УНИВЕРСИТЕТ ИТМО

Теория вероятностей и математическая статистика

Лекция 7. ***Проверка статистических гипотез***

Задачи статистической проверки гипотез

Одна из часто встречающихся на практике задач, связанных с применением статистических методов, состоит в решении вопроса о том, должно ли на основании данной выборки быть принято или, напротив, отвергнуто некоторое предположение (гипотеза) относительно генеральной совокупности (случайной величины).

Например, новое лекарство испытано на определенном числе людей. Можно ли сделать по данным результатам лечения обоснованный вывод о том, что новое лекарство более эффективно, чем применявшиеся ранее методы лечения? Аналогичный вопрос логично задать, говоря о новом правиле поступления в вуз, о новом методе обучения, о пользе быстрой ходьбы, о преимуществах новой модели автомобиля или технологического процесса и т. д.

Процедура сопоставления высказанного предположения (гипотезы) с выборочными данными называется *проверкой гипотез*.

Задачи статистической проверки гипотез

Задачи статистической проверки гипотез ставятся в следующем виде: относительно некоторой генеральной совокупности высказывается та или иная гипотеза H. Из этой генеральной совокупности извлекается выборка. Требуется указать правило, при помощи которого можно было бы по выборке решить вопрос о том, следует ли отклонить гипотезу H или принять ее.

Следует отметить, что статистическими методами гипотезу *можно только опровергнуть или не опровергнуть*, но не доказать. *Например*, для проверки утверждения (гипотеза H) автора, что «в рукописи нет ошибок», рецензент прочел (изучил) несколько страниц рукописи.

Если он обнаружил хотя бы одну ошибку, то гипотеза H отвергается, в противном случае – не отвергается, говорят, что «результат проверки с гипотезой согласуется».

Выдвинутая гипотеза может быть правильной или неправильной, поэтому возникает необходимость ее проверки.

Статистическая гипотеза

Под *статистической гипотезой* (или просто *гипотезой*) понимают всякое высказывание (предположение) о генеральной совокупности, проверяемое по выборке.

Статистические гипотезы делятся на гипотезы о параметрах распределения известного вида (это так называемые *параметрические* гипотезы) и гипотезы о виде неизвестного распределения (*непараметрические* гипотезы).

Одну из гипотез выделяют в качестве *основной* (или *нулевой*) и обозначают H_0, а другую, являющуюся логическим отрицанием H_0, т. е. противоположную H_0 — в качестве *конкурирующей* (или *альтернативной*) гипотезы и обозначают H_1.

Статистическая гипотеза

Гипотезу, однозначно фиксирующую распределение наблюдений, называют *простой* (в ней идет речь об одном значении параметра), в противном случае *сложной*.

Например, гипотеза H_0, состоящая в том что математическое ожидание с. в. X равно a_0, т. е. $MX = a_0$, является простой. В качестве альтернативной гипотезы можно рассматривать одну из следующих гипотез: H_1: $MX > a_0$ (сложная гипотеза), H_1: $MX < a_0$ (сложная), H_1: $MX \neq a_0$ (сложная) или H_1: $MX = a_1$ (простая гипотеза).

Имея две гипотезы H_0 и H_1, надо на основе выборки $X_1, \ldots, X_n$ принять либо основную гипотезу H_0, либо конкурирующую H_1.

Статистический критерий

Правило, по которому принимается решение принять или отклонить гипотезу H_0 (соответственно, отклонить или принять H_1), называется *статистическим критерием* (или просто *критерием*) проверки гипотезы H_0.

Проверку гипотез осуществляют на основании результатов выборки $X_1, X_2, \ldots, X_n$, из которых формируют функцию выборки $T_n = T(X_1, X_2, \ldots, X_n)$, называемой *статистикой критерия*.

Основной принцип проверки гипотез состоит в следующем. Множество возможных значений статистики критерия T_n разбивается на два непересекающихся подмножества: *критическую область* S, т. е. область отклонения гипотезы H_0 и *область* $\overline{S}$ *принятия* этой гипотезы. Если фактически наблюдаемое значение статистики критерия (т. е. значение критерия, вычисленное по выборке: $T_{\text{набл}} = T(x_1, x_2, \ldots, x_n)$) попадает в критическую область S, то основная гипотеза H_0 отклоняется и принимается альтернативная гипотеза H_1; если же $T_{\text{набл}}$ попадает в $\overline{S}$, то принимается H_0, а H_1 отклоняется.

