



НЕПАРАМЕТРИЧЕСКИЕ МЕТОДЫ

Авторы: Ю. В. Прохоров

НЕПАРАМЕТРИЧЕСКИЕ МЕТОДЫ математической статистики, методы, не предполагающие знания функционального вида теоретич. распределения. Название «Н. м.» подчёркивает их отличие от классических (параметрических) методов, в которых предполагается, что неизвестное теоретич. распределение принадлежит к-л. семейству, зависящему от конечного числа параметров (напр., семейству нормальных распределений), и которые позволяют по результатам наблюдений оценивать неизвестные значения этих параметров и проверять те или иные гипотезы относительно их значений.

Одним из Н. м. является критерий Колмогорова проверки согласованности теоретич. и эмпирич. распределений. Пусть взаимно независимые случайные величины X_1, X_2, \dots, X_n (выборка объёма n) имеют теоретич. функцию распределения $F(x)$ и пусть $F_n(x)$ – эмпирич. функция распределения, построенная по наблюдениям над этими случайными величинами (F_n является несмещённой и состоятельной оценкой для F). Пусть D_n – наибольшее по абсолютной величине значение разности $F_n(x) - F(x)$. Случайная величина $\sqrt{n}D_n$ имеет, в случае непрерывности $F(x)$, функцию распределения $K_n(\lambda)$, не зависящую от F и стремящуюся при возрастании n к пределу $K(\lambda) = \sum_{j=-\infty}^{\infty} (-1)^j e^{-2j^2 \lambda^2}$.

Отсюда при достаточно больших n для вероятности $p_{n,\lambda}$ неравенства $\sqrt{n}D_n \geq \lambda$ получается приближённое выражение $p_{n,\lambda} \approx 1 - K(\lambda)$; функция $K(\lambda)$ табулирована. Её значения для некоторых λ приведены в таблице.

Таблица значений функции $K(\lambda)$						
λ	0,57	0,71	0,83	1,02	1,36	1,63
$K(\lambda)$	0,10	0,30	0,50	0,75	0,95	0,99

Равенство (*) используется для проверки гипотезы о том, что теоретич. распределением является распределение с заданной непрерывной функцией распределения $F(x)$: сначала по результатам наблюдений находят значение величины D_n , а затем по формуле (*) вычисляют вероятность получить отклонение F_n от F , большее или равное наблюдаемому. Если указанная вероятность достаточно мала, точнее – равна наперёд заданному малому положительному числу α (см. [Значимости уровень](#)), то в соответствии с общими принципами [статистических гипотез проверки](#) проверяемую гипотезу отвергают. В противном случае считают, что результаты опыта не противоречат проверяемой гипотезе. Аналогично проверяется гипотеза о том, что для двух независимых выборок объёмов n_1 и n_2 соответствующие (непрерывные) теоретич. функции распределения одинаковы (гипотеза однородности двух выборок). При этом вместо формулы (*) пользуются тем, что вероятность неравенства $D_{n_1, n_2} \sqrt{\frac{n_1 n_2}{n_1 + n_2}} < \lambda$ имеет пределом $K(\lambda)$, где D_{n_1, n_2}

есть наибольшее по абсолютной величине значение разности $F_{n_1}(x) - F_{n_2}(x)$. Приведённые примеры относятся к Н. м., основанным на разностях теоретич. и эмпирич. или двух эмпирич. функций распределения.

Значит. место в совр. математич. статистике занимают Н. м., в которых используются не сами эмпирич. функции распределения, а некоторые функции от порядковых статистик – членов вариационного ряда. Если используются порядковые номера результатов наблюдений (они называются рангами), то такие Н. м. называются ранговыми, они, как правило, являются критериями однородности. Напр., пусть X_1, X_2, \dots, X_n и Y_1, Y_2, \dots, Y_m – взаимно независимые элементы двух выборок с непрерывными функциями распределения. Для проверки гипотезы о том, что эти функции распределения одинаковы, можно использовать ранговый критерий, основанный на значениях функций $W = s(r_1) + s(r_2) + \dots + s(r_m)$, где r_j – ранги случайных величин Y_j , $j = 1, 2, \dots, m$, в общем вариационном ряду X_i и Y_j , а $s(1), s(2), \dots, s(n+m)$ – одна из возможных перестановок чисел $1, 2, \dots, n+m$. Выбор перестановки может быть осуществлён оптимальным образом.

Литература

Лит.: Смирнов Н. В., Дунин-Барковский И. В. Курс теории вероятностей и математической статистики для технических приложений. 3-е изд. М., 1969; Кендалл М. Дж., Стьюарт А. Статистические выводы и связи. М., 1973; Большев Л. Н., Смирнов Н. В. Таблицы математической статистики. 3-е изд. М., 1983.

Processing math: 0%