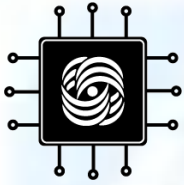


Имитационное моделирование в исследовании и разработке информационных систем

Лекция 5

Примеры систем моделирования
(продолжение)

Статистическая обработка результатов
эксперимента

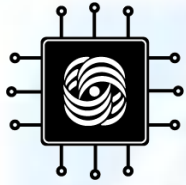


Откуда случайность?

- Натурные эксперименты и измерения – влияние внешних факторов
- Имитационные модели:
 - случайность потоков запросов
 - случайность действий (время, результат)

На выходе:

- последовательность результатов отдельных экспериментов;
- случайный процесс



Оценка случайной величины

$$\bar{X}(n) = \frac{\sum_{i=1}^n X_i}{n}$$

**оценка мат.
ожидания**

μ

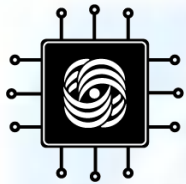
$$S^2(n) = \frac{\sum_{i=1}^n [X_i - \bar{X}(n)]^2}{n-1}$$

**Оценка
дисперсии**

σ^2

$$\hat{\text{Var}}[\bar{X}(n)] = \frac{S^2(n)}{n} = \frac{\sum_{i=1}^n [X_i - \bar{X}(n)]^2}{n(n-1)},$$

**Оценка
дисперсии
оценки мат.
ожидания**

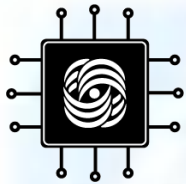


Сколько нужно экспериментов для оценки м.о.?

**Доверительный интервал длиной 2ε ,
в который μ укладывается с
вероятностью γ
задано γ, ε , найти n**

**Ц.П.Т.: Нормированная оценка м.о.
для n выборок сходится к**

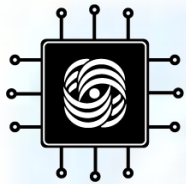
$$\Phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^z e^{-y^2/2} dy \quad \text{при } -\infty < z < \infty.$$



Оценка числа выборок (2)

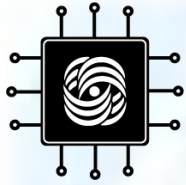
- Для нормированного распределения находим $u(\gamma)$ по таблице
- Далее, $\epsilon = u(\gamma) * \sqrt{\sigma/n}$
- Определяем n исходя из требований к ϵ

См. подробнее [1], с. 192



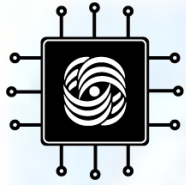
Если число выборок невелико

Если X_i – нормально распределённые,
то вместо таблицы нормального
распределения используем таблицу t -
распределения с $n-1$ степенями
свободы [3, с. 306]



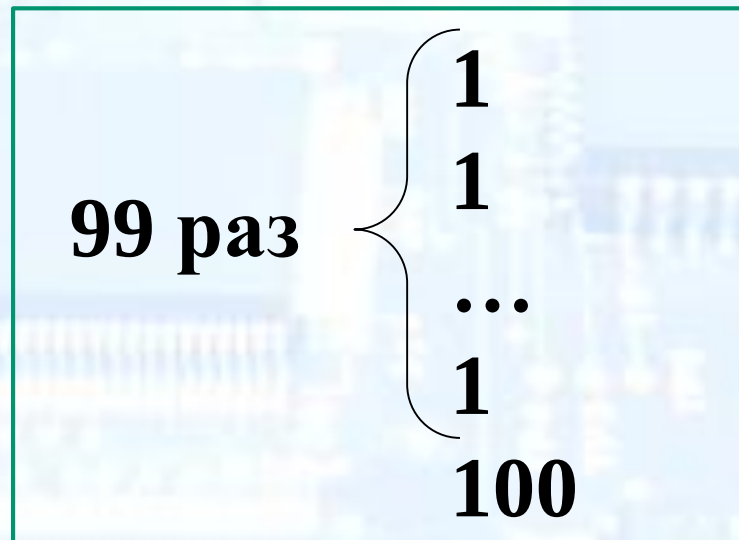
Проверка статистических гипотез

- По материалам диссертации П.Е. Шестова «Совместное планирование вычислений и обменов в информационно-управляющих системах реального времени», пункт «Методика статистической обработки результатов экспериментов»
- По учебнику [1]

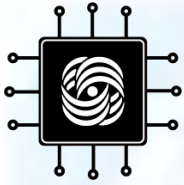


Результаты эксперимента

- 100 прогонов, замеряем "x"

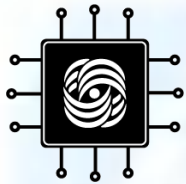


Как это обработать?
Какой вывод сделать из
полученных данных?



