

Математические основы искусственного интеллекта.

Проверка статистических гипотез

Солодушкин Святослав Игоревич

Кафедра вычислительной математики и компьютерных наук,
УрФУ имени первого Президента России Б.Н. Ельцина

Март 2022

Пусть изучается генеральная совокупность X , пусть $\vec{X}_n = (X_1, \dots, X_n)$ — случайная выборка из генеральной совокупности X , а $\vec{x}_n = (x_1, \dots, x_n)$ — выборка. Имея выборку, можно выдвинуть несколько взаимоисключающих гипотез о распределении генеральной совокупности, одну из которых следует предпочесть остальным.

Определение

Предположение о распределении генеральной совокупности X или параметрах этого распределения называется статистической гипотезой.

Примеры статистических гипотез

Примеры статистических гипотез:

- 1 генеральная совокупность распределена по нормальному закону;
- 2 математическое ожидание генеральной совокупности равно 100;
- 3 две случайные величины не коррелированы.

Эти гипотезы вытекают из следующих содержательных задач:

- 1 правда ли, что среднемесячный товарооборот с соседним государством составляет 500 млн руб.;
- 2 правда ли, что повышение деловой активности в Европе приводит к росту цен на российскую нефть.

Предположение, состоящее в том, что на Марсе есть жизнь, не является примером статистической гипотезы.

Наряду с основной гипотезой рассматривается противоречащая ей гипотеза. Нулевой (основной) называют выдвинутую гипотезу, ее обычно обозначают как H_0 . Конкурирующей (альтернативной) называют гипотезу, которая противоречит нулевой, ее обычно обозначают как H_1 .

Выдвинутая гипотеза нуждается в проверке на основе наблюдаемых значений (выборки).

Процедура проверки статистических гипотез — последовательность действий, позволяющих с той или иной степенью достоверности подтвердить или опровергнуть утверждение гипотезы.

Отвергая основную гипотезу, мы отдаем предпочтение альтернативной.

		Гипотеза, верная на самом деле	
		H_0	H_1
Принятая гипотеза	H_0	Верное решение	Ошибка II рода
	H_1	Ошибка I рода	Верное решение

В результате проверки возможно принятие двух верных решений и двух ошибочных:

- 1 гипотеза H_0 верна и ее приняли в результате проверки;
- 2 гипотеза H_0 неверна и ее отвергли в результате проверки, приняв H_1 ;
- 3 гипотеза H_0 верна, но в ходе проверки ее ошибочно отвергли, приняв H_1 ;
- 4 гипотеза H_0 неверна, но в ходе проверки ее ошибочно приняли.

Ошибки первого и второго рода

		Гипотеза, верная на самом деле	
		H_0	H_1
Принятая гипотеза	H_0	Верное решение	Ошибка II рода
	H_1	Ошибка I рода	Верное решение

Ошибка первого рода состоит в том, что будет отвергнута правильная гипотеза H_0 . Ошибка второго рода состоит в том, что будет принята неправильная гипотеза H_0 .

Практическая невозможность маловероятных событий

В основе проверки статистических гипотез лежит принцип практической невозможности маловероятных событий, который гласит: «В единичном испытании маловероятное событие не должно появиться».



Пример из жизни

В открытом море перекадывают груз с одного корабля на другой. Колесный робот ездит по палубе первого корабля и манипулятором («механической рукой») захватывает контейнеры с палубы второго корабля, перекадывая их к себе. Первоначально робот был хорошо откалиброван, но в процессе эксплуатации мог испортиться.

Из официальной документации известно, что сила, с которой манипулятор хватает груз, распределена по закону, близкому к нормальному, и составляет в среднем 600 Н. Ясно, что если прилагаемая сила будет много меньше 600 Н, то контейнер выскользнет из манипулятора и утонет в море. Если прилагаемая сила будет много больше 600 Н, то контейнер погнется.

На ежегодном техобслуживании проведено 25 замеров силы манипулятора. Результаты представлены в таблице. Аналитики желают проверить, согласуются ли данные фактических замеров с заявленными в документации характеристиками. При этом отклонения и в ту и в другую сторону опасны.

Пример из жизни. Данные

624	598	609	592	588
578	598	604	616	628
634	605	590	628	632
584	627	612	606	620
641	585	641	637	654

Это выборка x_1, x_2, \dots, x_n из генеральной совокупности X , т. е. результат наблюдения случайной выборки X_1, X_2, \dots, X_n .

Заявленное значение силы $m_0 = 600$.

Среднее значение силы по результатам 25 измерений равно $m_1 = 613$, т. е. $\bar{x} = m_1$.

Очевидно, что расчетное среднее больше 600, но возникает вопрос, насколько это значимо. Возможно, эта разница в 13 находится, как говорят, «в пределах статистической погрешности».

Этап 1. Формулировка основной и альтернативной гипотез

Гипотеза H_0 состоит в том, что $EX = m_0$.

Вопрос состоит в том, не противоречат ли наблюдаемые результаты выдвинутой гипотезе? То есть, правда ли то, что имеющее место отклонение в развиваемой силе ($\bar{x} = m_1$ вместо m_0) объясняется случайными факторами.

