Школа лингвистики, 2024-25 уч. год

Теория вероятностей

Дискретные случайные величины: биномиальное распределение, геометрическое распределение, распределение Пуассона (18.02.2025)

И. В. Щуров, Д. А. Филимонов, Р. Я. Будылин

Задача 1. Монетка имеет смещённый центр тяжести и выпадает орлом вверх с вероятностью $p \in (0, 1)$. Монетку подкинули n раз. Случайная величина ξ — число выпавших орлов. Как называется такое распределение?

- (a) Найти ряд распределения
- ξ
- для
- $n = 4$
- ,
- $n = 5$
- ,
- $n = 6$
- для честной монеты.

0	1	2	3	4
1/16	4/16	6/16	4/16	1/16

0	1	2	3	4	5
1/32	5/32	10/32	10/32	5/32	1/32

0	1	2	3	4	5	6
1/64	6/64	15/64	20/64	15/64	6/64	1/64

- (b) Написать ряд распределения
- $\text{Bin}(3, \frac{1}{3})$
- , т. е. для
- ξ
- при
- $n = 3$
- и
- $p = \frac{1}{3}$
- .

0	1	2	3
8/27	12/27	6/27	1/27

- (c) Найти
- $E(\xi)$
- ,
- $D(\xi)$
- и
- $\sigma(\xi)$
- для
- $n = 3$
- ,
- $n = 4$
- ,
- $n = 5$
- для честной монеты.

► $E\xi = 1.5$; $D\xi = 0.75$; $\sigma(\xi) = \sqrt{0.75}$

$E\xi = 2$; $D\xi = 1$; $\sigma(\xi) = 1$

$E\xi = 2.5$; $D\xi = 1.25$; $\sigma(\xi) = \sqrt{1.25}$ ◄

- (d) Найти
- $E(\xi)$
- ,
- $D(\xi)$
- и
- $\sigma(\xi)$
- для
- $n = 100$
- и
- $p = \frac{1}{4}$
- .

► $E\xi = 25$; $D\xi = 75/4 = 18.75$; $\sigma(\xi) = 5\sqrt{3}/2$ ◄

Задача 2. Пусть мы бросаем игральный кубик 100 раз, и ξ — число выпавших шестёрок. Найти математическое ожидание и дисперсию ξ .

► $E(\xi) = 100/6 = 50/3$; $D(\xi) = 125/9$ ◄

Задача 3. У меня в кармане две монетки: одна нормальная, а другая — с двумя орлами. Я достаю случайную монетку и подбрасываю её. Если выпала решка, я прекращаю подбрасывание, а если орёл, возвращаю монетку в карман и продолжаю процесс. Случайная величина ξ — число выпавших орлов до первого выпадения решки.

- (a) Как называется такое распределение? Найти значения параметров.

► Геометрическое, $p = 1/4$, $q = 3/4$ ◄

- (b) Какие значения может принимать
- ξ
- ?

► $\mathbb{N} \cup \{0\}$ ◄

- (c) Чему равняется
- $P(\xi = 0)$
- ?
- $P(\xi = 1)$
- ?
- $P(\xi = k)$
- для произвольного
- k
- ?

► $1/4$, $3/16$, $3^k/4^{k+1}$ ◄

- (d) Найти математическое ожидание этой случайной величины.

► 3 ◄

- (e) Чему равняется
- $P(\xi < 4)$
- ?

► $175/256$ ◄

- (f) Чему равняется
- $P(\xi > 3)$
- ?

► $81/256$ ◄

- (g) Что поменяется, если мы, наоборот, будем подбрасывать монетку до выпадения первого орла?
 ► p и q поменяются местами ◀
- (h) (*) Чему равняется $P(\xi > 5 | \xi > 2)$ (то есть вероятность того, что ξ оказалось больше 5 при условии, что $\xi > 2$)?
 ► $q^3 = 27/64$ ◀

Задача 4. Частота русской буквы «о» — примерно 0.1.

