

Математика для Data Science. Теория вероятностей.

Условия задач

Содержание

Арифметика случайных величин и нормальное распределение	2
Задача 2	2
Нормальное распределение	2
Задача 1	2
Задача 2	2
Задача 3	2
Задача 4	2
Статистический тест	2
Задача 1	2
Задача 2	3
ЗБЧ и ЦПТ	3
Задача 1	3
Задача 2	3

Замечание. Вот этим цветом отмечены ссылки на страницы внутри этого файла.

Арифметика случайных величин и нормальное распределение

Задача 2

Даны совместно независимые случайные величины X_1, X_2, X_3 , такие что

- $E[X_1] = 0, Var(X_1) = 1$,
- $E[X_2] = 11, Var(X_2) = 3$,
- $E[X_3] = 8, Var(X_3) = 4$.

Найдите математическое ожидание и дисперсию случайной величины $\frac{2X_1+4X_2-X_3}{6} - 4$.

Нормальное распределение

Задача 1

Независимые случайные величины X и Y имеют распределения $N(4, 5)$ и $N(3, 9)$ соответственно. Найдите распределение случайной величины $X + \frac{Y}{3} - 4$.

Задача 2

Независимые случайные величины X, Y и Z имеют распределения $N(1, 2), N(3, 4)$ и $N(5, 3)$ соответственно. Найдите распределение случайной величины $\frac{X+Y+Z}{3} + 2$.

Задача 3

Все совместно независимые случайные величины X_1, \dots, X_n имеют одинаковое распределение $N(\mu, \sigma^2)$. Найдите распределение случайной величины $X_1 + \dots + X_n$.

Задача 4

Все совместно независимые случайные величины X_1, \dots, X_n имеют одинаковое распределение $N(\mu, \sigma^2)$. Найдите распределение случайной величины $\frac{X_1 + \dots + X_n}{n}$.

Статистический тест

Задача 1

Докажите, что на прошлом шаге мы правильно нашли распределение статистики T при условии, что верна гипотеза H_0 . А именно, докажите такую последовательность утверждений:

- $x_1 + \dots + x_n$ имеет распределение $N(n\mu_0, n\sigma^2)$
- $\frac{x_1 + \dots + x_n}{n}$ имеет распределение $N\left(\mu_0, \frac{\sigma^2}{n}\right)$
- $\frac{x_1 + \dots + x_n}{n} - \mu_0$ имеет распределение $N\left(0, \frac{\sigma^2}{n}\right)$
- $T(x_1, \dots, x_n) := \frac{\frac{x_1 + \dots + x_n}{n} - \mu_0}{\sigma/\sqrt{n}}$ имеет распределение $N(0, 1)$

Задача 2

Найдите распределение T при условии, что выполнена H_1 . Это нужно нам для нахождения вероятности ошибки второго рода

Ясно, что из-за того, что μ_1 и μ_0 не указаны, нельзя указать вероятность ошибки второго рода. Обсудите с преподавателем, как будет вести себя β (вероятность ошибки второго рода) с увеличением $\mu_1 - \mu_0$.

Комментарий. Обратите внимание, что вероятность ошибки первого рода не зависит от μ_1 , ведь вероятность ошибки первого рода всегда равна $\alpha = 0.05$. А вот вероятность ошибки второго рода зависит от того, насколько μ_1 далеко от μ_0 , то есть от $\mu_1 - \mu_0$. Это логично: чем дальше μ_1 от μ_0 , тем легче должно быть отличить H_1 от H_0 . Можно сказать, что чем больше $\mu_1 - \mu_0$, тем мощнее наш критерий (как мы помним, мощность критерия это $(1 - \beta)$, где β это вероятность ошибки второго рода).

ЗБЧ и ЦПТ

Задача 1

Рассмотрим несимметричную монетку — она имеет распределение Бернулли с параметром $p = 0.25$. То есть $P(\xi'_i = 0) = 0.75$ и $P(\xi'_i = 1) = 0.25$. Воспользуйтесь ЦПТ и найдите параметры нормального распределения, к которому близко распределение величины $\eta'_{100} := \sum_{i=1}^{100} \xi'_i$.

Задача 2

Докажите, что стандартное отклонение случайной величины, имеющей распределение Бернулли, не превосходит 0.5.

Замечание. Это задача скорее на матан, чем тервер.