Математика и статистика, часть 2
Сумма нормально распределенных случайных величин. Теорема Муавра-Лапласа. (20.03.2020)
А. А. Макаров, А. А. Тамбовцева, Н. А. Василёнок

Задача 1. X и Z – независимые случайные величины. Случайная величина X имеет нормальное распределение $N(3, \sigma^2 = 5)$, а Z имеет стандартное нормальное распределение. Для случайной величины $U = X + 2Z - 4$:

(a) Укажите, какое распределение будет иметь эта случайная величина и каковы его параметры;

Во-первых, обратим внимание, что случайная величина Z имеет стандартное нормальное распределение с параметрами: $Z \sim N(0, \sigma^2 = 1)$. Случайная величина U, как и X и Z, будет иметь нормальное распределение. Чтобы вычислить его параметры, вспомним про свойства математического ожидания и дисперсии:

$$E(U) = E(X + 2Z - 4) = E(X) + 2E(Z) - 4 = 3 + 2 \times 0 - 4 = -1$$

Так как случайные величины независимы, то дисперсия их суммы будет равна сумме дисперсий:

$$D(U) = E(X + 2Z - 4) = D(X) + 2^2 \times D(Z) = 5 + 4 \times 1 = 9$$

Получаем: $U \sim N(-1, \sigma^2 = 9)$ или $U \sim N(-1, \sigma = 3)$

(b) Рассчитайте вероятность, что U попадет в промежуток $\pm$ 2 стандартных отклонения от среднего;

Обратим внимание, что стандартное отклонение U равно 3, а значит, нас интересует вероятность $P((-1) - 2 \times 3 \le U \le (-1) + 2 \times 3) = P(-7 \le U \le 5)$. Теперь можно либо вспомнить про «правило трех сигм»: мы знаем, что для нормального распределения в границах ± 2 стандартных отклонений от среднего лежит примерно 95% плотности. Либо можно вычислить эту вероятность напрямую:

$$P(-7 \le U \le 5) = P(-2 \le Z \le 2) =$$

$$F(2) - (1 - F(2)) = 2F(2) - 1 = 2 \times 0.9772 - 1 = 0.9544$$

(c) Найдите квантиль u_p уровня $p = 0.3$.

Для начала найдем квантиль станлартной нормальной величины z_p уровня $p = 0.3$. Мы знаем, что это отрицательное число. Найдем симметричный ему квантиль в положительной части распределения:

$$z_{(1-p)} = z_{0.7} = \sim 0.52$$

и обратно «отразим» его в отрицательную часть распределения, умножив на (-1): $z_{0.3} = -0.52$.
Теперь рассчитаем u_p по формуле: $u_{0.3} = z_{0.3} \times 3 + (-1) = -2.57$

Задача 2. Пусть S – число успехов в $n = 10$ испытаниях Бернулли при $p = 0.4$. Вычислите точную вероятность события $5 \leq S \leq 7$. Затем вычислите приближенную вероятность того же события, используя теорему Муавра–Лапласа (*далее – теорема М-Л*). Сравните полученные результаты. Достаточно ли число испытаний n, чтобы пользоваться приближенными формулами?

(a) Рассчитаем точное значение $P(5 \leq S \leq 7) = P(S = 5) + P(S = 6) + P(S = 7)$ по формуле Бернулли.

$$P(S = 5) = C_{10}^5 \times 0.4^5 \times 0.6^5 = 252 \times 0.01 \times 0.0778 = 0.2$$

Остальные вероятности вычислите самостоятельно. Итоговая вероятность равна: $P(5 \leq S \leq 7) = 0.3546$

(b) Рассчитаем приближенное значение, используя теорему М-Л. Для этого рассчитаем $E(S) = 10 \times 0.4 = 4$ и $D(S) = 10 \times 0.4 \times 0.6 = 2.4$.

$$P(5 \leq S \leq 7) = P(\frac{5-4}{\sqrt{2.4}} \leq Z \leq \frac{7-4}{\sqrt{2.4}}) = P(0.65 \leq Z \leq 1.94) = 0.2382$$

(c) Используя приближение через нормальное распределение, мы недооценили точную вероятность на $0.3546 - 0.2382 = 0.1164$. Это довольно большое значение: $n = 10$ недостаточно, чтобы использовать теорему М-Л.

Задача 3. Всероссийский центр изучения общественного мнения в 2013 г. проводил опрос на тему «Российская Конституция: первые 20 лет». Согласно полученным данным, только 14% россиян ответили, что хорошо знают основные положения Конституции и читали её. Используя теорему Муавра–Лапласа, найдите вероятность того, что в выборке объема 1600 человек окажется от 250 до 350 человек, которые действительно знают основной закон государства.

Случайная величина X – количество людей в выборке, которые читали Конституцию, имеет биномиальное распределение с параметрами $X \sim Bin(n = 1600, p = 0.14)$. Рассчитаем ее математическое ожидание $E(X) = 1600 \times 0.14 = 224$ и стандартное отклонение $\sqrt{D(X)} = \sqrt{1600 \times 0.14 \times (1 - 0.14)} = \sqrt{192.64} = 13.88$.

Теперь мы можем рассчитать нужную вероятность, приближая ее через нормальное распределение (так как n достаточно велико):

$$P(250 \leq X \leq 350) = P(\frac{250 - 224}{13.88} \leq Z \leq 350 - 22413.88) = 1 - 0.9695 = 0.0305$$