24.05.24. Лекция 13

Пример

1) $H_0 = \{X$ имеет з-н расп $N(1,2)\} = \{f_X(t) \equiv \frac{1}{\sqrt{2\pi}\sqrt{2}} e^{-\frac{(t-1)^2}{4}}\}$ – простая

$H_1 = \neg H_0$

2) $H_0 = \{X$ имеет норм. з-н расп-я$\} = \{(\exists m \in R)(\exists \sigma \in R^+)(\forall t \in R)$

$(f_X(t) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-m)^2}{2\sigma^2}})\}$ – сложная.

$H_1 = \neg H_0$.

Проверка осн. гипотезы H_0 сводится к оценке величины $\rho(F_n, F_0)$ рассогласования предполагаемой ф-ии F_0 расп. с.в. X и эмпирической ф-ии распределения, отвечающей имеющейся выборке $\vec{x}$ из ген. совок. X.

Опр. Критерием согласия наз. стат. критерий, служащий для проверки корректности гипотезы о том, что з-н распределения с.в. X совпадает с предполагаемым з-ном распределения, заданным ф-ей распределения F_0.

Замечание: (о способах выдвижения осн. гип. H_0)

1) Проанализировать априорную информацию о физ. интер. сл.в. X и сопоставить ее с исходными предпосылками построения стат. вер. моделей (норм., ...)

2) Построить гистограмму по выборке $\vec{x}$ и сравнить ее с ф-ями плотности типовых з-нов распределения (норм., экспоненц., равномерного, ...)

3) Построить эмпирическую ф-ю распределения и сравнить ее с ф-ями расп. типовых законов расп. (норм, ...)

② Критерий Колмогорова для простой гипотезы

Пусть 1) X – непр. с.в.

2) $\vec{X}$ – сл. выборка из ген. сов-ти X.

Рассмотрим задачу проверки гипотезы $H_0 = \{F_X(t) \equiv F_0(t)\}$ против $H_1 = \neg H_0$,

где F_0 – предполаг. ф. распределения с.в. X,

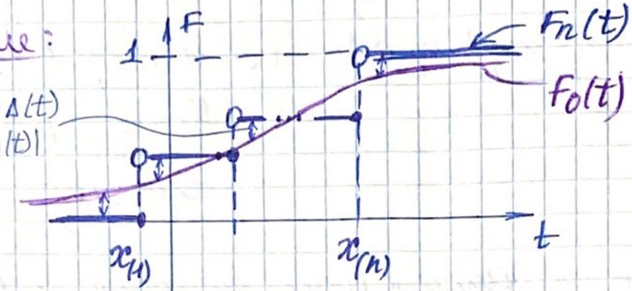
F_X – теоретич. ф. расп. с.в. X.

Замечание: F_0 – полностью изв. ф-я, в частности F_0 не зависит ни от каких неизв. пар-в $\Rightarrow$ H_0 – простая

Рассмотрим статистику $\Delta(\vec{X})$, значение которой опр. соотношением:
$$\Delta(\vec{x}) = \sup_{t\in\mathbb{R}} |F_n(t) - F_0(t)|,$$
где F_n – эмпирич. ф. расп., постр. на выборке $\vec{x}$,
n – объем выборки $\vec{x}$.

Замечание:

$\Delta(\vec{x}) = \sup_{t\in\mathbb{R}} |\Delta(t)|$



Очевидно, что в пользу гипотезы H_0 свид-ют "малые", "близкие к 0" значения статистики Δ, а в пользу H_1 – "большие" значения Δ. По этой причине крит. мн-во для рассм. задачи проверки гипотезы можно записать в виде
$$W = \{\vec{x} \in X_n : \Delta(\vec{x}) \geq \delta_{1-\alpha}\},$$
где $\delta_{1-\alpha}$ – квантиль ур-ня $1-\alpha$ ф-ия расп-я с. в. $\Delta(\vec{X}_n)$; при ист-ти H_0.
α – ур-нь значимости критерия.

