Лабораторная работа №3
Тема: Последовательности испытаний. Формула Бернулли. Наивероятнейшее число событий. Точечные оценки неизвестных параметров. Метод максимального правдоподобия (MLE).
Цель работы:
1. Научиться вычислять вероятности в схеме Бернулли и находить наивероятнейшее число событий.
2. Понять суть метода максимального правдоподобия (ММП) для нахождения точечных оценок.
3. Реализовать программно расчеты по формуле Бернулли и визуализировать распределение вероятностей.
4. Провести эксперимент по оценке параметров распределения методом ММП на сгенерированных данных.

Инструментарий: Python (библиотеки numpy, matplotlib, scipy.stats, math).
Краткая теоретическая справка
1. Схема Бернулли
Если проводится серия из n независимых испытаний, в каждом из которых событие A (успех) может наступить с вероятностью p или не наступить с вероятностью $q = 1 - p$, то такая схема называется **схемой Бернулли**.
Вероятность того, что в n испытаниях событие A наступит ровно k раз, вычисляется по **формуле Бернулли**:
$$P_n(k) = C_n^k p^k q^{n-k}$$
где
$C_n^k = \frac{n!}{k!(n-k)!}$ — число сочетаний.
2. Наивероятнейшее число событий
Число k_0 (число успехов) называется наивероятнейшим, если вероятность его наступления $P_n(k_0)$ не меньше вероятности наступления любого другого числа успехов.
Оно определяется из двойного неравенства:
$$np - q \leq k_0 \leq np + p$$
Если $np - q$ — целое число, то наивероятнейших чисел два: k_0 и $k_0 + 1$.
- Если $np - q$ — дробное, то k_0 — единственное целое число, попадающее в этот интервал.

3. Метод максимального правдоподобия (MLE)
Точечная оценка — это число, вычисляемое на основе выборки, которое принимается за приближенное значение неизвестного параметра генеральной совокупности.
Метод максимального правдоподобия (ММП) заключается в поиске такого значения неизвестного параметра θ, при котором вероятность получить имеющуюся выборку максимальна.
Для этого составляется **функция правдоподобия** $L(\theta)$. Если $x_1, x_2, \ldots, x_n$ — независимые наблюдения, то:
$$L(\theta) = f(x_1, \theta) \cdot f(x_2, \theta) \cdot \ldots \cdot f(x_n, \theta)$$
где $f(x, \theta)$ — плотность распределения (или вероятность в дискретном случае).
Для удобства вычислений обычно ищут максимум логарифма функции правдоподобия ($\ln L$), так как логарифм — монотонная функция, и точка максимума у них совпадает.
Алгоритм:
1. Составить $L(\theta)$.
2. Найти логарифмическую функцию $\mathcal{L}(\theta) = \ln\ L(\theta)$.
3. Найти производную по параметру: $\frac{d\mathcal{L}}{d\theta}$.
4. Приравнять производную к нулю и найти θ.

Пример (Оценка вероятности успеха p в схеме Бернулли):

Проведено n испытаний, получено k успехов.
Функция правдоподобия: $L(p) = p^k(1-p)^{n-k}$ (множитель C_n^k не зависит от p, его можно опустить при максимизации).
Логарифм: $\ln\, L = k\ln\, p + (n-k)\ln\,(1-p)$
Производная: $\left(\ln\, L\right)'_p = \frac{k}{p} - \frac{n-k}{1-p}$.
Приравниваем к 0: $\frac{k}{p} = \frac{n-k}{1-p} \Rightarrow k(1-p) = p(n-k) \Rightarrow k - kp = np - kp \Rightarrow k = np.$
Оценка: $\hat{p} = \frac{k}{n}$. (Что логично: частота события есть оценка его вероятности).

Ход работы

Часть А: Аналитические задачи (решаются на бумаге)

Задача 1 (Формула Бернулли):

Стрелок делает 5 выстрелов по мишени. Вероятность попадания при одном выстреле равна 0.7. Найдите вероятность того, что:
а) Стрелок попадет ровно 3 раза.
б) Стрелок попадет менее 2 раз.

Задача 2 (Наивероятнейшее число):

На заводе производится партия деталей. Вероятность брака составляет 0.05 (5%). В партии 100 деталей. Найдите наивероятнейшее число бракованных деталей в этой партии и вычислите вероятность этого числа.

