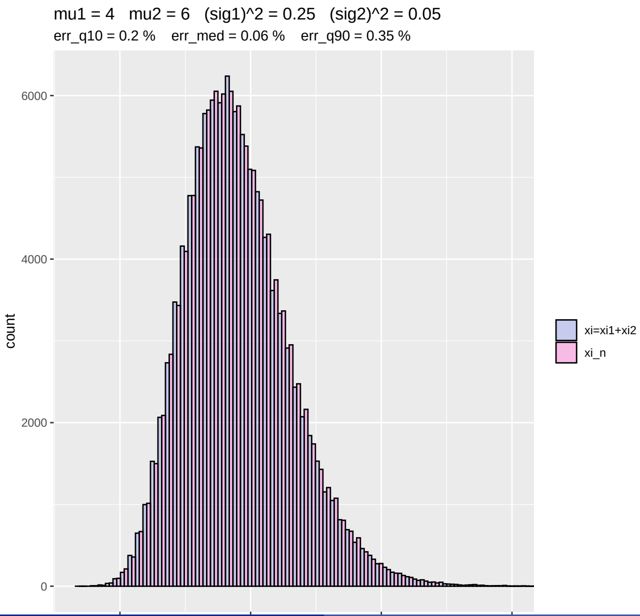
Также построим гистограммы для η и η_n, когда ошибки имеют очень маленькие значения и когда достаточно большие.



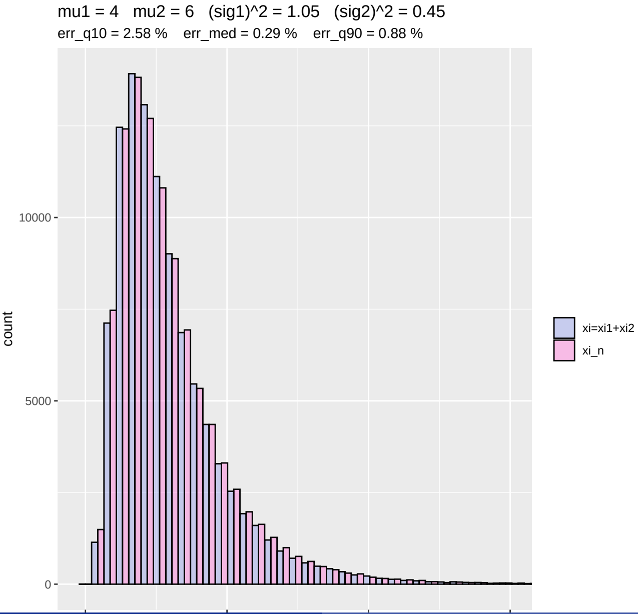
Точность аппроксимации



Точность аппроксимации



Точность аппроксимации



Результаты:

1. Методы аппроксимации нормального, логнормального распределений.
2. Условие на σ для аппроксимации.
3. Точность аппроксимации логнормального правилом 30-40-30.
4. Метод аппроксимации произведения логнормальных распределений.
5. Метод аппроксимации суммы логнормальных распределений.
6. Точность аппроксимации суммы логнормальных распределений.

Проблемы:

1. Аппроксимировать дискретным распределением получается только при ограниченных σ.
2. Для суммы логнормальных результат имеет ошибки, так как сумма логнормальных не логнормальная.