- (a) Сколько в среднем букв «о» встретится в случайно взятом тексте на русском языке длиной 10000 знаков?
 ► 1000 ◀
- (b) Найти дисперсию числа букв «о».
 ► 900 ◀
- (c) Каково стандартное отклонение числа букв «о» в таком тексте?
 ► 30 ◀

Задача 5. Найдите среднее число бросаний игрального кубика, если мы останавливаемся, получив шестёрку.

► 6 ◀

Задача 6. Пусть вероятность аварии на участке дороги длиной в 1 км в среднем равна 0.001. Сколько в среднем можно проехать по такой дороге без аварий?

► 999 км ◀

Задача 7. Найдите среднее число бросаний монеты, если мы останавливаемся, получив двух орлов (не обязательно подряд). А если останавливаемся, получив 10 орлов?

► 4, 20 ◀

Задача 8. Найдите дисперсию числа бросаний монеты до выпадения 2 орлов (не обязательно подряд). А для 10 орлов?

► 4, 20 ◀

Задача 9. В ходе исследования вопроса об авторстве «Федералиста» (сборник статей в поддержку ратификации Конституции США) использовались статистические методы [Mosteller and Wallace (1984, 2.3)]: в частности, подсчёт частотности вхождения тех или иных слов. Для сравнения был взят длинный текст, принадлежащий перу Александра Гамильтона. Его разбили на большое количество фрагментов примерно одинаковой длины. Выбрали случайный фрагмент и посчитали, сколько раз в нём встречается слово «апу». Оказалось, что полученная случайная величина распределена по Пуассону¹ с параметром $\lambda = 0.67$.

- (a) С какой вероятностью в выбранном фрагменте слово «апу» встречалось три раза?
 ► $\frac{0.67^3 e^{-0.67}}{3!} = 0.02565$ ◀
- (b) Сколько в среднем раз встречалось слово «апу» в фрагменте?
 ► 0.67 ◀
- (c) С какой вероятностью в фрагменте встретилось меньше трёх слов «апу»?
 ► 0.9694 ◀

¹Вообще говоря, частотность слов далеко не всегда распределена по Пуассону, но в данном случае это оказалось так.

Задача 10. На заводе для приготовления булочек с изюмом на 10 тонн теста добавляют миллион изюминок. Какова вероятность, что в одной булочке весом 100 грамм нет ни одной изюминки?

► $e^{-10} = 0.0000454$ ◀

Задача 11. Примерно четвертая часть носителей русского языка считает, что слово «виски» среднего рода. Вы решили проверить это и сделали форму, где респонденту предлагается вписать окончание прилагательного. Далее программа выбирает случайного человека из 1000 и предлагает ему ответить на вопрос. Программа собрала 10 ответов, однако так получилось, что один и тот же человек мог ответить на вопрос более одного раза. Предположим, что гипотеза верна; какое распределение имеет количество людей, считающих, что слово «виски» среднего рода? А какое у этой случайной величины математическое ожидание и стандартное отклонение?

► *Биномиальное.* $E(\xi) = 10/4 = 2.5$, $\sigma(\xi) = \sqrt{15/8} = \sqrt{1.875} = 1.3693$ ◀

Задача 12. В условиях предыдущей задачи исправим программу так, чтобы каждый человек мог ответить только один раз. Как изменится распределение количества людей, считающих, что слово «виски» среднего рода? А среднее?

► *Станет гипергеометрическим.* $E(\xi) = 2.5$, не изменится. ◀

Дополнительные задачи

Задача 13. В условиях предыдущей задачи найти стандартное отклонение. Записать его в общем виде для вероятности успеха p , количества человек N и количества ответов n .

► $D\xi = npq \frac{N-n}{N-1} = 10 \cdot \frac{1}{4} \cdot \frac{3}{4} \cdot \frac{1000-10}{1000-1} = 1.8581$; $\sigma(\xi) = 1.3631$ ◀

Задача 14. За час рыбак ловит в среднем 6 рыб. Какова вероятность, что он за 10 минут поймает больше 2 рыб?

► 0.08 ◀