Задача 3 (Метод максимального правдоподобия) – усложненный вариант:

Пусть случайная величина X имеет показательное (экспоненциальное) распределение с плотностью $f(x) = \lambda e^{-\lambda x}$ (при $x \geq 0$).
Имеется выборка из n наблюдений: $x_1, x_2, \ldots, x_n$.
Используя метод максимального правдоподобия, найдите формулу для оценки параметра λ.
Подсказка: Составьте функцию $L(\lambda)$, прологарифмируйте, возьмите производную по λ и выразите λ через среднее арифметическое выборки $\bar{x}$.

Задача 3 (Метод максимального правдоподобия) – облегченный вариант:

В лекции приведены оценки максимального правдоподобия для параметров нормального распределения $N(\mu, \sigma^2)$: $\hat{\mu} = \bar{x}$, $\sigma^2 = S^2$. Объясните своими словами, что означают эти формулы. Почему выборочное среднее является хорошей оценкой для истинного среднего?
Ответ пишем в тетради

Часть Б: Практическое задание (выполняется в Python / Jupyter Notebook)

Задание 1. Моделирование схемы Бернулли и визуализация

1. Импортируйте необходимые библиотеки (scipy.stats, matplotlib.pyplot, numpy).
2. Задайте параметры: $n = 50$ (количество испытаний), $p = 0.4$ (вероятность успеха).
3. **Аналитика в Python:**
   - Создайте объект распределения binom из scipy.stats.
   - Сгенерируйте массив всех возможных исходов k от 0 до n.
   - Вычислите вероятности $P(k)$ для каждого k (функция .pmf).
   - Найдите теоретическое наивероятнейшее число (используя формулу из справки) и выведите его на экран.
4. **Визуализация:**
   - Постройте график (столбчатую диаграмму), где по оси X — количество успехов k, а по оси Y — вероятность $P(k)$.
   - Отметьте на графике другим цветом столбец, соответствующий наивероятнейшему числу.

Задание 2. Оценка параметра методом максимального правдоподобия

1. Экспоненциальное распределение

Цель: Проверить аналитический вывод из Задачи 3 на практике.

1. Задайте истинное значение параметра $\lambda_{true} = 2.0$.

2. Сгенерируйте выборку из $N = 1000$ случайных чисел, распределенных экспоненциально с параметром λ_{true}.
    - *Примечание:* В numpy.random.exponential параметр scale равен $1/\lambda$. То есть scale = 0.5.
3. Постройте гистограмму полученной выборки.
4. **Расчет оценки:**
    - Вычислите выборочное среднее $\bar{x}$ (метод .mean()).
    - Вычислите оценку параметра $\hat{\lambda}$ по формуле, которую вы вывели в Задаче 3 (Часть А).
5. **Сравнение:** Выведите на экран истинное значение λ и полученную оценку $\hat{\lambda}$. Вычислите относительную ошибку в процентах.
6. (Дополнительно) Повторите п. 2-5 для выборки размером $N = 10000$. Как изменилась точность оценки?

2. Нормальное распределение

1. Сгенерируйте выборку из 1000 чисел из нормального распределения с истинными параметрами μ_true = 10 и σ_true = 2.
2. Вычислите оценки максимального правдоподобия: μ_mle (как выборочное среднее) и σ2_mle (как выборочную дисперсию np.var()).
3. Сравните полученные оценки с истинными значениями.

Контрольные вопросы

1. Какие условия должны выполняться, чтобы можно было применить формулу Бернулли?
2. Как меняется форма распределения вероятностей в схеме Бернулли при изменении p (например, $p = 0.1$ против $p = 0.5$)?
3. В чем физический смысл функции правдоподобия? Почему мы ищем её максимум?
4. Почему при использовании ММП часто переходят к логарифму функции правдоподобия?
5. Всегда ли наивероятнейшее число единственно? Приведите пример, когда это не так.

Требования к отчету

Отчет должен содержать:

1. Титульный лист.
2. Цель работы.
3. **Часть А:** Показываем в тетради на занятии
4. **Часть Б: Загружаем на Гугл Диск отчет в Jupyter Lab**
    - Код программы.
    - График распределения вероятностей (Задание 1) с отмеченным наивероятнейшим числом.
    - Гистограмма экспоненциального и нормального распределения (Задание 2).
    - Сравнение истинного параметра λ и его оценки $\hat{\lambda}$ для разных размеров выборки. Вывод о влиянии объема выборки на точность оценки.
5. Ответы на контрольные вопросы.