В качестве альтернативной гипотезы выбирается $EX \neq m_0$.

С содержательной точки зрения альтернативная гипотеза означает, что средняя сила, развиваемая манипулятором, отличается от 600 (возможно, в большую, а возможно, и в меньшую сторону¹).

¹Это будет важно, когда будем строить критическую область. В данном примере она будет двусторонняя.

Этап 2. Выбор статистического критерия

Рассмотрим специально подобранную вещественнозначную функцию $\varphi(\vec{X}_n)$. Будучи функцией от случайных величин, φ сама является случайной величиной, а значит, имеет функцию распределения. Выбирать вид φ надо так, чтобы ее точное или приближенное распределение было известно. Эта функция потребуется для проверки нулевой гипотезы.

Определение

Статистическим критерием (или просто критерием) называют случайную величину $\varphi(\vec{X}_n)$, которая служит для проверки нулевой гипотезы.

Распределение статистического критерия.

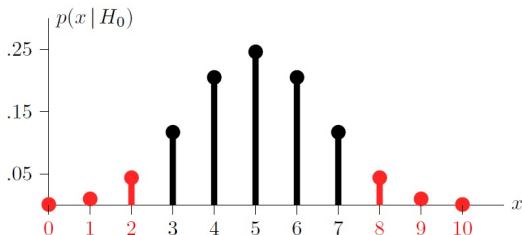
Иллюстрация

Монета с вероятностью выпадения орла p .

Критерий ϕ = количество орлов в 10 бросках.

H_0 : монета правильная, т. е. $p = 0.5$.

H_1 : монета неправильная, т. е. $p \neq 0.5$.



x	0	1	2	3	4	5	6	7	8	9	10
$p(x H_0)$.001	.010	.044	.117	.205	.246	.205	.117	.044	.010	.001

Если проверяют гипотезу о равенстве дисперсий s_1 и s_2 в двух нормальных генеральных совокупностях, то в качестве φ выбирают отношение исправленных выборочных дисперсий:

$$\varphi = \frac{s_1}{s_2}.$$

Поскольку в различных опытах выборочные дисперсии принимают различные значения, наперед неизвестно какие, то и их отношение будет величиной случайной. Можно показать, что φ будет распределена по закону Фишера–Снедекора.

При проверке гипотезы о равенстве нулю коэффициента корреляции полагают

$$\varphi = r \frac{\sqrt{n-2}}{1-r^2},$$

которая при справедливости нулевой гипотезы имеет распределение Стьюдента с $n-2$ степенями свободы (здесь n — объем выборки, r — значение выборочного коэффициента корреляции).

Выбор статистического критерия

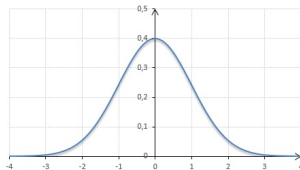
Можно рассмотреть функцию такого вида:

$$\varphi = \frac{X_1 + X_2 + \dots + X_n}{n} - 600,$$

которая распределена по нормальному закону и при условии справедливости нулевой гипотезы имеет математическое ожидание, равное нулю, здесь $n = 25$ — число измерений.

Исходя из математических соображений, рассмотрим величину:

$$\varphi = \frac{X_1 + X_2 + \dots + X_n - 600n}{n\sigma}.$$



Этап 3. Выбор уровня значимости

Нужно определить вероятность отправки на (ненужный) ремонт нормального робота. Взамен получим некоторую уверенность в том, что на корабль не поставят разболтанного робота.

Определение

Уровнем значимости называется вероятность ошибки первого рода, т. е. вероятность ошибочно отклонить верную нулевую гипотезу.

Уровень значимости обычно обозначают буквой α и берут малым, например, $\alpha = 0.05$ или $\alpha = 0.01$.

Этап 4. Построение критической области

После выбора определенного критерия и уровня значимости все множество возможных значений критерия разбивается на два непересекающихся подмножества: одно из них содержит значения критерия, при которых нулевая гипотеза отвергается, другое — значения, при которых она принимается.

Определение

Критической областью называют множество значений критерия, при которых нулевую гипотезу отвергают.

Для нахождения критической области достаточно найти ее границы, следовательно, возникает вопрос, как их найти.

Границы критической области

Находят границы критической области исходя из требования:

$$P(\varphi < \varphi_{\text{кр. лев}}) = \alpha/2, \quad P(\varphi > \varphi_{\text{кр. прав}}) = \alpha/2,$$

где $\varphi_{\text{кр. лев}}$ — левая граница критической области,

$\varphi_{\text{кр. прав}}$ — правая граница критической области.

В случае если распределение случайной величины φ симметрично относительно нуля, то $\varphi_{\text{кр. лев}} = -\varphi_{\text{кр. прав}}$, т. е. достаточно найти одну границу, которую будем обозначать $\varphi_{\text{кр}}$.

