Лекция 7
(продолжение)

Проверка статистических гипотез

Критерий согласия хи-квадрат Пирсона

Разработан первоначально для дискретных распределений:

$$\begin{array}{c|cccc} X & x_1 & \cdots & x_L \\ \hline P & p_1 & \cdots & p_L \end{array}$$

Статистический ряд:

$$\begin{array}{c|cccc} X & x_1 & \cdots & x_L \\ \hline \nu & \nu_1 & \cdots & \nu_L \end{array} \qquad \nu_1 + \cdots + \nu_L = N$$

Нулевая гипотеза: исследуемая случайная величина имеет заданный закон распределения.

Статистика критерия:

$$\chi^2 = \chi^2\left(X_1, X_2, \ldots, X_L\right) = \sum_{l=1}^{L} \frac{\left(v_l - Np_l\right)^2}{Np_l}$$

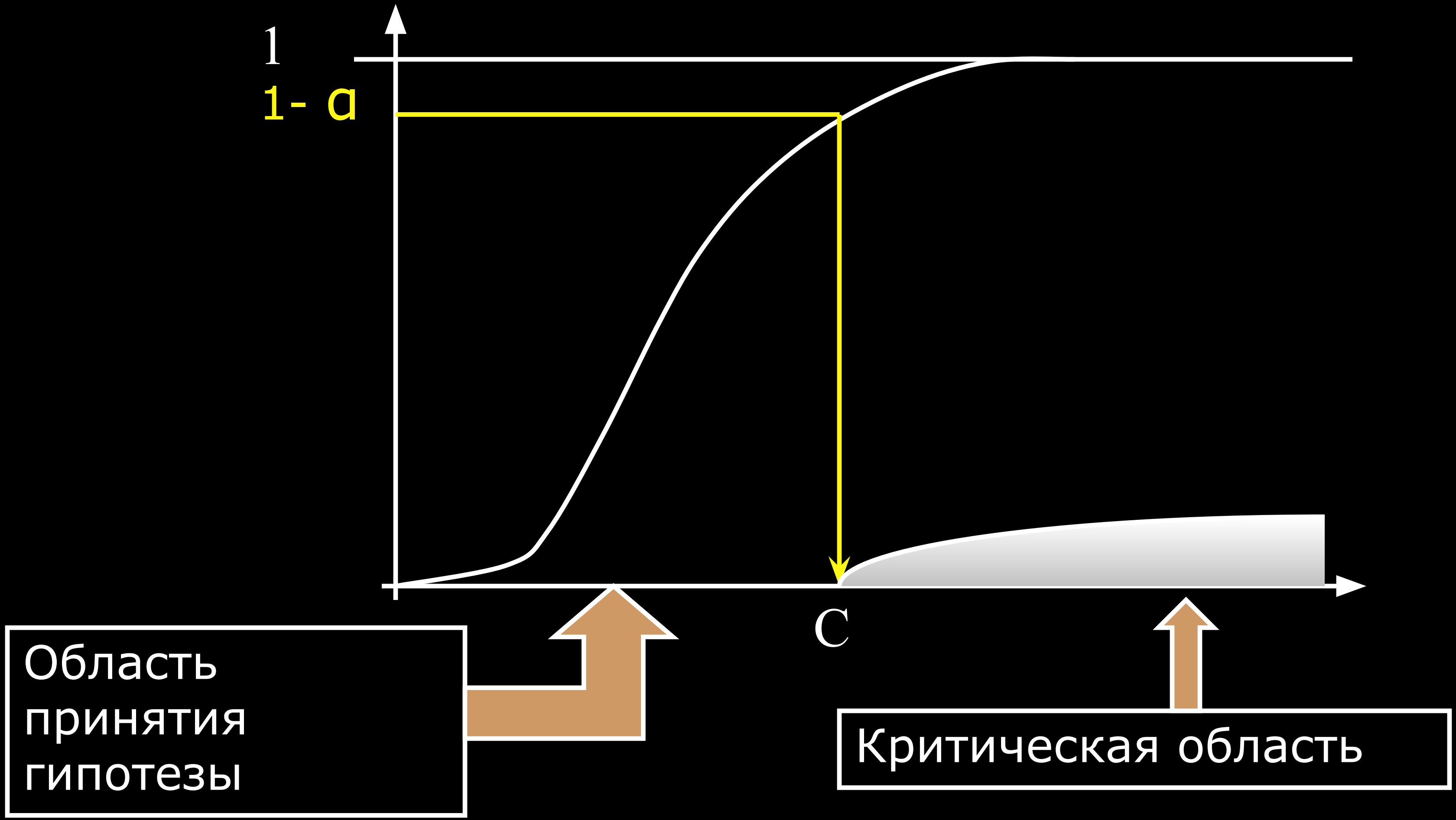
Является мерой близости теоретических вероятностей $\boldsymbol{P_l}$ и эмпирических (экспериментальных) частот v_l

Имеет асимптотическое (при n -->oo) распределение хи-квадрат.

Число степеней свободы равно:

- $\boldsymbol{L}$-1, если распределение полностью задано.
- L - 1 - r, если дополнительно оценивается r неизвестных параметров распределения.

Для нахождения критической области необходимо по заданной вероятности ошибки первого рода (уровню значимости критерия) α найти квантиль хи-квадрат распределения на уровне 1- α .



Подсчитываем значение статистики критерия и сравниваем его с критической точкой. Если

$$\chi^2 > C$$

То нулевая гипотеза отвергается.

В противном случае она принимается на уровне значимости α

Критерий легко приспосабливается и для непрерывных распределений путем их *дискретизации*.

Проверку гипотезы удобно совмещать с построением гистограмм.

Пять шагов проверки гипотезы

1. Сформулировать нулевую H_0 и альтернативную H_1 гипотезы.
2. Выбрать статистику критерия $T(X)$ и уяснить её закон распределения.
3. Задать уровень значимости критерия. По таблицам квантилей распределения статистики найти критические точки и указать критическую область.
4. Подсчитать значение статистики критерия и проверить условие попадания в критическую область.
5. Сделать вывод о принятии нулевой или альтернативной гипотезы.

Простейшие параметрические гипотезы

Гипотезы о среднем значении гауссовской случайной величины

Дано: Проведено две серии независимых испытаний одинакового объема, по результатам которых получены оценки математического ожидания a_0 и a_1.

Проверить нулевую гипотезу: $a_0 = a_1$.

Случай 1. Дисперсия известна и равна σ^2

Статистика критерия

$$a_1 = \overline{X_1}, \quad > \quad a_0 = \overline{X_0} \quad T = \frac{a_1 - a_0}{\sqrt{D\left[a_1 - a_0\right]}} = \frac{a_1 - a_0}{\sqrt{2\frac{\sigma^2}{n}}} = \frac{a_1 - a_0}{\sqrt{2}\sigma}\sqrt{n}$$

Имеет стандартное распределение

Выбор критической области зависит от вида альтернатив.

Альтернатива первая: $H_1:\ a_0 \neq a_1$