При этом решающее правило имеет вид:
$$\vec{x} \in W \Rightarrow \begin{cases} \text{откл. } H_0 \\ \text{прин. } H_1 \end{cases}$$
$$\vec{x} \in X_n \setminus W \Rightarrow \begin{cases} \text{откл. } H_1 \\ \text{прин. } H_0. \end{cases}$$

Замечание: (о ф-ие расп. $\Delta(\vec{X})$ при ист. H_0.)

1) Построенный критерий требует исп-я квантилей с.в. $\Delta(\vec{X})$. Вообще говоря, ф-я расп. этой статистики зависит от:
   а) используемого предполагаемого ф-ия расп. F_0 с.в. X.;
   б) теор. ф-ия расп. с.в. X.

2) Однако в МС доказано след. утв-е:

Th

Пусть 1) с. в. $Y \sim R[0,1]$;
2) $\vec{Y}_n$ – сл. выборка из ген. сов-ти Y;
3) $R(t, \vec{Y}_n)$ – выборочная ф-я расп., постр. по сл. выборке $\vec{Y}_n$.

Тогда при ист-ти H_0 ф-я расп. стат-ки $\Delta(\vec{X}_n)$ совпадает с ф-ей расп. с. в. $\xi(n) = \sup_{t\in[0,1]} |R(t, \vec{Y}_n) - t|$

а) $Y \sim R[0,1]$ – з-н распр. изв.

б) $\vec{Y}_n$ – з-н распр. изв.

в) $Z(n) = \sup\limits_{t \in [0,1]} |R(t, \vec{Y}_n) - t|$ – с.в., ф. зав. от сл.в-ра $\vec{Y}_n$

$\Rightarrow$ з-н распр. с.в. $Z(n)$ для к. n известен.

3) Т.о. для всевозможных теоретических з-нов распр. с.в. X и всевозм. предполаг. з-нов распр. с.в. X, заданных ф-ей распр. F_0, (при спр-ти H_0) з-н распр-я с.в. $Z(n)$ известен.

4) Т.о. в кач. закона распр. с.в. $\Delta(\vec{X}_n)$ (в кач. квантилей $\delta_{1-\alpha}$) можно исп-ть з-н распр. (квантили) с.в. $Z(n)$.

Для $n \le 100$ составлены таблицы квантилей законов распределения с.в. $Z(n)$ их можно найти в лит-ре и прикладных пакетах матем. программ.

5) Что делать, если $n > 100$?

а) Колмогоров доказал, что

$$\lim_{n \to \infty} P\{\sqrt{n}\, Z(n) < t\} = K(t), \quad t > 0,$$

где $$K(t) = \sum_{j=-\infty}^{+\infty} (-1)^j e^{-2k^2 t^2}, \quad t \in R.$$

Это означает, что посл-ть с.в. $\sqrt{n}\, Z(n)$ слабо сходится к с.в. A, ф-я распр. которой

$$F_A(t) = \begin{cases} 0, & t \le 0 \\ K(t), & t > 0 \end{cases}$$

б) Т.о. при достаточно больших n квантиль $\delta_{1-\alpha}$ распр-я сл.вел. $Z(n)$ приблизительно равна:

$$\delta_{1-\alpha} \approx \frac{a_{1-\alpha}}{\sqrt{n}}, \quad (*)$$ где $a_{1-\alpha}$ – квантиль ур. $1-\alpha$ с.в. A.

Таблицы зн-й $K(t)$ составлены, их можно найти в лит-ре.

6) Как показывает практика, приближ. соотнош. (*) можно пользоваться при $n \ge 20$.

3 Критерий χ^2 для проверки простой гипотезы.

Пусть 1) X – дискр. с.в.

2) X может принимать конечное мн-во значений $a_1, \dots, a_l$ с неизв. вероятностями $p_1, \dots, p_l$ соотв-но.

Требуется проверить гипотезу $H_0 = \{p_1 = p_1^0, \dots, p_l = p_l^0\}$

против $H_1 = \bar{H}_0 = \{(\exists j \in \{1, \dots, l\})(p_j \ne p_j^0)\}$.

где $p_1^0 + \ldots + p_l^0 = 1$,

$p_i^0 \geq 0$, $i = \overline{1;l}$ - предполагаемые вер-ти событий

$P\{X = a_i\}$.

Для решения этой задачи исп-ся критерий χ^2.

"ЗЫ ЁНД"