Возможные выводы

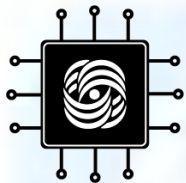
- В среднем $x=1.99$
 - ни в одном прогоне x не равнялся 1.99
 - почти 100% отклонение от 1
- В 99% случаях $x=1$
 - а если при дальнейших прогонах всегда $x=100$?
- Как сделать обобщённые выводы?



Статистическая гипотеза

- X – случайная величина
- Замеры x – *выборка* случайной величины
- Гипотеза:

С вероятностью p_0 значение x принадлежит отрезку $[x_{\min}; x_{\max}]$ ($x_{\min}$ и $x_{\max}$ уточняются по выборке)

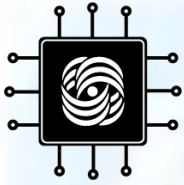


Гипотеза и альтернативная гипотеза

$p = P(x \in [x_{\min}; x_{\max}])$ – вероятность,
что x принадлежит отрезку

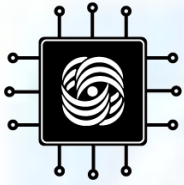
$$H_0 : p \geq p_0$$

$$H_1 : p < p_0$$



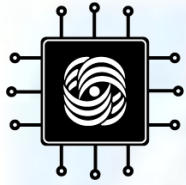
Уровень значимости

- α – уровень значимости или вероятность ошибки первого рода, т.е. вероятность, что гипотеза H_0 , будучи верной, будет отвергнута в пользу H_1
- Обычно $\alpha=0.05$
- Ошибка второго рода: принята H_1 , а на самом деле верна H_0



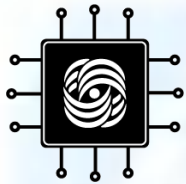
Если по-простому

- Статистически обосновывается, что с уровнем значимости 0.05 верна гипотеза, что с вероятностью не меньше 0.9 значение x лежит на заданном отрезке x



Статистический Критерий

- Зависит от выборки X
- Определяет «степень соответствия» выборки гипотезе
- Функция с известным распределением

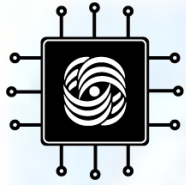


Критическая область

φ – критерий, ω – критическая область

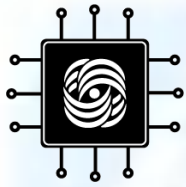
$$P(\varphi \in \omega) = \alpha.$$

Минимизация ошибки второго рода



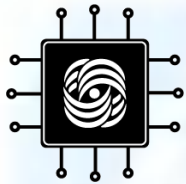
Примеры типовых стат. гипотез (по [1])

- значение МО нормального распределения при неизвестной дисперсии;
- равенство МО двух норм. распред.
- вид закона распределения случайной величины;



Статистический критерий

- m – число экспериментов, в которых
- $x \in [x_{\min}; x_{\max}]$
- $P(m = k) = C_n^k p^k (1 - p)^{n-k}$
- $\frac{m}{n}$ – «эмпирическая вероятность»



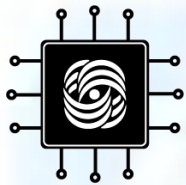
Критическая область

- Если критерий m принадлежит критической области, то H_0 отвергается

$$P(m < m_{\alpha}^{кр}) = \alpha$$

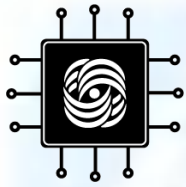
$$\sum_{i=0}^{m_{\alpha}^{кр}} \left(C_n^i p^i (1-p)^{n-i} \right) \leq \alpha$$

- $p = p_0$



Критическая область и границы отрезка

- Гипотеза H_0 принимается, если не менее $m_{\alpha}^{кр} + 1$ значений $x \in [x_{\min}; x_{\max}]$
- Теперь известно, какими свойствами должны обладать границы отрезка



Подбор границ отрезка

- Упорядочить элементы выборки x по возрастанию:

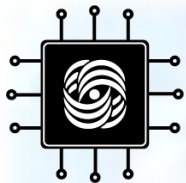
$$x_1, x_2, \dots, x_n$$

- Выбрать любые

$$x_i, x_j : j - i = m_{\alpha}^{kp}$$

- Обычно

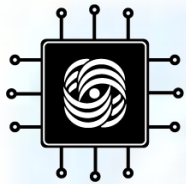
$$x_{\max} - x_{\min} \rightarrow \min$$



