2) Рассм. с.в.

$$Y_n = \frac{\overline{X}_n - M\overline{X}_n}{\sqrt{D\overline{X}_n}} \underset{\text{прибл., т.к. } n \gg 1}{\sim} N(0,1)$$

$$P\{k_1 \le k \le k_2\} = P\{k_1 \le \sum_{i=1}^{n} X_i \le k_2\} =$$

$$= P\{\frac{k_1}{n} \le \overline{X}_n \le \frac{k_2}{n}\} = P\{\frac{k_1}{n} - p \le \overline{X}_n - p \le \frac{k_2}{n} - p\} =$$

$$= P\{\underbrace{\frac{\frac{k_1}{n} - p}{\sqrt{pq}/\sqrt{n}}}_{x_1} \le \underbrace{\frac{\overline{X}_n - p}{\sqrt{pq}/\sqrt{n}}}_{=Y_n} \le \underbrace{\frac{\frac{k_2}{n} - p}{\sqrt{pq}/\sqrt{n}}}_{x_2}\} \approx$$

$$\underset{Y_n \sim N(0,1)}{\approx} \Phi_0(x_2) - \Phi_0(x_1), \text{ чтд.}$$

~~Замечание~~

Пример.

В эксперименте Пирсона по подбрасыванию правильной монеты из $n = 24000$ подбрасываний герб выпал 12012 раз. Какова вер-ть того, что при повт. проведении эксперимента отклонение набл. частоты выпадения герба от теор. окажется таким же или больше?

Решение:

01.03.24. Лекция 3.

1) Рассм. схему Бернулли, где:
- "успех" — выпад. герба;
- "неудача" — " — решки;

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

$A = \{$отклонение относит. частоты не меньше, чем в проведенном эксп.$\}$

$\overline{A} = \{$отклонение отн. частоты меньше, чем — " — $\}$

2) $P\{\overline{A}\} = P\{11988 < k < 12012\} \approx$

$\approx \Phi_0(x_2) - \Phi_0(x_1) \ominus$

$\frac{k}{n}$ — отн. частота

$\left|\frac{k}{n} - \frac{1}{2}\right| < \frac{12012}{n}$

$$x_2 = \frac{12012 - np}{\sqrt{npq}} = \frac{12012 - 12000}{\sqrt{24000/4}} = \frac{12}{\sqrt{6000}} \approx 0.155$$

$$x_1 = \frac{11988 - np}{\sqrt{npq}} \approx -0.155$$

$\ominus\ 2\Phi_0(0.155) \approx 2 \cdot 0.0615 = 0.123$

$P(A) = 1 - P(\overline{A}) \approx 1 - 0.123 = \boxed{0.877}$ ← Ответ.

МАТЕМАТИЧЕСКАЯ СТАТИСТИКА.

Основные понятия выборочной теории.

① Основные определения

Теория вероятностей – одна из ветвей "чистой" математики, которая строится дедуктивно исходя из вполне определенного набора аксиом.

Мат. статистика – раздел прикладной математики, кот. строится индуктивно, т.е. от наблюдения к гипотезе, при этом аргументация основана на выводах теории вероятностей.

Пример.

1) Типичная задача ТВ:
Вероятность выпадения герба при одном подбр. монеты равна p. Какова вер-ть того, что после n бросков герб выпадет m раз?

2) Типичная задача МС:
После n бросков монеты герб выпал m раз. Чему равна вер-ть p появления герба при одном броске?

- Основная задача МС:
разработка методов получения научно обоснованных выводов о массовых процессах или явлениях по данным наблюдений или экспериментов.

При этом такие выводы относятся не к отд. экспериментам, а представляют собой утв-я отн. вероятностных характеристик изучаемого процесса или явления.

! Замечание! Упрощенно можно считать, что "общая" задача МС состоит в следующем.

Пусть X – с.в., ф-я распр-я которой неизвестна. Требуется по результатам наблюдений (т.е. по имеющимся реализациям) сл. вел. X сделать вывод о ее ф-ие распр-я.

Обычно эта задача встречается в 2^x вариантах.

– I задача МС: ф-я распр-я с.в. X неизвестна "вообще", т.е. неизвестен ее тип (класс). Требуется на основании наблюдений определить тип этого ф-ия распределения и, возможно, значение пар-ров этого ф-ия.

– II задача МС: известен тип (класс) ф-ия распр-я с.в. X, но неизв. значение одного или неск. параметров. Требуется на осн. набл-я оценить значение этих пар-ров.

Пример. (E)

1) Известно, что с.в. $X \sim N(m, \sigma^2)$, где значение σ^2 - изв-но. Требуется оценить значение пар-ра m.

2) Известно, что с.в. $X \sim R[a, b]$, где a, b - неизв-ны. Требуется оценить значение пар-в a, b.

Пусть X - случ. вел.

Опр. Множество всех возм. значений сл. вел. X наз. генеральной совокупностью X.

