Математическая статистика

16 октября 2023

1 задача:

При уровне значимости $\alpha = 0,05$ проверить гипотезу $H_0 : a = 100$, против альтернативной гипотезы $H_1 : a \neq 100$.

Гипотеза о среднем нормальной совокупности при неизвестной дисперсии:

Возьмём $K = \sqrt{n}\frac{\overline{x} - a_0}{S}$
Для уровня значимости α найдём t такую, что
$P(|t_{n-1}| \geqslant t) = \alpha$
t – квантиль уровня α двустороннего распределения $|T_{n-1}|$.
$t = \text{СТЬЮДЕНТ.ОБР.2Х}(\alpha; n-1) = 2,01$
$|K| = 2,51$
$|K| \geqslant t$ – значит, принимаем гипотезу H_1.

2 задача:

В точке $(0, 1)$ находится источник излучения, равномерно испускающий частицы во все стороны. При пересечении с осью OX частица регистрируется. Случайная величина ξ – точка пересечения частицы с осью OX. Найти её закон распределения $F_\xi(x)$ и вычислить математическое ожидание $E\xi$.

φ – угол

$F\xi(x) = P(\xi < x) = P(-\frac{\pi}{2} < \varphi < \arctan(x)) = \frac{\arctan(x) - \frac{\pi}{2}}{\pi} = \frac{1}{2} + \frac{1}{\pi}\arctan(x)$ $f_\xi(x) = \frac{1}{\pi}\frac{1}{1+x^2}$

$E\xi = \int_{-\infty}^{+\infty} x f_\xi(x)dx = \frac{1}{2\pi}\int_{-\infty}^{+\infty} \frac{d(1+x^2)}{1+x^2} = \frac{1}{2\pi}\ln(1+x^2)|_0^\infty \to +\infty$ – абсолютной сходимости нет, $E\xi$ не существует.

Допустим, что источник излучения сдвинули по оси OX на θ.

$f_\xi(x) = \frac{1}{\pi}\frac{1}{1+(x-\theta)^2}$

Требуется дать оценку неизвестному параметру θ на основе выборки объёма n.

1) $\theta^* = \overline{x}$ – не работает, т.к. не существует математического ожидания.

2) т.к. распределение симметричное, то $\theta = Me$ и имеет смысл в качестве оценки взять $\theta^* = Me^*$.

Эта оценка будет состоятельной согласно теореме:

$Me^* \to^p Me$, причём со скоростью $\frac{1}{\sqrt{n}}$.

(Условие конечности первого момента необязательно).

Использование медианы

Если распределение симметричное, то для оценки центра медиану имеет смысл использовать в ситуации "жирных хвостов"и возможных выбросов.

Пример:

Распределение Коши: $f(x) = \dfrac{1}{\pi(1 + (x - \theta)^2)}$

$(-\infty; \theta - t)$ и $(\theta + t; +\infty)$ – хвосты распределения.

Недостатки медианы:

1. Хорошо работает только в случае симметричных распределений.

2. В случае нормального распределения сходимость приблизительно на 20% медленнее, чем у выборочного среднего, и до 40% в некоторых других случаях.

Другие подходы

I. Метод усечённого среднего

   Откидываем в выборке снизу и сверху по 5 единиц, $5 - 10\%$.

   Пусть имеется выборка объёма n. Уберём по k значений сверху и снизу.

   В результате получаем новую выборку объёма $n-2k$: $(X_{k+1}, ..., X_{n-k})$

   И для неё вычисляем выборочное среднее: $\frac{1}{n-k}\sum_{t=k+1}^{n-k} X_t$

   Эту величину можно считать компромиссом между выборочным средним и медианой.

   При $k = 0$ получаем $\overline{x}$.

   При $n - k = 1$ или $n - k = 2$ получаем Me.

II. Метод средних Уолша (1980).

   $X = (X_1, ..., X_n)$

   Для каждой пары данных берём среднее арифметическое:

   $y_k = \dfrac{X_i + X_j}{2}$, $1 \leqslant k \leqslant \dfrac{n(n-1)}{2}$.

   $Y = (Y_1, ..., Y_{\frac{n(n-1)}{2}})$ и для этой выборки вычисляем выборочную медиану Me^*.

Достоинства:

1. Сглаживаются выбросы
2. Скорость сходимости этой медианы падает по сравнению со средним выборочным приблизительно на 12% в худших случаях.