ПРИКЛАДНАЯ СТАТИСТИКА

Первая домашка

26 февраля 2023 г.

— Куда ты теперь, Док? Назад в будущее?

— Нет, там я уже был.

«Назад в будущее 3»

Решение работы нужно сдать либо в виде затеханого pdf-файла либо в виде рукописной pdf-ки. Если вы решаете задачи и оформляете их решение на листочке, пишите максимально разборчивым почерком. Если у вас плохой почерк, техайте решения.

Свойства оценок

Задача 1 (10 баллов). У Дианы есть выборка $X_1, \ldots, X_n \sim iid N(0, \sigma^2)$. Она собирается использовать для параметра σ нестандартную оценку $\hat{\sigma} = \frac{1}{n} \sum_{i=1}^{n} |x_i|$. Вручную:

1. Найдите плотность распределения для случайной величины $|X|$
2. Проверьте, является ли оценка $\hat{\sigma}$ смещённой, найдите её $bias$, если оценка смещена, то скорректируйте её
3. Найдите дисперсию скорректированной оценки
4. Определите, является ли оценка состоятельной
5. Найдите для несмещённой оценки MSE

Задача 2 (5 баллов). Мы знаем, что для выборки $X_1, \ldots, X_n \sim iidN(\mu, \sigma^2)$ хорошими оценками параметров будут

$$\hat{\mu} = \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$
$$\hat{\sigma}^2 = s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2.$$

У этих оценок есть ещё одна полезная особенность. Случайные величины $\hat{\mu}$ и $\hat{\sigma}^2$ независимы. Докажите это.

Hint: если ковариация между двумя случайными величинами нулевая, они будут независимы только, если эти две случайные величины нормально распределены. Воспользуйтесь этим. Для начала найдите $Cov(x_1, x_1 - \bar{x})$.

Асимптотические доверительные интервалы

Задача 3 (10 баллов). Аналитесса Аня наугад поймала 400 покемонов-девочек и 100 покемонов-мальчиков. Среди девочек 250 оказались ядовитыми, среди мальчиков 60.

1. Методом моментов найдите точечную оценку для доли ядовитых покемонов среди мальчиков.
2. Постройте 95% доверительный интервал для ядовитых покемонов-мальчиков. Какая у него получилась длина?
3. Постройте 95% доверительный интервал для ядовитых покемонов-девочек. Какая у него получилась длина?
4. Постройте 95% доверительный интервал для разницы долей ядовитых покемонов среди девочек и мальчиков. Какой длины он оказался? Попадает ли ноль в этот интервал? Что это означает?
5. Нужно ли предположение о нормальности X_i и Y_i для решения предыдущих пунктов? А какие предположения нужны? Выпишите их.

Задача 4 (5 баллов). Аналитесса Аня пытается выяснить какая доля жителей Москвы любит кофе. Она хочет построить оценку так, чтобы ширина 95% доверительного интервала для неё оказалась не больше 0.1.

1. Предположим, что Аня ничего не знает о доле. Сколько наблюдений Ане необходисо собрать для достижения такой точности?
2. Аналитик Богдан проговорился Ане, что в прошлом году уже опрашивал так людей. У него оценка доли получилась в районе 0.7. Аня верит Богдану как себе. Сколько наблюдений ей надо собрать для достижения той же точности с учётом этой информации?

ЦПТ и ЗБЧ

Задача 5 (5 баллов). Для идеальной картинки режиссёру рекламы грузовиков Volvo требуется 60 дублей. В каждом Жан Клод ван Дамм садится на шпагат. Пусть $U_1, U_2, \ldots, U_{60}$ – независимо распределённые случайные величины, сколько сантиметров актёру не хватает до идеальной параллели, $U_i \sim U(0, 1)$, и $X = U_1 + \ldots + U_{60}$.

1. К какому распределению очень близко распределение случайной величины X? Укажите его параметры.
2. Найдите, чему приблизительно равна $(X > 20)$.