Ошибки первого и второго рода

При проверке гипотезы может быть принято неправильное решение, т. е. могут быть допущены ошибки двух родов:

Ошибка первого рода состоит в том, что отвергается нулевая гипотеза H_0, когда на самом деле она верна.

Ошибка второго рода состоит в том, что отвергается альтернативная гипотеза H_1, когда она на самом деле верна.

Рассматриваемые случаи наглядно иллюстрирует следующая таблица.

Гипотеза H_0	Отвергается	Принимается
верна	ошибка 1-го рода	правильное решение
неверна	правильное решение	ошибка 2-го рода

Вероятность ошибки 1-го рода (обозначается через α) называется *уровнем значимости критерия*.

Очевидно, $\alpha = P(H_1|H_0)$. Чем меньше α, тем меньше вероятность отклонить верную гипотезу. Допустимую ошибку 1-го рода обычно задают заранее.

Мощность критерия

В одних случаях считается возможным пренебречь событиями, вероятность которых меньше 0,05 ($\alpha = 0{,}05$ означает, что в среднем в 5 случаях из 100 испытаний верная гипотеза будет отвергнута), в других случаях, когда речь идет, например, о разрушении сооружений, гибели судна и т. п., нельзя пренебречь обстоятельствами, которые могут появиться с вероятностью, равной 0,001.

Обычно для α используются стандартные значения: $\alpha = 0{,}05$; $\alpha = 0{,}01$; $0{,}005$; $0{,}001$.

Вероятность ошибки 2-го рода обозначается через β, т. е. $\beta = P(H_0|H_1)$.

Величину $1 - \beta$, т. е. вероятность недопущения ошибки 2-го рода (отвергнуть неверную гипотезу H_0, принять верную H_1), называется *мощностью критерия.*

Очевидно, $1 - \beta = P(H_1|H_1) = P\big((x_1, x_2, \ldots, x_n) \in S|H_1\big)$.

Чем больше мощность критерия, тем вероятность ошибки 2-го рода меньше, что, конечно, желательно (как и уменьшение α).

Ошибки первого и второго рода

Последствия ошибок 1-го, 2-го рода могут быть совершенно различными: в одних случаях надо минимизировать α, в другом — β. Так, применительно к радиолокации говорят, что α — вероятность пропуска сигнала, β — вероятность ложной тревоги; применительно к производству, к торговле можно сказать, что α — риск поставщика (т. е. забраковка по выборке всей партии изделий, удовлетворяющих стандарту), β — риск потребителя (т. е. прием по выборке всей партии изделий, не удовлетворяющей стандарту); применительно к судебной системе, ошибка 1-го рода приводит к оправданию виновного, ошибка 2-го рода — осуждению невиновного.

Отметим, что *одновременное уменьшение ошибок 1-го и 2-го рода возможно лишь при увеличении объема выборок.* Поэтому обычно при заданном уровне значимости α отыскивается критерий с наибольшей мощностью.

Методика проверки гипотез

Методика проверки гипотез сводится к следующему:

1. Располагая выборкой $X_1, X_2, \ldots, X_n$, формируют нулевую гипотезу H_0 и альтернативную H_1.
2. В каждом конкретном случае подбирают статистику критерия $T_n = T(X_1, X_2, \ldots, X_n)$, обычно из перечисленных ниже: U — нормальное распределение, χ^2 — распределение хи-квадрат (Пирсона), t — распределение Стьюдента, F — распределение Фишера-Снедекора.
3. По статистике критерия T_n и уровню значимости α определяют критическую область S (и $\overline{S}$). Для ее отыскания достаточно найти критическую точку $t_{\text{кр}}$, т. е. границу (или квантиль), отделяющую область S от $\overline{S}$.

Методика проверки гипотез

Границы областей определяются, соответственно, из соотношений: $P(T_n > t_{кр}) = \alpha$, для правосторонней критической области S (рис. 76); $P(T_n < t_{кр}) = \alpha$, для левосторонней критической области S (рис. 77); $P(T_n < t_{кр}^{л}) = P(T_n > t_{кр}^{п}) = \dfrac{\alpha}{2}$, для двусторонней критической области S (рис. 78).

