Замена непрерывного распределения на дискретное для применения на практике

Нагуманова Карина Ильнуровна, 19Б.04-мм

Санкт-Петербургский государственный университет
Прикладная математика и информатика

Санкт-Петербург
2023г.

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**.

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.

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 нефти в разных скважинах, нужно оценить суммарные запасы.

Задача: находить аппроксимацию суммы и произведения 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.

Часть 1: Общий подход к трехточечной аппроксимации

ξ — непрерывная случайная величина

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

$F(x)$ — 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

Предложение (Swanson)

Пусть верно

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

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

Если $\xi \sim N(\mu, \sigma)$ имеет нормальное распределение, то $\hat{\xi}$ имеет нормальное стандартное распределение, поэтому можно написать систему, которая не зависит от μ и σ.

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

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**.

Часть 3: Связь логнормального распределения с нормальным

Моменты m, s^2 логнормального распределения выражаются через моменты μ, σ^2 соответствующего нормального распределения:

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

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

Параметр σ выражается как

$$\sigma = \frac{\log\left(\dfrac{x_{\pi_j}}{x_{\pi_i}}\right)}{\Phi^{-1}(\pi_j) - \Phi^{-1}(\pi_i)}, \qquad i \neq j.$$

Значение σ одинаковое для любых пар i и j.

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

Результат не зависит от i.

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 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 $\tilde{\xi}$.

Часть 3: Условие на параметр σ

Мною доказаны следующие предложения.

Предложение

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

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

где $c = \Phi^{-1}(\pi)$.

Предложение

При уменьшении значения π ограничение на σ становится слабее, то есть 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

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

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

   Если p_1, p_2, p_3 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 метода Свонсона для логнормального распределения

Проблема: метод Свонсона, применяемый к нормальному распределению, используют для логнормального распределения.
Вопрос: какова точность аппроксимации m и s^2?

Предложение

Ошибка аппроксимации мат.ожидания логнормального распределения по методу Свонсона, применяемому к нормальному распределению, 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.

Предложение

Ошибка аппроксимации дисперсии логнормального распределения по методу Свонсона, применяемому к нормальному распределению, равна

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

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

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

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

Часть 3: Точность метода Свонсона для логнормального распределения

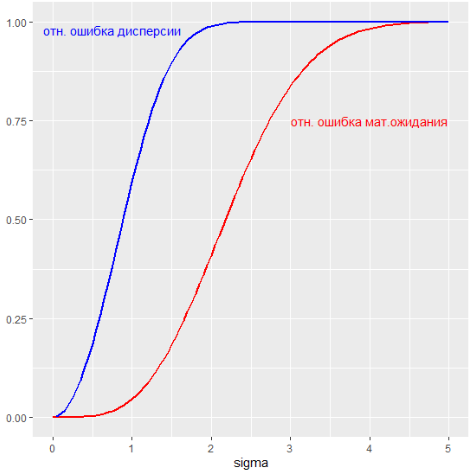

Рис.: Относительная ошибка аппроксимации мат.ож. и дисперсии

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

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* $(\ln(x_{0.5}y_{0.5}), 0)$, *где*

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

Часть 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.

Далее 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. $x_\pi,\ x_{0.5},\ x_{1-\pi} \to \mu_1,\ \sigma_1$
2. $y_\pi,\ y_{0.5},\ y_{1-\pi} \to \mu_2,\ \sigma_2$
3. $\mu_i,\ \sigma_i \to m_i,\ s_i^2$
4. $m = m_1 + m_2$
5. $s^2 = s_1^2 + s_2^2$
6. $m,\ s^2 \to \mu,\ \sigma$
7. $\mu,\ \sigma \to z_\pi,\ z_{0.5},\ z_{1-\pi}$
8. $z_\pi,\ z_{0.5},\ z_{1-\pi} \to p_1,\ p_2,\ p_3$

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. Точность аппроксимации

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

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

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

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

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

Часть 5: Сумма двух логнормальных распределений. Точность аппроксимации

Рассмотрим $\ln(\xi_1) \sim N(4, \sigma_1^2)$, $\ln(\xi_2) \sim N(6, \sigma_2^2)$ 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**	**2.5**
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

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_{90} (%) в зависимости от σ_1^2 (строка) и σ_2^2 (столбец)

	0.25	**0.75**	**1.25**	**1.75**	**2.5**
0.25	1.01	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

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)

	0.15	**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

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)

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

	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

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.

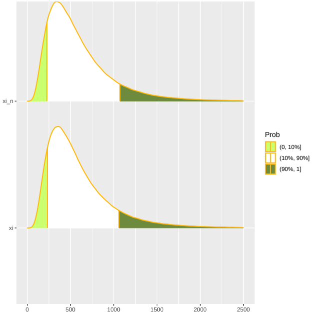

Рис.: $\sigma_1^2 = 1.05$, $\sigma_2^2 = 0.45$,
err_{med} = 0.09%,
$err_{q_{10}} = 2.4\%$,
$err_{q_{90}} = 1.1\%$.

Часть 5: Сумма двух логнормальных распределений. Точность аппроксимации

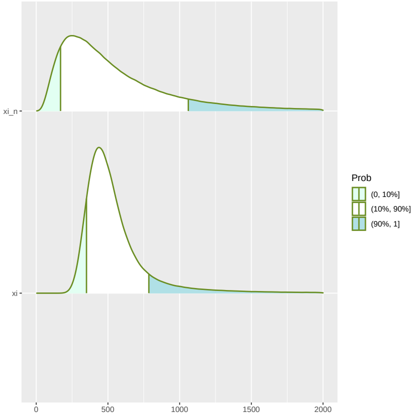

Рис.: $\sigma_1^2 = 2.25$, $\sigma_2^2 = 0.05$, $err_{med} =$ 9.92%, $err_{q_{10}} = 51.34\%$, $err_{q_{90}} = 39.88\%$.

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.