Замена непрерывного распределения на дискретное для применения на практике

Нагуманова Карина Ильнуровна, 19Б.04-мм

Санкт-Петербургский государственный университет
Математико-механический факультет
Кафедра статистического моделирования

Научный руководитель: д.ф.-м.н., доцент Голяндина Н. Э.
Рецензент: лектор Кардиффского университета Пепелышев А. Н.

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**.

Аппроксимируемые случайные величины складывают и умножают.

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

План работы.

1. Рассмотреть общий подход к трехточечной аппроксимации,трехточечную аппроксимацию нормального распределения, метод Свонсона и вывод правила 30-40-30.
2. Рассмотреть трехточечную аппроксимацию логнормального распределения и её свойства.
3. Построить алгоритм аппроксимации произведения двух логнормальных распределений.
4. 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,
$\tilde{\xi}$ — дискретная случайная 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

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

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

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

Рассмотрим частный случай $\xi \sim N(\mu, \sigma), \pi = 0.1$, имеем

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

$$p_1 \approx 0.305, \qquad p_2 \approx 0.390, \qquad p_3 \approx 0.305.$$

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

Пусть $\xi = \ln(\eta)$ и $\xi \sim N(\mu, \sigma)$. Параметры $m = \mathbf{E}(\eta)$, $s^2 = \mathbf{D}(\eta)$ логнормального распределения выражаются как

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

Обратная функция распределения η имеет вид

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

Предложение

Параметр μ выражается как $\mu = \log(x_{\pi_i}) - \sigma\Phi^{-1}(\pi_i)$ и результат не зависит от i. Параметр σ выражается через любые два квантиля как

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

Часть 2: 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. Выражаем параметры μ и σ математическое ожидание и дисперсию соответствующего нормального распределения через известные $x_{\pi_1}, x_{\pi_2}, x_{\pi_3}$.
2. Вычисляем значения математического ожидания m и дисперсии s^2 случайной величины η, используя μ и σ.
3. С помощью системы уравнений 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 $\tilde{\xi}$.

В реальных задачах в нефтяной промышленности используются следующие диапазоны параметров: $\mu \leq 12$, $\sigma \leq 1.5$.

Часть 2: Условие на параметр σ

Мною было доказано следующее предложение.

Предложение

Неотрицательные вероятности p_1, p_2, p_3 для аппроксимации логнормальной случайной величины η в точках π-, 0.5- и $(1-\pi)$- квантилей 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 $c = \Phi^{-1}(\pi)$.

Следствие

При уменьшении значения π диапазон существования дискретной вероятностной аппроксимации увеличивается.

Например, для $\pi = 0.1$ получаем ограничение $\sigma \leq 0.6913$.

Часть 2: Точность метода Свонсона для логнормального распределения

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

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

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

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

Часть 2: Точность метода Свонсона для логнормального распределения

Предложение

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

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

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

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

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

Часть 2: Точность метода Свонсона для логнормального распределения

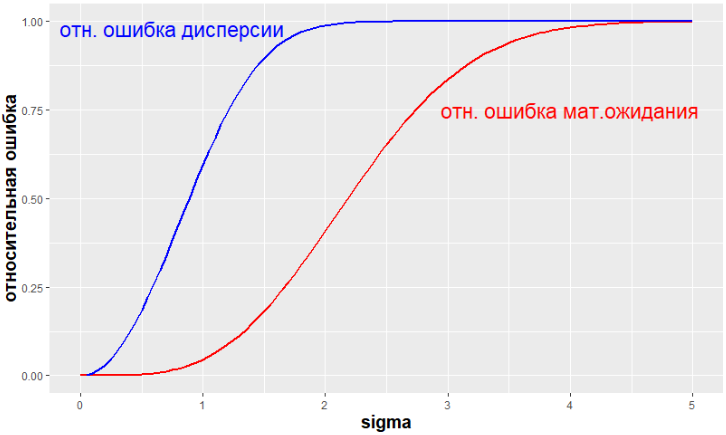

Рис.: Относительная ошибка аппроксимации математического ожидания и дисперсии

Часть 3: 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: Сумма двух логнормальных распределений

Рассмотрим сумму двух логнормальных случайных величин

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

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

$$\xi = \xi_1 + \xi_2,$$

где ξ_1 и ξ_2 заданы своими квантилями.

Поставим задачу аппроксимации суммы логнормальным распределением $\ln(\eta) \sim N(\mu, \sigma)$.