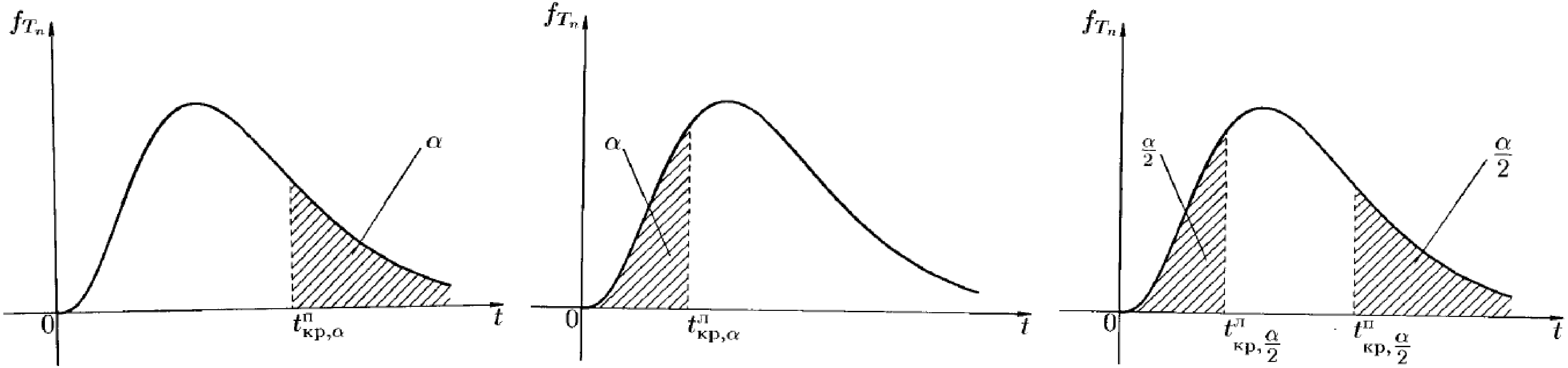

Рис. 76 *Рис. 77* *Рис. 78*

Методика проверки гипотез

Для каждого критерия имеются соответствующие таблицы, по которым и находят критическую точку, удовлетворяющую приведенным выше соотношениям.

4. Для полученной реализации выборки $x = (x_1, x_2, \ldots, x_n)$ подсчитывают значение критерия, т. е. $T_{\text{набл}} = T(x_1, x_2, \ldots, x_n) = t$.

5. Если $t \in S$ (например, $t > t_{\text{кр}}$ для правосторонней области S), то нулевую гипотезу H_0 отвергают; если же $t \in \overline{S}$ $(t < t_{\text{кр}})$, то нет оснований, чтобы отвергнуть гипотезу H_0.

Проверка гипотез о законе распределения

Во многих случаях закон распределения изучаемой случайно величины неизвестен, но есть основания предположить, что он имеет вполне определенный вид: нормальный, биномиальный или какой-либо другой.

Пусть необходимо проверить гипотезу H_0 о том, что с. в. X подчиняется определенному закону распределения, заданному функцией распределения $F_0(x)$, т. е. H_0: $F_X(x) = F_0(x)$. Под альтернативной гипотезой H_1 будем понимать в данном случае то, что просто не выполнена основная (т. е. H_1: $F_X(x) \neq F_0(x)$).

Для проверки гипотезы о распределении случайной величины X проведем выборку, которую оформим в виде статистического ряда:

x_i	x_1	x_2	...	x_m
n_i	n_1	n_2	...	n_m

(*)

где $\sum_{i=1}^{m} n_i = n$ — объем выборки.

Проверка гипотез о законе распределения

Требуется сделать заключение: согласуются ли результаты наблюдений с высказанным предположением. Для этого используем специально подобранную величину — критерий согласия.

Критерием согласия называют статистический критерий проверки гипотезы о предполагаемом законе неизвестного распределения. (Он используется для проверки согласия предполагаемого вида распределения с опытными данными на основании выборки.)

Существуют различные критерии согласия: Пирсона, Колмогорова, Фишера, Смирнова и др.

Критерий согласия Пирсона — наиболее часто употребляемый критерий для проверки простой гипотезы о законе распределения.

