Математические основы искусственного интеллекта
Описательные статистики. Метод моментов. Точечные и интервальные оценки параметров генеральной совокупности

Солодушкин Святослав Игоревич

Кафедра вычислительной математики и компьютерных наук,
УрФУ имени первого Президента России Б.Н. Ельцина

Октябрь 2021

Задачи

1. Из генеральной совокупности извлечена выборка, заданная вариантами x_i и соответствующими им частотами. Найдите несмещенную оценку генеральной средней и дисперсии.

x_i	2	5	7	10
n_i	16	12	8	14

2. По выборке объема N=41 найдена смещенная оценка генеральной дисперсии равная 8. Найдите несмещенную оценку дисперсии генеральной совокупности.

Задачи

3. Найти доверительный интервал для оценки с надежностью $P = 0,95$ неизвестного математического ожидания m нормально распределенной генеральной совокупности X, если даны генеральное среднее квадратическое отклонение $\sigma = 5$, выборочная средняя $\overline{X} = 15$, а объем выборки $n = 25$.
4. По данным $n = 9$ независимых равноточных измерений некоторой физической величины найдены среднее арифметическое результатов измерений $\overline{X} = 10$ и исправленное среднее квадратическое отклонение $S^2 = 6$. Оценить истинное значение измеряемой величины при помощи доверительного интервала с надежностью $=0.95$.
5. Как изменится решение предыдущей задачи, если построить доверительный интервал с надежностью равной 0.99? А если при этом еще увеличить объем выборки до $n = 16$?

Решение задачи № 1

Множество всех объектов, подлежащих изучению, называется генеральной совокупностью. Множество случайно отобранных объектов называется выборочной совокупностью или выборкой.

Для оценки неизвестных параметров теоретического распределения служат статистические оценки. Статистическая оценка, определяемая одним числом, называется точечной оценкой.

Точечная статистическая оценка, математическое ожидание которой равно оцениваемому параметру при любом объеме выборки, называется несмещенной оценкой. Статистическая оценка, математическое ожидание которой не равно оцениваемому параметру является смещенной.

Решение задачи № 1

Несмещенной оценкой генеральной средней (математического ожидания) служит выборочная средняя

$$\overline{x} = \frac{\sum_{i=1}^{k} x_i n_i}{n},$$

где k — число различных вариант, $n = \sum_{i=1}^{k} n_i$ — обеъем выборки.
Объем данной выборки равен $n = 16 + 12 + 8 + 14 = 50$.
Далее по формуле вычисляем несмещенную оценку генеральной средней:

$$\overline{x} = \frac{2 \cdot 16 + 5 \cdot 12 + 7 \cdot 8 + 10 \cdot 14}{50} = 5.76$$

Смещенной оценкой генеральной дисперсии служит выборочная дисперсия

$$D\text{выб} = \frac{1}{n}\sum_{i=1}^{k}(x_i - \bar{x})^2 ,$$

Несмещенной оценкой генеральной дисперсии является «исправленная дисперсия»

$$S^2 = \frac{n}{n-1}D\text{выб} = \frac{1}{n-1}\sum_{i=1}^{k}(x_i - \bar{x})^2$$

Таким образом, получаем искомую несмещенную оценку дисперсии генеральной совокупности:

$$\frac{41}{41-1}8 = 8.2,$$

Доверительный интервал для математического ожидания нормальной выборки

Случайная величина

$$Z = \frac{\bar{X} - m}{\sigma/\sqrt{n}}$$

имеет стандартное нормальное распределение $\mathrm{N}(0, 1)$.

