Выборка. Выборочное пространство

Грауэр Л.В.

Примеры задач

Какова доля брака в продукции завода?

Правильно ли настроен станок?

Действует ли новое лекарство?

Однородны ли посетители магазина в выходные и в будние?

Сколько заказать товара на следующий месяц?

Согласованы ли оценки экспертов?

Основные задачи математической статистики

- приближенное определение вероятности события по относительной частоте
- нахождение приближенного закона распределения с.в.
- оценивание числовых характеристик или параметров распределения с.в.
- проверка статистических гипотез
- определение эмпирической зависимости между переменными, описывающими случайное явление

Выборка. Выборочное пространство

В одномерном случае
$\xi(\omega) : \Omega \longrightarrow \mathbb{R}$

$(\mathbb{R}, \mathcal{B}(\mathbb{R}), P_\xi)$

В многомерном случае
$(\xi_1, \dots, \xi_m) : \Omega \longrightarrow \mathbb{R}^m,$

$(\mathbb{R}^m, \mathcal{B}(\mathbb{R}^m), P_\xi)$

Совокупность взаимно независимых реализаций случайной величины ξ образует выборку $X_{[n]}$ объема n:

$$X_{[n]} = (X_1, \ldots, X_n)\,,$$

где X_i — числовая реализация случайной величины ξ в i-ом эксперименте ($i = 1, \ldots, n$).

Если $\xi = (\xi_1, \dots, \xi_m)^T$ — случайный вектор, то

$$X_1 = \begin{pmatrix} X_{11} \\ X_{21} \\ \vdots \\ X_{m1} \end{pmatrix}, \dots, X_n = \begin{pmatrix} X_{1n} \\ X_{2n} \\ \vdots \\ X_{mn} \end{pmatrix}.$$

Случайная величина ξ, реализации которой мы наблюдаем, часто называется *генеральной совокупностью*.

Примеры

$$\xi \sim B(N, p),\ P(\xi = k) = C_N^k p^k (1-p)^{N-k}$$

$$X_{[10]} = (0, 0, 0, 1, 0, 0, 0, 0, 0, 0)$$

Примеры

$$\xi \sim B(N,p),\ P(\xi = k) = C_N^k p^k (1-p)^{N-k}$$

$$X_{[10]} = (0,0,0,1,0,0,0,0,0,0)$$

$$\xi \sim N(a,\sigma)$$

$$X_{[10]} = (1.34, 0.70, 1.21, 0.46, 1.40, 1.47, 1.56, 1.09, 1.62, 1.56)$$

Простой случайный выбор

Процесс составления выборки называется *выбором*

Простым случайным выбором *называется выбор с возвращением в урновой модели, когда из конечного множества каждый элемент выбирается независимо и равновероятно с другими элементами*

Свойства простого случайного выбора

С теоретической точки зрения элементы выборки — случайные величины

Свойства простого случайного выбора

С теоретической точки зрения элементы выборки — случайные величины

Все элементы выборки могут быть выбраны и равновероятно

Свойства простого случайного выбора

С теоретической точки зрения элементы выборки — случайные величины

Все элементы выборки могут быть выбраны и равновероятно

Каждый элемент конкретной выборки получен в равных условиях выбора

Виды реальных выборов

Механический выбор

Серийный выбор

Типический выбор

Выбор на основе суждения

Функция распределения выборки:

$$F_{X_{[n]}}(x_1, \ldots, x_n) =$$

Функция распределения выборки:

$F_{X_{[n]}}(x_1, \ldots, x_n) =$

Выборкам объема n соответствует *выборочное пространство*

$$\left(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n), P_{X_{[n]}}\right)$$

При $n \to \infty$ рассмотрим бесконечномерное пространство:

$$\left(\mathbb{R}^\infty, \mathcal{B}(\mathbb{R}^\infty), P_{X_{[\infty]}}\right).$$

$\left(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n), P_{X_{[n]}}\right)$ — подпространство $\left(\mathbb{R}^\infty, \mathcal{B}(\mathbb{R}^\infty), P_{X_{[\infty]}}\right)$, соответствующее первым n координатам.

Эмпирическая вероятностная мера

Эмпирическим распределением назовем вероятностную меру, определенную следующим образом

$$P_n^*(B) = \frac{\nu(B)}{n},$$

где $B \in \mathcal{B}(\mathbb{R})$, а $\nu(B)$ — количество элементов выборки, попавших в B.

Эмпирической функцией распределения называется функция

$$F_n^*(x) = P_n^*(-\infty; x) = \frac{\nu(-\infty; x)}{n}, \quad x \in \mathbb{R}.$$

$$(X_1, X_2, \ldots, X_n) \Rightarrow X_{(1)} < X_{(2)} < \ldots < X_{(n)}$$

$$F_n^*(x) = \begin{cases} \quad, & \text{если } x \leqslant X_{(1)}; \\ \quad, & \text{если } X_{(1)} < x \leqslant X_{(2)}; \\ \quad, & \text{если } X_{(2)} < x \leqslant X_{(3)}; \\ \ldots & \\ \quad, & \text{если } X_{(k)} < x \leqslant X_{(k+1)}; \\ \ldots & \\ \quad, & \text{если } x > X_{(n)}. \end{cases}$$

Теорема Гливенко-Кантелли. Теорема о предельном распределении эмпирических вероятностей

Теорема

Для $\forall\ B \in \mathcal{B}(\mathbb{R})$ выполняется:

$$P_n^*(B) \xrightarrow[n \longrightarrow \infty]{\text{п.н.}} P_\xi(B).$$

и для $\forall\ x \in \mathbb{R}$ выполняется:

$$F_n^*(x) \xrightarrow[n \longrightarrow \infty]{\text{п.н.}} F_\xi(x).$$

Теорема Гливенко-Кантелли

Пусть заданы ф.р. $F_\xi(x)$ и эмпирическая ф.р. $F_n^*(x)$, тогда

$$\sup_{x\in\mathbb{R}} |F_n^*(x) - F_\xi(x)| \xrightarrow[n\longrightarrow\infty]{\text{п.н.}} 0$$

Теорема

Для $\forall\ B \in \mathcal{B}(\mathbb{R})$ выполняется:

$$\frac{\sqrt{n}\,(P_n^*(B) - P_\xi(B))}{\sqrt{P_\xi(B)(1 - P_\xi(B))}} \xrightarrow[n\longrightarrow\infty]{d} \zeta,$$

где $\zeta \sim N(0,1)$.