Критерий χ^2 Пирсона

Для проверки гипотезы H_0 поступают следующим образом.

Разбивают всю область значений с. в. X на m интервалов Δ_1, $\Delta_2, \ldots, \Delta_m$ и подсчитывают вероятности p_i $(i = 1, 2, \ldots, m)$ попадания с. в. X (т. е. наблюдения) в интервал Δ_i, используя формулу $P\{\alpha \leqslant X \leqslant \beta\} = F_0(\beta) - F_0(\alpha)$. Тогда теоретическое число значений с. в. X, попавших в интервал Δ_i, можно рассчитать по формуле $n \cdot p_i$.

Таким образом, имеем статистический ряд распределения с. в. X (*) и теоретический ряд распределения:

Δ_1	Δ_2	$\ldots$	Δ_m
$n'_1 = np_1$	$n'_2 = np_2$	$\ldots$	$n'_m = np_m$

Если эмпирические частоты (n_i) сильно отличаются от теоретических $(np_i = n'_i)$, то проверяемую гипотезу H_0 следует отвергнуть; в противном случае — принять.

Критерий χ^2 Пирсона

Каким критерием, характеризующим степень расхождения между эмпирическими и теоретическими частотами, следует воспользоваться? В качестве меры расхождения между n_i и np_i для $i = 1, 2, \ldots, m$

К. Пирсон (1857–1936; англ. математик, статик, биолог, философ) предложил величину («критерий Пирсона»):

$$\chi^2 = \sum_{i=1}^{m} \frac{(n_i - np_i)^2}{np_i} = \sum_{i=1}^{m} \frac{n_i^2}{np_i} - n. \qquad (**)$$

Согласно теореме Пирсона, при $n \to \infty$ статистика $(**)$ имеет χ^2-распределение с $k = m - r - 1$ степенями свободы, где m — число групп (интервалов) выборки, r — число параметров предполагаемого распределения. В частности, если предполагаемое распределение нормально, то оценивают два параметра (a и σ), поэтому число степеней свободы $k = m - 3$.

Критерий χ^2 Пирсона

Правило применения критерия χ^2 сводится к следующему:

1. По формуле (**) вычисляют $\chi^2_{\text{набл}}$ — выборочное значение статистики критерия.
2. Выбрав уровень значимости α критерия, по таблице χ^2-распределения находим критическую точку (квантиль) $\chi^2_{\alpha,k}$.
3. Если $\chi^2_{\text{набл}} \leqslant \chi^2_{\alpha,k}$, то гипотеза H_0 не противоречит опытным данным; если $\chi^2_{\text{набл}} > \chi^2_{\alpha,k}$, то гипотеза H_0 отвергается.

Необходимым условием применения критерия Пирсона является наличие в каждом из интервалов не менее 5 наблюдений (т. е. $n_i \geqslant 5$). Если в отдельных интервалах их меньше, то число интервалов надо уменьшить путем объединения (укрупнения) соседних интервалов.

Критерий χ^2 Пирсона

Пример Измерены 100 обработанных деталей; отклонения от заданного размера приведены в таблице:

$[x_i, x_{i+1})$	$[-3,-2)$	$[-2,-1)$	$[-1,0)$	$[0,1)$	$[1,2)$	$[2.3)$	$[3,4)$	$[4,5)$
n_i	3	10	15	24	25	13	7	3

Проверить при уровне значимости $\alpha = 0{,}01$ гипотезу H_0 о том, что отклонения от проектного размера подчиняются нормальному закону распределения.

Число наблюдений в крайних интервалах меньше 5, поэтому объединим их с соседними. Получим следующий ряд распределения ($n = 100$):

$[x_i, x_{i+1})$	$[-3,-1)$	$[-1,0)$	$[0,1)$	$[1,2)$	$[2,3)$	$[3,5)$
n_i	13	15	24	25	13	10

Критерий χ^2 Пирсона

Случайную величину — отклонение — обозначим через X. Для вычисления вероятностей p_i необходимо вычислить параметры, определяющие нормальный закон распределения (a и σ). Их оценки вычислим по выборке: $\bar{x} = \frac{1}{100} \cdot (-2 \cdot 13 + (-0{,}5) \cdot 15 + \ldots + 4 \cdot 10) = 0{,}885 \approx 0{,}9$, $D_{\text{в}} = \frac{1}{100}(4 \cdot 13 + 0{,}25 \cdot 15 + \ldots + 16 \cdot 10) - (0{,}885)^2 \approx 2{,}809$, $\sigma \approx 1{,}676 \approx 1{,}7$.

