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 $N(2, \sigma^2 = 3)$, а Z имеет стандартное нормальное распределение. Для случайной величины $U = 2X + 2Z - 5$:

(a) Укажите, какое распределение будет иметь эта случайная величина и каковы его параметры;
(b) Рассчитайте вероятность, что U попадет в промежуток $\pm$ 2 стандартных отклонения от среднего;
(c) Найдите квантиль u_p уровня $p = 0.2$.
(d) 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:

$$E(U) = E(2X + 2Z - 5) = 2E(X) + 2E(Z) - 5 = 2 \times 2 + 2 \times 0 - 5 = -1$$

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

$$D(U) = E(2X + 2Z - 5) = 2^2 \times D(X) + 2^2 \times D(Z) = 4 \times 3 + 4 \times 1 = 16$$

Получаем: $U \sim N(-1, \sigma^2 = 16)$ или $U \sim N(-1, \sigma = 4)$

(e) 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 $P((-1) - 2 \times 4 \leq U \leq (-1) + 2 \times 4) = P(-9 \leq U \leq 7)$. 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:

$$P(-9 \leq U \leq 7) = P(-2 \leq Z \leq 2) =$$
$$F(2) - (1 - F(2)) = 2F(2) - 1 = 2 \times 0.9772 - 1 = 0.9544$$

(f) Найдите квантиль u_p уровня $p = 0.2$.

Для начала найдем квантиль стандартной нормальной случайной величины z_p

уровня $p = 0.2$. Мы знаем, что это отрицательное число. Найдем симметричный ему квантиль в положительной части распределения:

$$z_{(1-p)} = z_{0.8} =\sim 0.84$$

и обратно «отразим» его в отрицательную часть распределения, умножив на (-1): $z_{0.2} = -0.84$.

Теперь рассчитаем u_p по формуле: $u_{0.2} = z_{0.2} \times 4 + (-1) = -4.37$

Задача 2. Пусть S – число успехов в $n = 10$ испытаниях Бернулли при $p = 0.6$. 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.

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

Остальные вероятности вычислите самостоятельно. Итоговая вероятность равна: $P(5 \leq S \leq 7) = 0.6665$

(b) Рассчитаем приближенное значение, используя теорему М-Л. Для этого рассчитаем $E(S) = 10 \times 0.6 = 6$ и $D(S) = 10 \times 0.6 \times 0.4 = 2.4$.

$$P(5 \leq S \leq 7) = P(\frac{5-6}{\sqrt{2.4}} \leq Z \leq \frac{7-6}{\sqrt{2.4}}) = P(-0.65 \leq Z \leq 0.65) = 0.4814$$

(c) Используя приближение через нормальное распределение, мы недооценили точную вероятность на $0.6665 - 0.4814 = 0.1851$. Это довольно большое значение: $n = 10$ недостаточно, чтобы использовать теорему М-Л.

Задача 3. По данным Росстата на 2011 год вероятность того, что школьник Российской Федерации получает горячее питание, составляет 0.84. Найдите с помощью теоремы Муавра–Лапласа вероятность того, что из 1500 случайно выбранных российских школьников от 200 до 300 получают горячее питание.

Случайная величина X – количество школьников, получающих горячее питание, имеет биномиальное распределение с параметрами $X \sim Bin(n = 1500, p = 0.84)$. Рассчитаем ее математическое ожидание $E(X) = 1500 \times 0.84 = 1260$ и стандартное отклонение $\sqrt{D(X)} = \sqrt{1500 \times 0.84 \times (1 - 0.84)} = \sqrt{201.6} = 14.2$.

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

$$P(200 \leq X \leq 300) = P(\frac{300-1260}{14.2} \leq Z \leq \frac{200-1260}{14.2}) = 0$$