

Текст выступления «Замена непрерывного распределения на дискретное для применения на практике».

Слайд №2

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. Однако в ряде областей, например, в нефтяной промышленности распределением, описывающим запасы нефти, общепринятым является логнормальное распределение. Аппроксимация по методу Свонсона для нормального распределения используется в этих областях, хотя распределение и логнормальное. Также аппроксимируемые случайные величины складывают и умножают. Соответственно, реальной задачей является аппроксимация суммы и произведения логнормальных случайных величин по аппроксимациям исходных случайных величин.

Слайд №3

План работы состоит в следующем.

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.

Слайд №4

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. Пусть дана непрерывная случайная величина ξ с функцией распределения $F(x)$. Обозначим $m = \mathbf{E}(\xi)$, $s^2 = \mathbf{D}(\xi)$. Для неё заданы квантили $x_{\pi_1}, x_{\pi_2}, x_{\pi_3}$. Также есть случайная дискретная величина $\tilde{\xi}$, которая задана следующим образом. Для неё тоже делаем обозначения математического ожидания и дисперсии. Задача: аппроксимировать распределение случайной величины ξ дискретным распределением $\tilde{\xi}$, то есть найти p_1, p_2, p_3 такие, что следующие равенства верные.

Слайд №5

Рассмотрим предложение. Пусть верна следующая система, где $\hat{x}_{\pi_i} = \hat{F}^{-1}(\pi_i)$, $\hat{F}(y)$ — функция распределения $\hat{\xi} = \frac{\xi - m}{s}$. Тогда $m = \tilde{m}$ и $s^2 = \tilde{s}^2$. Это Предложение дает требуемую аппроксимацию дискретным распределением, если найденные вероятности p_i являются неотрицательными. Рассмотрим частный случай $\xi \sim N(\mu, \sigma)$, $\pi = 0.1$, получаем $p_1 \approx 0.305$, $p_2 \approx 0.390$, $p_3 \approx 0.305$. Эти вероятности примерно равны 0.3, 0.4, 0.3, поэтому это правило называют правилом 30-40-30.

Слайд №6

Пусть $\xi = \ln(\eta)$ и $\xi \sim N(\mu, \sigma)$. Знаем, как выразить параметры логнормального распределения через параметры нормального. По следующим формулам. Обратная функция распределения η имеет следующий вид. Также мною было доказано это предложение. Зная любые два квантиля, можно через них выразить параметры μ и σ .

Слайд №7

Имеем следующий алгоритм аппроксимации 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: $\mu \leq 12$, $\sigma \leq 1.5$. Поэтому мы будем обращать на них особое внимание.

Слайд №8

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. Мною было доказано следующее предложение, а также следствие из него. Таким образом, мы получили условие для существования дискретной вероятностной аппроксимации. Например, для $\pi = 0.1$ получаем ограничение $\sigma \leq 0.6913$, $\sigma^2 \leq 0.4779$.

Слайд №9

Возникает проблема: метод Свонсона выведенный для аппроксимации нормального распределения используют для логнормального. Какова точность аппроксимации m и s^2 ? Мною доказано предложение, формула для вычисления относительной ошибки аппроксимации математического ожидания представлена на слайде, результат не зависит от параметра μ .

Слайд №10

Также мною была доказана формула для вычисления ошибки аппроксимации дисперсии, она представлена на слайде, результат не зависит от параметра μ .

Слайд №11

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

Слайд №12

Мною было подробно написано обоснование алгоритма для нахождения трехточечной симметричной аппроксимации произведения логнормальных распределений. Верно следующее предложение. Зная квантили случайной величины ξ_1 и квантили случайной величины ξ_2 , можно найти квантили соответствующие квантили случайной величины $\xi_1\xi_2$ по следующим формулам, представленным на экране. Здесь a и b такие, что прямая $y = \frac{x-a}{b}$, проходит через точки $(\ln(x_\pi y_\pi), t)$ и $(\ln(x_{0.5} y_{0.5}), 0)$. Значение t выражается через известные квантили следующим образом.

Слайд №13

Рассмотрим сумму двух логнормальных случайных величин $\xi = \xi_1 + \xi_2$. 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. То есть нужно найти квантили $z_\pi, z_{0.5}, z_{1-\pi}$ случайной величины η .

Слайд №14

Имеем следующий алгоритм для решения задачи. Дано: квантили $x_\pi, x_{0.5}, x_{1-\pi}$ — квантили ξ_1 , $y_\pi, y_{0.5}, y_{1-\pi}$ — квантили ξ_2 .

1. По набору квантилей ξ_1 находим параметры μ_1, σ_1 нормального распределения.
2. По набору квантилей ξ_2 находим параметры μ_2, σ_2 нормального распределения.
3. Находим математические ожидания и дисперсии ξ_1 и ξ_2 .
4. Вычисляем математическое ожидание $\xi_1 + \xi_2$.
5. Вычисляем дисперсию $\xi_1 + \xi_2$.
6. Находим параметры нормального распределения.
7. Находим значения квантилей через μ и σ .
8. 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 суммы логнормальным распределением.

Слайд №15

Выразим ошибки аппроксимации квантилей q_π , $q_{0.5}$, $q_{1-\pi}$ случайной величины ξ через параметры μ_1 , μ_2 , σ_1^2 , σ_2^2 . Здесь значения z_π , $z_{0.5}$, $z_{1-\pi}$ вычисляются через обратную функцию распределения для η . Параметры μ , σ можно найти через параметры случайных величин ξ_1 , ξ_2 и вычисленные значения $m = \exp\left(\mu_1 + \frac{\sigma_1^2}{2}\right) + \exp\left(\mu_2 + \frac{\sigma_2^2}{2}\right)$, $s^2 = m_1^2(\exp(\sigma_1^2) - 1) + m_2^2(\exp(\sigma_2^2) - 1)$. Квантили η выражаются через функцию распределения $\xi = \xi_1 + \xi_2$, найденную с помощью формулы свертки.

Слайд №16

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_1^2 (строка) и σ_2^2 (столбец) с помощью моделирования, объемы выборок равны 10^6 . Ошибки представлены в этих двух таблицах. Также вычислим для данных значений σ_1 и σ_2 коэффициенты асимметрии и коэффициент эксцесса. Они представлены на следующей таблице.

Слайд №17

Построим оценки плотности для ξ и η при $\mu_1 = \mu_2 = 4$ и вычислим ошибки аппроксимации. Для $\sigma_1^2 = 0.25$, $\sigma_2^2 = 0.25$ получили $err_{med} = 0.12\%$, $err_{q_{10}} = 0.45\%$, $err_{q_{90}} = 0.28\%$.

Слайд №18

Теперь посмотрим на случай, когда ошибки аппроксимации достаточно большие. Для $\sigma_1^2 = 2.25$, $\sigma_2^2 = 0.25$ получили $err_{med} = 20.9\%$, $err_{q_{10}} = 66.7\%$, $err_{q_{90}} = 19.1\%$.

Слайд №19

Таким образом, 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.