Находим p_i $(i = \overline{1,6})$. Так как с. в. $X \sim N(a, \sigma)$ определена на $(-\infty, \infty)$, то крайние интервалы в ряде распределения заменяем, соответственно, на $(-\infty, -1)$ и $(3, +\infty)$. Тогда $p_1 = P\{-\infty < X < -1\} = \Phi_0\left(\frac{-1 - 0{,}9}{1{,}7}\right) - \Phi_0(-\infty) = \frac{1}{2} - \Phi_0(1{,}12) = 0{,}1314$. Аналогично получаем: $p_2 = 0{,}1667$, $p_3 = 0{,}2258$, $p_4 = 0.2183$, $p_5 = 0{,}1503$, $p_6 = p\{3 \leqslant X < \infty\} = \Phi_0(\infty) - \Phi_0\left(\frac{3 - 0{,}9}{1{,}7}\right) = 0{,}5 - \Phi_0(1{,}24) = 0{,}1075$.

Критерий χ^2 Пирсона

Полученные результаты приведем в следующей таблице:

$[x_i, x_{i+1})$	$(-\infty, -1)$	$[-1, 0)$	$[0, 1)$	$[1, 2)$	$[2, 3)$	$[3, \infty)$
n_i	13	15	24	25	13	10
$n' = np_i$	13,14	16.67	22,58	21,83	15,03	10,75

Вычисляем $\chi^2_{\text{набл}}$:

$$\chi^2_{\text{набл}} = \sum_{i=1}^{6} \frac{n_i^2}{np_i} - n = \left(\frac{13^2}{13,14} + \frac{15^2}{16,67} + \ldots + \frac{10^2}{10,75}\right) - 100 = 101{,}045 - 100,$$

т. е. $\chi^2_{\text{набл}} \approx 1{,}045$.

Находим число степеней свободы; по выборке рассчитаны два параметра, значит, $r = 2$. Количество интервалов 6, т. е. $m = 6$. Следовательно, $k = 6 - 2 - 1 = 3$. Зная, что $\alpha = 0{,}01$ и $k = 3$, по таблице χ^2-распределения находим $\chi^2_{\alpha,k} = 11.3$. Итак, $\chi^2_{\text{набл}} < \chi^2_{\alpha,k}$, следовательно, нет оснований отвергнуть проверяемую гипотезу.

Критерий Колмогорова

Критерий Колмогорова для простой гипотезы является наиболее простым критерием проверки гипотезы о виде закона распределения. Он связывает эмпирическую функцию распределения $F_n^*(x)$ с функцией распределения $F(x)$ непрерывной случайной величины X.

Пусть $x_1, x_2, \ldots, x_n$ — конкретная выборка из распределения с неизвестной непрерывной функцией распределения $F(x)$ и $F_n^*(x)$ — эмпирическая функция распределения. Выдвигается простая гипотеза H_0: $F(x) = F_0(x)$ (альтернативная H_1: $F(x) \neq F_0(x)$, $x \in \mathbb{R}$).

Сущность критерия Колмогорова состоит в том, что вводят в рассмотрение функцию

$$D_n = \max_{-\infty<x<\infty} |F_n^*(x) - F_0(x)|,$$

называемой *статистикой Колмогорова*, представляющей собой максимальное отклонение эмпирической функции распределения $F_n^*(x)$ от гипотетической (т. е. соответствующей теоретической) функции распределения $F_0(x)$.

Критерий Колмогорова

Колмогоров доказал, что при $n \to \infty$ закон распределения случайной величины $\sqrt{n} \cdot D_n$ независимо от вида распределения с. в. X стремится к *закону распределения Колмогорова*:

$$P\{\sqrt{n} \cdot D_n < x\} \to K(x),$$

где $K(x)$ — функция распределения Колмогорова, для которой составлена таблица, ее можно использовать для расчетов уже при $n \geqslant 20$:

α	0,1	0,05	0,02	0,01	0,001
x_0	1,224	1.358	1,520	1,627	1,950

Найдем D_0 такое, что $P(D_n > D_0) = \alpha$.