Пусть z_α — это α-квантиль стандартного нормального распределения, тогда в силу симметрии имеем:

$$\mathrm{P}\left(-z_{1-\frac{\alpha}{2}} \leq Z \leq z_{1-\frac{\alpha}{2}}\right) = 1 - \alpha.$$

После подстановки выражения для Z получаем:

$$\mathrm{P}\left(\bar{X} - z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \leq m \leq \bar{X} + z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha.$$

Решение задачи № 3

Интервальной оценкой называется интервал, покрывающий оцениваемый параметр. Доверительным интервалом является интервал, который с данной надежностью покрывает оцениваемый параметр.
В данной задаче имеем

$$P\left(-z_{1-\frac{\alpha}{2}} \leq Z \leq z_{1-\frac{\alpha}{2}}\right) = 1 - \alpha = 0.95.$$

Следовательно, обращая функцию Лапласа, находим $z_{1-\frac{\alpha}{2}} = 1.96$ Далее получаем

$$15 - 1.96\frac{5}{\sqrt{25}} \leq m \leq 15 + 1.96\frac{5}{\sqrt{25}}$$

В итоге, $13.04 \leq m \leq 16.96$.

Доверительный интервал для математического ожидания нормальной выборки

Пусть $X_1, \dots, X_n$ — случайная выборка из нормально распределенной генералной совокупности $X \sim \mathrm{N}(m, \sigma^2)$, где m, σ^2 — неизвестные константы.

Определим произвольное $\alpha \in (0, 1)$ и построим доверительный интервал для неизвестного среднего m.

Доверительный интервал для математического ожидания нормальной выборки

Случайная величина

$$T = \frac{\bar{X} - m}{S/\sqrt{n}}$$

имеет распределение Стьюдента с $n-1$ степенями свободы $\mathrm{t}(n-1)$, где S — несмещенное выборочное стандартное отклонение.

Пусть $t_{\alpha,n-1}$ — это α-квантиль распределения Стьюдента, тогда в силу симметрии имеем:

$$\mathrm{P}\left(-t_{1-\frac{\alpha}{2},n-1} \leq T \leq t_{1-\frac{\alpha}{2},n-1}\right) = 1-\alpha.$$

После подстановки выражения для T получаем:

$$\mathrm{P}\left(\bar{X} - t_{1-\frac{\alpha}{2},n-1}\frac{S}{\sqrt{n}} \leq m \leq \bar{X} + t_{1-\frac{\alpha}{2},n-1}\frac{S}{\sqrt{n}}\right) = 1-\alpha.$$

Таблица 4

Распределение Стьюдента с k степенями свободы

$$p = \frac{\Gamma((k+1)/2)}{\Gamma(k/2)\sqrt{\pi k}} \int_{-\infty}^{y_p} \left(1 + \frac{y^2}{k}\right)^{-\frac{k+1}{2}} dy$$

k \ p	0,9	0,95	0,975	0,99	0,995	0,999
1	3,078	6,314	12,71	31,82	63,66	318,3
2	1,886	2,920	4,303	6,965	9,925	22,33
3	1,638	2,353	3,182	4,541	5,841	10,21
4	1,533	2,132	2,776	3,747	4,604	7,173
5	1,476	2,015	2,571	3,365	4,032	5,893
6	1,440	1,943	2,447	3,143	3,707	5,208
7	1,415	1,895	2,365	2,998	3,500	4,785
8	1,397	1,860	2,306	2,897	3,355	4,501
9	1,383	1,833	2,262	2,821	3,250	4,297
10	1,372	1,813	2,228	2,764	3,169	4,144
12	1,356	1,782	2,179	2,681	3,055	3,930
14	1,345	1,761	2,145	2,625	2,977	3,787
16	1,337	1,750	2,120	2,584	2,921	3,686
18	1,330	1,734	2,101	2,552	2,878	3,611
20	1,3[illegible]	1,725	2,0[illegible]	2,52[illegible]	2,845	3,5[illegible]

Решение задачи № 4

В данной задаче имеем

$$P\left(\bar{X} - t_{1-\frac{\alpha}{2},n-1}\frac{S}{\sqrt{n}} \leq m \leq \bar{X} + t_{1-\frac{\alpha}{2},n-1}\frac{S}{\sqrt{n}}\right) = 1 - \alpha = 0.95.$$

по таблицам для квантилей распределения Стьюдента находим $z_{1-\frac{\alpha}{2}} = 2.306.$
Далее получаем

$$10 - 2.306\frac{6}{\sqrt{9}} \leq m \leq 10 - 2.306\frac{6}{\sqrt{9}}.$$

В итоге, $5.388 \leq m \leq 14.612.$