Примеры

- $p_0 = 0.9$
- $\alpha = 0.05$

n	$m^{кр}$
100	84
200	172
500	438
1000	883

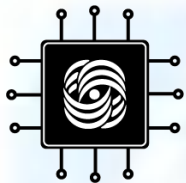


Ошибка второго рода

- β – вероятность ошибки второго рода, т.е. принять гипотезу H_0 тогда как верна

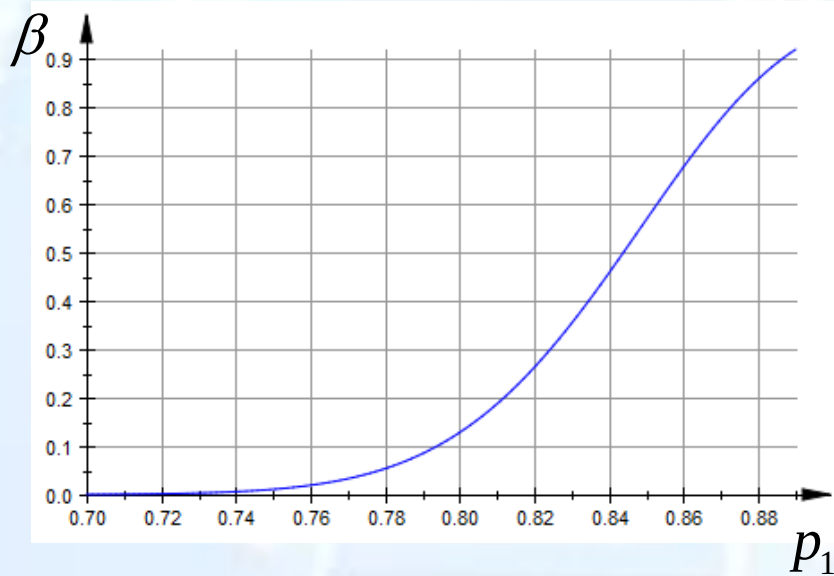
$$H_1 : p = p_1 < p_0$$

$$\beta = P(m > m_{\alpha}^{kp} \mid p = p_1) = \sum_{i=m_{\alpha}^{kp}+1}^n \left(C_n^i p_1^i (1-p_1)^{n-i} \right)$$

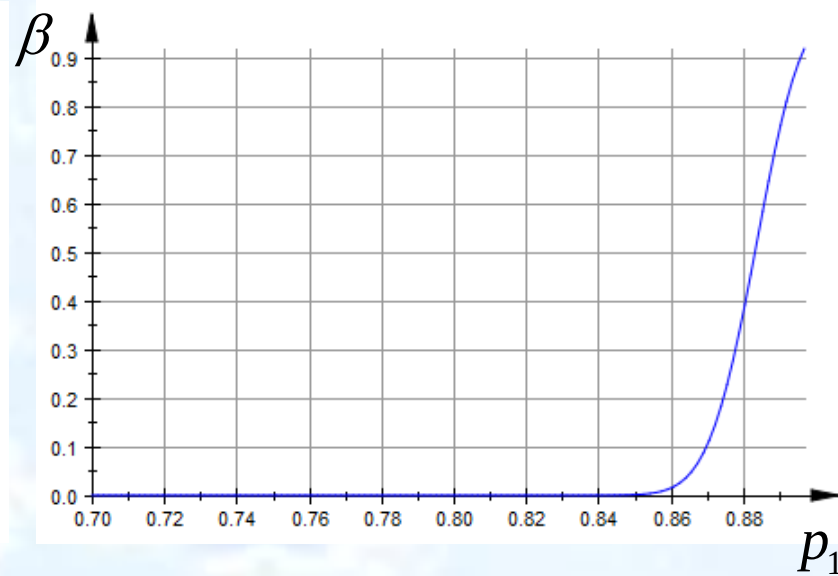


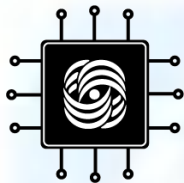
Пример

$n=100$



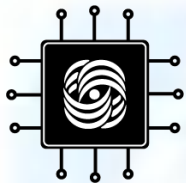
$n=1000$



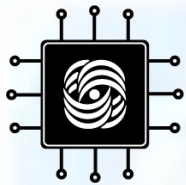


Литература

- Калинина В.Н., Панкин В.Ф.
Математическая статистика. М.:
Дрофа, 2002 год. 340 с.
- Гмурман В. Е. Теория вероятностей и
математическая статистика. М.:
Высшая школа, 2003. 479 с.
- Аверилл М.Лоу, В. Дэвид Кельтон.
Имитационное моделирование. 3-е
издание. // СПб:Питер, 2004. – 847 с.



Спасибо за внимание!



Применение стат. методов в имит. моделировании

- Проверка датчиков случайных чисел