Опр. Случайной выборкой объема n из ген. совокупности X наз. случайный вектор

$$\vec{X} = (X_1, \dots, X_n),$$

где 1) $X_1, \dots, X_n$ - незав. в совок-ти;
2) $X_i \sim X$, $i = \overline{1,n}$

Замечание. Пусть $F(t)$ - ф-я расп. с.в. X.
Тогда ф. расп. случ. выборки объема n из ген. сов-ти X:

$$F_{\vec{X}}(t_1, \dots, t_n) = P\{X_1 < t_1, \dots, X_n < t_n\} =$$
$$= \{X_i, i = \overline{1,n}, \text{ незав. в сов.}\} = F_{X_1}(t_1) \cdot \ldots \cdot F_{X_n}(t_n) =$$
$$= \{X_i \sim X \Rightarrow F_{X_i} = F, i = \overline{1,n}\} = F(t_1) \cdot \ldots \cdot F(t_n).$$

Опр. Выборкой объема n из ген. сов-ти X наз. (любая) реализация случайной выборки того же объема n из той же ген. сов-ти X:

$$\vec{x} = (x_1, \dots, x_n)$$

При этом x_k, $k = \overline{1,n}$ - k-й элемент выборки.

Опр. Мн-во всех возможных значений в-ра $\vec{X}$ объема n наз. выборочным пространством.

- Обозначение: X_n

Опр. Скалярная функция $g(\vec{X})$ от случайной выборки $\vec{X}$ наз. статистикой.
(выборочной характеристикой)

Закон расп-я с.в. $g(\vec{X})$ наз. выборочным з-ном расп-я.

Замечание. 1) $\vec{X}$ - с.в.
$g: \mathbb{R}^n \to \mathbb{R}$
Тогда $Y = g(\vec{X})$ - с.в.
2) Значение $g(\vec{x})$ - выборочное значение статистики
3) Пусть $\vec{x} = (x_1, \dots, x_n)$ - выборка объема n из ген. сов-ти X.

Можно считать, что з-н распр-я с.в. X моделируется з-ном распр-я дискр. с.в. $\tilde{X}$, ряд распр-я кот. выглядит так:

$\tilde{X}$	x_1	...	x_i	...	x_n
P	$1/n$	...	$1/n$	...	$1/n$

$$M[\tilde{X}] = \left\{ \sum_{i=1}^{n} x_i p_i \right\} = \frac{1}{n} \sum_{i=1}^{n} x_i \text{ — обозн. } \bar{x}$$

$$D[\tilde{X}] = \left\{ \sum_{i=1}^{n} p_i (x_i - M[\tilde{X}])^2 \right\} = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})^2$$

Значение $M[\tilde{X}]$, $D[\tilde{X}]$ оценивают зн. MX, DX.

Пусть $\vec{X} = (X_1, \dots, X_n)$ — сл. выборка объема n из ген. сов-ти X.

Опр. Выборочным средним наз. статистика

$$\hat{m}_1(\vec{X}) = \bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i.$$

Опр. Выборочной дисперсией наз. статистика

$$\hat{\sigma}^2(\vec{X}) = \frac{1}{n} \sum_{i=1}^{n} (X_i - \bar{X})^2.$$

Опр. Выборочным начальным моментом порядка k наз. статистика

$$\hat{m}_k(\vec{X}) = \frac{1}{n} \sum_{i=1}^{n} X_i^k$$

Опр. Выборочным центральным моментом пор. k наз. статистика

$$\hat{\mathring{m}}_k(\vec{X}) = \frac{1}{n} \sum_{i=1}^{n} (X_i - \bar{X})^k.$$

Замечание: $\hat{\mathring{m}}_2(\vec{X}) = \hat{\sigma}^2(\vec{X})$.

② Предварительная обработка результатов эксперимента.

I. Вариационный ряд.

Пусть $\vec{x} = (x_1, \dots, x_n)$ — выборка из скал. ген. сов-ти.

Расположим значения x_i, $i = \overline{1, n}$, в порядке неубывания:

$$x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)}.$$

Опр. Вариационным рядом выборки $\vec{x}$ наз. кортеж

$$(x_{(1)}, \dots, x_{(n)})$$

Опр. Вариационным рядом случ. выборки $\vec{X} = (X_1, \dots, X_n)$ наз. случ. вектор

$$(X_{(1)}, \dots, X_{(n)})$$

такой, что для любой реализации $\vec{x}$ сл. выб. $\vec{X}$ соотв. реал. $(x_{(1)}, \dots, x_{(n)})$ этого вектора совпадает с вар. рядом выборки $\vec{x}$.

Замечание Пусть 1) $\vec{X} = (X_1, \dots, X_n)$ – случ. выб. из ген. сов-ти X.

2) $F(t)$ – ф-я распр. с. в. X.

Тогда

$$F_{X_{(1)}}(t) = P\{X_{(1)} < t\} = 1 - P\{X_{(1)} \geqslant t\} \underset{\text{н.з.}}{=}$$

$$= 1 - P\{X_1 \geqslant t, X_2 \geqslant t, \dots, X_n \geqslant t\} =$$

$$F_{X_{(n)}}(t) = P\{X_{(n)} < t\} = P\{X_1 < t, \dots, X_n < t\} =$$

$$= \{X_i, i = \overline{1;n} \text{ – незав}\} = P\{X_1 < t\} \cdot \ldots \cdot P\{X_n < t\} =$$

$$= F_{X_1}(t) \cdot \ldots \cdot F_{X_n}(t) = \left\{ \begin{matrix} X_i \sim X \\ F_i = F \end{matrix} \right\} = [F(t)]^n$$

$$= 1 - P\{X_1 \geqslant t\} \cdot P\{X_2 \geqslant t\} \cdot \ldots \cdot P\{X_n \geqslant t\} =$$

$$= 1 - (1 - P\{X_1 < t\}) \cdot \ldots \cdot (1 - P\{X_n < t\}) = \{F_i = F\} =$$

$$= 1 - (1 - F(t))^n$$