Задача 6 (5 баллов). Пусть $X_1, X_2, \ldots$ независимо распределённые случайные величины, число дублей в каждом съёмочном дне, которые не идеальны по вине водителя левого грузовика Volvo, $E(X_i) = 2$, $Y_1, Y_2, \ldots$ независимо распределённые случайные величины, сколько дублей испортил водитель второго грузовика за каждый съёмочный день, $E(Y_i) = 4$. К какому числу сходится случайная величина $\frac{X_1+X_2+\ldots+X_n}{Y_1+Y_2+\ldots+Y_n}$, если режиссёр-перфекционист не ограничен во времени и деньгах, то есть при $n \to \infty$?

Задача 7 (5 баллов). В рекламе задействовано два грузовика. Расход топлива за i-ый день (в л/100 км) у левого – случайная величина X_i, у правого – случайная величина Y_i, X_i и Y_i независимы $\forall i$. Известно, что $\mathbb{E}(X_i) = 20$, $\mathrm{Var}(X_i) = 1$, $\mathbb{E}(Y_i) = 18$, $\mathrm{Var}(Y_i) = 4$.

За один съёмочный день грузовики проезжают по 100 км. Поскольку сначала надо отрепетировать трюк без Жан Клода ван Дамма, а потом с ним, грузовики будут нужны в 40 съёмочных днях. Какова вероятность того, что за всё время репетиций и съёмок понадобится более 1550 литров топлива?

Задача 8 (5 баллов). После выхода рекламы компания Volvo провела опрос 10000 дальнобойщиков. У каждого спросили, считает ли он модель Volvo XC40 более стильной, чем модель Volvo V90. Дальнобойщики отвечали «да» или «нет» равновероятно. Оцените вероятность того, что число положительных ответов отличалось от 5000 меньше, чем на 100.

Бонусные задачи из инсайдов к домашке

Задача 9 (5 баллов). Пусть $X \sim N(0, \sigma^2)$. Пусть $\Phi(x)$ — функция распределения для $N(0, 1)$. Нужно найти математическое ожидание для случайной величины $\Phi(X)$.

Найдите предел для функции распределения случайной величины $\Phi(X)$ при $\sigma \to \infty$. К чему сходится по распределению случайная величина $\Phi(X)$?

Задача 10 (5 баллов). Пусть X_1, ..., X_{100} независимы и равномерны на $[0;1]$. Пусть $L = \max\{X_1, X_2, \ldots, X_{80}\}$ а $R = \max\{X_{81}, X_{82}, \ldots, X_{100}\}$ и $M = \max\{X_1, \ldots, X_{100}\}$.

1. Найдите вероятности $\mathbb{P}(L > R|L)$ и $\mathbb{P}(L > R|R)$ и $\mathbb{P}(L > R|M)$, $\mathbb{P}(L > R|L, M)$.
2. Найдите ожидания $\mathbb{E}(X_1|L)$, $\mathbb{E}(X_1| \min\{X_1, \ldots, X_{100}\})$.
3. Найдите ожидание $\mathbb{E}(\min\{X_1, \ldots, X_{100}\}| \max\{X_1, \ldots, X_{100}\})$.
4. Найдите ожидание $\mathbb{E}(\min\{X_1, \ldots, X_{100}\}|X_1)$.
5. Нарисуйте условную функцию распределения $\mathbb{P}(X_1 \leqslant t|L)$.

Задача 11 (5 баллов). Пусть μ для каждой случайной величины – это математичесое ожидание, а σ^2 – дисперсия.

Фил взял n случайных величин 😎 с очень маленьким близким к нулю μ, сложил их и получил лучайную величину 🍏. Дальше Фил устремил $n \to \infty$ и получил случайную величину 🍍, для которой $\mu = \sigma^2$.

После Фил взял n случайных величин 🍍, сложил их, устремил $n \to \infty$ и получил случайную величину 🍒. Дальше он вычел из 🍒 математическое ожидание, поделил на стандартное отклонение и получил 🍓.

Поделив одну 🍓 на другую 🍓, Фил получил случайную величину 🍋. Ещё он возвёл 🍓 в квадрат и получил 🍔. Потом он взял функцию распределения случайной величины 🍒 от случайной величины 🍒 и получил случайную величину 🍕.

Найдите 😎, 🍏, 🍍, 🍒, 🍓, 🍋, 🍔, 🍕.