Мотивационное письмо

Куценко Антон

Коротко обо мне

- **Группа**: БЭК 1812
- **Оценка по курсу ТВиС**: 9 за модуль Теории вероятности, сейчас накопленная оценка - 5,94 без учета экзамена
- **Программирование**: Уверенный пользователь Python: можно тыкнуть и посмотреть код, в котором тестируются гипотезы из нашего курса, строятся прикольные (или не очень) графики и находится p-value.

  Летом планирую освоить R и SQL, так что к учебному году - в полной боевой готовности.
- **LaTeX**: Владею на уровне написания формул и текста, создания таблиц, вставки картинок и т.п. В общем, имеются базовые навыки.
- **Контакты**:
  - Телефон: 8(927)904-04-20
  - Telegram: @ontenkutsenko
  - Почта: askutsenko@edu.hse.ru

Почему я хочу стать ассистентом?

Кто вообще не хочет стать ассистентом по курсу ТВиС???

Если честно, то это был лучший курс за два года в Вышке. Во-первых, преподаватели и ассистенты очень доброжелательны к студентам, что мотивирует учиться и позволяет получать удовольствие от процесса. Во-вторых, организация курса (отдельное спасибо за организацию курса на карантине) и полученные знания - все на высочайшем уровне! Поэтому хотелось бы в следующем году присоединиться к команде и заражать энтузиазмом и любовью к предмету будущих второкурсников.

Также у меня уже есть опыт ассистентства на курсе «История экономики», так что я вполне осознаю требования к ассистентам и готов к работе.

Задачка

Один очень несчастный студент пришел к мудрому старцу с вопросом. Студент говорил: "Я несчастен, потому что всегда опаздываю на пары и пропускаю важный материал. Но проблема не во мне, а в том, что на моей станции поезда метро приходит слишком редко. Помоги мне найти станцию, отправляясь с которой я не буду опаздывать". Старец любил иносказательность, и поэтому сказал: "Распределение времени ожидания прибытия поездов на станции, которая тебе нужна будет -lnX, где $X \sim U(0{,}1)$ - равномерно распределенная на [0, 1] случайная величина". Несчастный студент отправился на поиски. Давайте ему поможем.

1. Найдите функцию распределения и функцию плотности времени ожидания поезда на нужной студенту станции. Посчитайте вероятность того, что студент опоздает на пару, отправляясь с этой станции, если мы знаем, что он опаздывает, когда ждет поезд больше 2 минут.

2. Студент провел 5 независимых наблюдений прибытия поезда на станции метро "Случайная"и получил следующие значения: 1, 0.5, 0.7, 3, 1.2. Проверьте с помощью критерия Колмогорова на 10 % уровне значимости гипотезу о том, что студент попал на станцию, о которой говорил старец.

3. Студент получил еще больше наблюдений, которые попали в следующие интервалы:

Интервал	0-0.2	0.2-0.5	0.5 - 1	1 - 2	2 - ∞
Количество наблюдений	20	25	21	20	14

Проверьте с помощью критерия Пирсона на уровне значимости 5 % гипотезу о том, что студент попал на станцию, о которой говорил старец. Найдите p-value.

4. Какова вероятность того, что студент, проведя три независимых наблюдения на станции, описанной старцем, получит значение более 4 минут.

Решение

1. Для $X \sim U(0{,}1)$ функция плотности имеет вид:

$$f_X(x) = \begin{cases} 1, & \text{если } x \in [0{,}1] \\ 0, & \text{иначе} \end{cases}$$

$F_X(x) = P(X \leqslant x)$ - функция распределения С.В. X по определению

Тогда С.В. Y будет иметь распределение:

$$Y \stackrel{d}{=} -lnX$$

Найдем функцию распределения для С.В. Y:

$F_Y(x) = P(-lnX \leqslant x) = P(lnX > -x) = P(X > e^{-x}) = 1 - P(X \leqslant e^{-x}) = 1 - F_X(e^{-x}) = 1 - \int_0^{e^{-x}} 1 \quad dt = 1 - e^{-x}$, где $x \in [0, \infty)$ - носитель С.В. Y

Функция распределения С.В. Y:

$$F_Y(x) = \begin{cases} 0, & x < 0 \\ 1 - e^{-x}, & \text{если } x \in [0{,}\infty) \end{cases}$$

Функция плотности С.В. Y:

$$f_Y(x) = \begin{cases} 0, & x < 0 \\ e^{-x}, & \text{если } x \in [0,\infty) \end{cases}$$

Мы видим, что $Y \sim EXP(1)$

Посчитаем вероятность, что студент опоздает на пару:

$P(Y > 2) = 1 - P(Y \leqslant 2) = e^{-2} \approx 0.135$

2. Посчитаем статистику теста Колмогорова, сравнив значения теоретической и эмпирической функции плотности.

x	0.5	0.7	1	1.2	3
$\hat{F}_Y(x)$	0.2	0.4	0.6	0.8	1
$F_Y(x)$	0.393	0.503	0.632	0.699	0.95

Супремум будет достигаться в точке $x = 0.5 - \epsilon$:

$\hat{F}_Y(0.5-\epsilon) = 0$, а $F_Y(0.5-\epsilon) \approx 0.393 \Rightarrow$ значение статистики Колмогрова будет равно: $D_n = 0.393$

Критическое значение: $D_5^{0.1} = 0.509$

$0.393 < 0.509 \Rightarrow H_0$ не отвергается. Это действительно может быть станция, описанная старцем.

3. Для решения посчитаем истинные вероятности попадания в заданные интервалы:

Интервал	0-0.2	0.2-0.5	0.5 - 1	1 - 2	2 - ∞
Количество наблюдений	20	25	21	20	14
Вероятности попадания	0.181	0.212	0.239	0.233	0.135

Количество наблюдений в выборке: n = 100

Рассчитаем значение тестовой статистики:

$$T = \frac{(20 - 0.181 * 100)^2}{(0.181 * 100)} + \frac{(25 - 0.212 * 100)^2}{(0.212 * 100)} + \frac{(21 - 0.239 * 100)^2}{(0.239 * 100)} + \frac{(14 - 0.135 * 100)^2}{(0.135 * 100)} \approx 1.25$$

Критическое значение: $\chi^2_{5-1,1-0.05} \approx 9.48 \Rightarrow H_0$ не отвергается

$p - value \approx 0.87$

4. Поскольку $Y \sim EXP(1)$, то сумма трех наблюдений будет иметь Гамма-распределение, а точнее его частный случай - распредление Эрланга с параметрами: k = 3, $\lambda = 1$

Назовем наше новое распредление Z, тогда вероятность будет представима в следующем виде:
P(Z > 4) = $\int_4^{+\infty} 0.5x^2 * e^{-x} dx \approx 0.238$

***** Значения для заданий 3 и 4 можно рассчитать в Python (import scipy.stats as sts):

3. sts.chi2(df=4).ppf(0.95) # критическое значение

sts.chi2(df=4).sf(1.25) #p-value

4. sts.erlang(3, scale=1).sf(4) # P(Z > 4)