Семинар 1

Суслова Ирина

5 февраля 2024

1 Введение

Курсы до мат. статистики:

Обычная задача в теорвере: Задан набор случайных величин, сказали, какое у них распределение, как они соотносятся друг с другом. Из этих случайных величин сделали какой-то новый набор случайных величин (взяли функцию или произведение например). В теорвере задается вопрос: Что за случайные величины мы получили, с какими характеристиками, например, какое распределение, какое мат. ожидание и.т.д

Мат. статистика

Решается немного обратная задача: Дали случайную величину, распределение неизвестно, но задается вопрос "Как она распределена?"
Мат. статистика разрабатывает методы восстановления численных характеристик с.в. по их измерениям. Например, если дали кубик с шестью гранями. Интерпретируем это как случайную величину ξ. С какой вероятностью на кубике выпадает единица? А шестерка? Допустим, мы подкинули кубик и получили 3 - это наше измерение.
Стоит отметить, что **нельзя** абсолютно точно восстановить численные характеристики, распределение по измерениям. Даже если возьмете бесконечно большое число измерений. Но тем не менее мы можем оценить наше распределение, аппроксимировать по измерениям. В нашем курсе мы будем изучать методы, как это можно сделать, как можно оценить.

Задача 1.1 (Пример с монеткой)**.** Рассмотрим стандартную монетку с орлом и решкой. Допустим, если выпадает решка, то мы будем интерпретировать это как единица, а если выпадет орел - 0. По сути мы имеем дело с какой-то случайной величиной $\xi \in Be(p)$. p - некий неизвестный нам параметр, вероятность выпадения единицы. Поставим задачу оценить p каким-то образом

Решение:
Рассматрим два случая:

1. $p = \frac{1}{2}$

2. Подкинем монетку n раз. Посчитаем, сколько раз выпала единичка и поделим на общее количество бросков.

Рассмотрим второй случай. Пусть X_1 - результат первого броска (с вероятностью p это единица, с вероятностью $1-p$ 0, т.е случайная величина), X_2 - результат второго броска, ..., X_n - результат n-го броска.
Будем считать, что $X_1, \ldots, X_n$ независимы в совокупности. Тогда предлагаемая оценка:

$$\hat{p} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

здесь мы рисуем крышку, чтоб отличать истинное значение p от нашей оценки $\hat{p}$, являющейся случайной величиной.
Сравним две предложенные оценки:

$$\mathbb{E}\hat{p} = \frac{1}{n}\mathbb{E}\sum_{i=1}^{n} X_i = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}X_i = p$$

Значит наша оценка принимает в среднем значение оцениваемого параметра p
А оценка $\tilde{p} = \frac{1}{2}$ непонятно как соотносится с оцениваемым параметром.

$$\mathbb{D}\hat{p} = \frac{1}{n^2}\sum_{i=1}^{n}\mathbb{D}X_i = \frac{p(1-p)}{n} \xrightarrow[n\to\infty]{} 0$$

То есть чем больше измерений, тем ближе наша оценка к истинному значению. А оценка $\tilde{p}$ не стремится к истинному: она либо попала, либо нет. Это плюсы в пользу оценки $\hat{p}$.

2 Выборка и оценка функции распределения

Определение 2.1.
Пусть $\xi \in \mathbb{R}^k$ - случайный вектор с функцией распределения $F_\xi(x)$. Тогда набор случайных векторов $X = (X_1, \ldots, X_n)$, распределение которых совпадает с распределением ξ, называется выборкой из распределения случайного вектора ξ. n - объем выборки

Определение 2.2.
Выборка $X = (X_1, \ldots, X_n)$ называется простой, если все $X_1, \ldots, X_n$ независимы в совокупности

Будем предполагать далее, что везде под выборкой имеется в виду именно простая выборка. Также предполагаем, что ξ является случайной величиной

Определение 2.3.
Любая измеримая функция выборки является статистикой

Статистика является случайной величиной.
Пусть дана случайная величина ξ и выборка $X = (X_1, \ldots, X_n)$ с неизвестной функцией распределения $F_\xi(x) = F_{X_1}(x) = \ldots = F_{X_n}(x)$.

$$F_\xi(x) = \mathbb{P}(\xi < x)$$

. Попробуем оценивать через отношение благоприятных исходов к общему числу измерений:

$$\hat{F}_n(x) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{I}(X_i < x)$$

$\hat{F}_n(x)$ называется эмпирической функцией распределения (ЭФР)-случайная функция (случайный процесс).
ЭФР обладает свойствами:

1. $0 \leq \hat{F}_n(x) \leq 1$
2. $n\hat{F}_n(x) \in \mathrm{Bi}(n, F_\xi(x))$
3. $\mathbb{E}\hat{F}_n(x) = F_\xi(x)$
4. $\mathbb{D}\hat{F}_n(x) = \frac{F_\xi(x)(1-F_\xi(x)}{n} \xrightarrow[n\to\infty]{} 0$
5. ЗБЧ по Хинчину $\hat{F}_n(x) \xrightarrow[n\to\infty]{\mathbb{P}} F_\xi(x)$ в каждой точке $x \in \mathbb{R}$
6. УЗБЧ $\hat{F}_n(x) \xrightarrow[n\to\infty]{..} F_\xi(x)$ в каждой точке $x \in \mathbb{R}$
7. $\frac{n\hat{F}_n(x)-nF_\xi(x)}{\sqrt{nF_\xi(x)(1-F_\xi(x)}} \xrightarrow[n\to\infty]{d} \xi_{N(0,1)}$ в каждой точке $x \in \mathbb{R}$

Введем величину

$$D_n = \sup_{x\in\mathbb{R}} |\hat{F}_n(x) - F_\xi(x)|$$

Получили случайную величину, не зависящую от x, которая называется статистикой Колмогорова-Смирнова

Теорема 2.1 (Теорема Гливенко)**.**
$D_n \xrightarrow[n\to\infty]{п.н} 0$

Теорема 2.2 (Теорема Колмогорова (о скорости сходимости D_n к нулю))**.**
Если $F_\xi(x)$ непрерывна, то при любом $t > 0$

$$\lim_{n\to\infty} \mathbb{P}(\sqrt{n}D_n \leq t) = K(t) = \sum_{j=-\infty}^{\infty} (-1)^j\, exp(-2j^2t^2)$$

При $t \leq 0$ считаем $K(t) = 0$

K(t) - функция распределения Колмогорова.
Таким образом, статистика Колмогорова-Смирнова стремится к нулю как $\frac{1}{\sqrt{n}}$.

3 Порядковые статистики

Определение 3.1.
Пусть дано вероятностное пространство $(\Omega, \mathcal{F}, \mathbb{P})$, на котором определены элементы выборки $X_1, \dots, X_n$ и $x_i = X_i(\omega)$, $\omega \in \Omega$. Пронумеруем последовательность $\{x_i\}$ в порядке неубывания: $x_{(1)} \le x_{(2)} \le \dots \le x_{(n)}$. Тогда функция $X_{(k)}(\omega) = x_{(k)}$ называется k-ой порядковой статистикой

$X_{(1)} = \min(X_1, \dots, X_n)$, $X_{(n)} = \max(X_1, \dots, X_n)$.
Алгоритм отрисовки графика ЭФР:

1. Измеряем ξ n раз и получаем $X_1, \dots, X_n$
2. Сортируем выборку по неубыванию
3. Рисуем кусочно-постоянную функцию со скачками в $X_{(k)}$ на величину $\frac{1}{n}$