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

Часть 4: 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 $\xi_1 + \xi_2$.

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

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

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

Часть 4: Точность аппроксимации

Рассмотрим $\ln(\xi_1) \sim N(4, \sigma_1^2)$, $\ln(\xi_2) \sim N(4, \sigma_2^2)$ и найдем ошибки в зависимости от σ_1 (строка) и σ_2 (столбец) с помощью моделирования, объемы выборок равны 10^6.

Таблица: Ошибка аппроксимации q_{10} (%) (слева) и q_{90} (%) (справа)

	0.5	**1.5**
0.5	0.67	67.15
1.5	66.62	26.01

	0.5	**1.5**
0.5	0.29	18.04
1.5	20.29	3.83

Таблица: Коэффициент асимметрии суммы (голубой) и аппроксимации (розовый)

	0.5	**1.5**
0.5	1.77	16.68
	1.53	29.70
1.5	11.18	23.61
	3.30	25.13

Построим оценки плотности для ξ и η при $\mu_1 = \mu_2 = 4$ и вычислим ошибки аппроксимации. Получили $err_{med} = 0.12\%$, $err_{q10} = 0.45\%$, $err_{q90} = 0.28\%$.

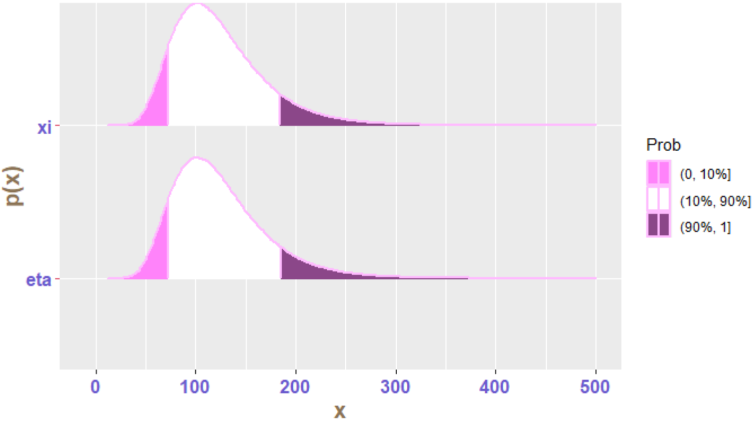

Рис.: $\sigma_1 = 0.5$, $\sigma_2 = 0.5$.

Теперь посмотрим на случай, когда ошибки аппроксимации достаточно большие. Получили $err_{med} = 20.9\%$, $err_{q_{10}} = 66.7\%$, $err_{q_{90}} = 19.1\%$.

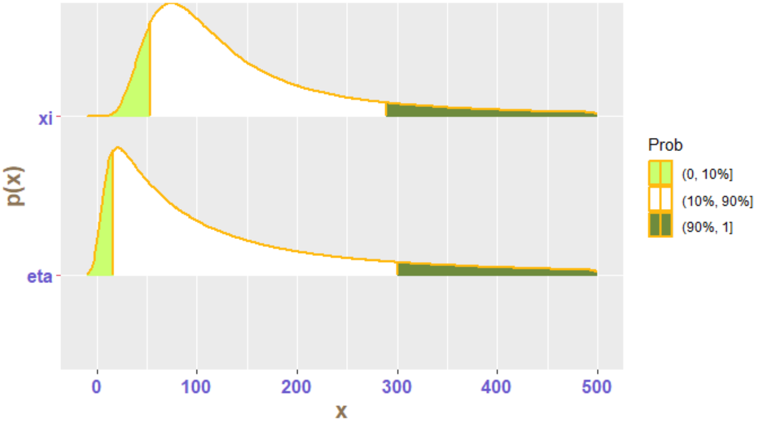

Рис.: $\sigma_1 = 1.5$, $\sigma_2 = 0.5$.

Заключение

Мною были получены следующие результаты:

1. Получено условие на σ для существования трехточечной симметричной вероятностной аппроксимации логнормального распределения.
2. Численно оценена точность аппроксимации математического ожидания и дисперсии логнормального распределения с помощью метода Свонсона, применяемого к нормальному распределению.
3. Формально и полно написано обоснование алгоритма для нахождения трехточечной симметричной аппроксимации произведения логнормальных распределений.
4. Построен алгоритм для нахождения трехточечной симметричной аппроксимации суммы логнормальных распределений.
5. Численно оценена точность трехточечной симметричной аппроксимации суммы логнормальных распределений.