Рассмотрим уравнение $K(x) = 1-\alpha$. С помощью функции Колмогорова найдем корень x_0 этого уравнения. Тогда по теореме Колмогорова, $P\{\sqrt{n} \cdot D_n < x_0\} = 1 - \alpha$, $P\{\sqrt{n} \cdot D_n > x_0\} = \alpha$, откуда $D_0 = \dfrac{x_0}{\sqrt{n}}$.

Если $D_n < D_0$, то гипотезу H_0 нет оснований отвергать; в противном случае — ее отвергают.

Критерий Колмогорова

Пример Монету бросали 4040 раз (Бюффон). Получили $n_1 = 2048$ выпадений герба и $n_2 = 1992$ выпадений решки. Проверить, используя а) критерий Колмогорова; б) критерий Пирсона, согласуются ли эти данные с гипотезой H_0 о симметричности монеты ($\alpha = 0.05$).

Случайная величина X принимает два значения: $x_1 = -1$ (решка) и $x_2 = 1$ (герб). Гипотеза H_0: $P\{X = -1\} = P\{X = 1\} = \frac{1}{2}$.

а) По таблице распределения Колмогорова находим корень уравнения $K(x) = 1 - \alpha$ при $\alpha = 0{,}05$. Следует $x_0 = 1{,}358$. Тогда $D_0 = \frac{x_0}{\sqrt{n}} = \frac{1{,}358}{\sqrt{4040}} \approx 0{,}021$.

Критерий Колмогорова

Пример Монету бросали 4040 раз (Бюффон). Получили $n_1 = 2048$ выпадений герба и $n_2 = 1992$ выпадений решки. Проверить, используя а) критерий Колмогорова; б) критерий Пирсона, согласуются ли эти данные с гипотезой H_0 о симметричности монеты ($\alpha = 0.05$).

Для нахождения по выборке D_n строим функции $F_0(x)$ и $F_n^*(x)$ и вычисляем величину $D_n = \max\limits_x |F_n^*(x) - F_0(x)|$.

x_i	решка $x_1 = -1$	герб $x_2 = 1$
p_i	0,5	0,5

$$\longrightarrow F_0(x) = \begin{cases} 0, & \text{при } x \leqslant -1, \\ 0{,}5, & \text{при } -1 < x \leqslant 1, \\ 1, & \text{при } 1 < x. \end{cases}$$

x_i	решка $x_1 = -1$	герб $x_2 = 1$
n_i	1992	2048
p_i^*	$\sim 0{,}493$	$\sim 0{,}507$

$$\longrightarrow F_n^*(x) = \begin{cases} 0, & \text{при } x \leqslant -1, \\ 0{,}493, & \text{при } -1 < x \leqslant 1, \\ 1, & \text{при } 1 < x. \end{cases}$$

Критерий Колмогорова

Пример Монету бросали 4040 раз (Бюффон). Получили $n_1 = 2048$ выпадений герба и $n_2 = 1992$ выпадений решки. Проверить, используя а) критерий Колмогорова; б) критерий Пирсона, согласуются ли эти данные с гипотезой H_0 о симметричности монеты ($\alpha = 0.05$).

Максимальное отклонение $F_0(x)$ от $F_n^*(x)$ равно 0,007, т. е. $D_n = 0{,}007$. Поскольку $D_n < D_0$, то нет оснований отвергать гипотезу H_0; опытные данные согласуются с гипотезой H_0 о симметричности монеты.

б) Вычисляем статистику χ^2:

$$\chi^2_{\text{набл}} = \sum_{i=1}^{2} \frac{n_i^2}{np_i} - n = \frac{1992^2}{\frac{1}{2} \cdot 4040} + \frac{2048^2}{\frac{1}{2} \cdot 4040} - 4040 \doteq 0{,}776.$$

По таблице χ^2-распределения находим критическую точку $\chi^2_{\alpha,k} = \chi^2_{0{,}05;1} = 3{,}8$. Так как $\chi^2_{\text{набл}} < \chi^2_{0{,}05;1}$, то опытные данные согласуются с гипотезой о симметричности монеты. ●