С учетом этого условие для нахождения границ критической области можно записать так:

$$P(\varphi < -\varphi_{\text{кр}}) + P(\varphi > \varphi_{\text{кр}}) = 2P(\varphi > \varphi_{\text{кр}}) = \alpha,$$

$$\int_{\varphi_{\text{кр}}}^{+\infty} f_{\varphi}(x) dx = \frac{\alpha}{2},$$

где $f_{\varphi}(x)$ — плотность распределения случайной величины φ .

$$\int_{\varphi_{\text{кр}}}^{+\infty} f_{\varphi}(x) dx = \frac{\alpha}{2},$$

где $f_{\varphi}(x)$ — плотность распределения случайной величины φ .

Для рассматриваемого примера данное нелинейное уравнение приближенно решим на компьютере и получим численное значение для $\varphi_{\text{кр}} = 2.06$.

Этап 5. Расчет наблюдаемого значения

В результате эксперимента случайная величина \vec{X}_n примет значение \vec{x}_n , а значит, функция $\varphi(\vec{X}_n)$ примет значение $\varphi_{\text{набл}} = \varphi(x_n)$.

Определение

Наблюдаемым значением критерия $\varphi_{\text{набл}}$ называют значение критерия, вычисленное по выборке.

Если наблюдаемое значение критерия $\varphi_{\text{набл}}$ попадает в критическую область, нулевую гипотезу отвергают как противоречащую экспериментальным данным. Если наблюдаемое значение критерия попадает в область принятия гипотезы, нет оснований отвергать нулевую гипотезу.

Этап 5. Расчет наблюдаемого значения

Если в результате 25 наблюдений отклонение средней силы окажется настолько большим, что $|\varphi_{\text{набл}}| > \varphi_{\text{кр}}$, то гипотезу H_0 надо отвергнуть. Если же отклонение средней силы окажется таким, что $|\varphi_{\text{набл}}| \leq \varphi_{\text{кр}}$, то не будет оснований отвергнуть гипотезу H_0 .

Пусть в данном примере аналитик получил $\varphi_{\text{набл}} = 3.87$. Это означает, что отклонения статистически значимы. Таким образом, аналитик отвергает нулевую гипотезу о том, что робот откалиброван хорошо, и делает заключение, что робот развивает силу более 600 Н.

При заданном уровне значимости α можно построить разные критические области. Мы в разобранном в предыдущем пункте примере построили ее симметричной относительно нуля, при этом никак не обосновывали такое решение. Целесообразно ввести в рассмотрение вероятность попадания критерия в критическую область при условии, что основная гипотеза неверна и справедлива конкурирующая гипотеза.

Определение

Мощностью критерия называют вероятность попадания критерия в критическую область при условии, что справедлива конкурирующая гипотеза.

Другими словами, мощность критерия есть вероятность того, что основная гипотеза будет отвергнута, если верна конкурирующая гипотеза.

Пусть для проверки гипотезы принят определенный уровень значимости и выборка имеет определенный фиксированный объем. Покажем, что строить критическую область надо так, чтобы мощность критерия была максимальной.

Если вероятность ошибки второго рода (принять неверную гипотезу H_0) равна β , то мощность критерия равна $1 - \beta$.

Пусть мощность $1 - \beta$ критерия возрастает, следовательно, уменьшается вероятность β совершить ошибку второго рода, что, конечно, желательно.

Чем меньше вероятность ошибок первого и второго рода, тем критическая область «лучше». Однако при заданном объеме выборки уменьшить одновременно и α и β невозможно: если уменьшать α , то β будет возрастать.

Например, если принять $\alpha = 0$, то будут приниматься все гипотезы, в том числе и неправильные, т. е. возрастает вероятность ошибки второго рода.

Единственный способ одновременного уменьшения вероятности ошибок первого и второго рода состоит в увеличении объема выборки.

Часто приходится слышать: «Если уровень значимости меньше 0.05, то нулевую гипотезу отвергаем».

Это высказывание лишено смысла, так как уровень значимости — это величина, которая фиксируется самим исследователем, а не вычисляется на основе выборочных данных.

Статистические гипотезы бывают простыми и сложными. Статистическая гипотеза, однозначно определяющая распределение генеральной совокупности, называется простой. Статистическая гипотеза, утверждающая принадлежность распределения к некоторому семейству распределений, называется сложной.

- 1 Верно ли утверждение. Мощность критерия есть вероятность того, что не будет допущена ошибка второго рода.
- 2 При проверке статистических гипотез назначается уровень значимости. Уровнем значимости называется вероятность ошибки первого рода, т. е. вероятность ошибочно отклонить верную нулевую гипотезу. Студент Василий не изучал статистику и рекомендует взять уровень значимости равным нулю, чтобы ошибок при проверке гипотез не было вовсе. К каким негативным последствиям приведет такой выбор?

В учебном пособии по спортивной метрологии В. В. Афанасьева² находим следующий вопрос для самоконтроля: «Укажите, при каком уровне значимости принимается гипотеза H_0 ». И далее предложены варианты:

1) $\alpha \geq 0.05$; 2) $\alpha \leq 0.05$; 3) $\alpha \geq 0.5$; 4) $\alpha \leq 0.5$.

В ответах находим:

«Если $\alpha \geq 0.05$, то принимается гипотеза H_0 ».

Объясните, почему этот вопрос некорректно сформулирован и, по сути, лишен смысла. Как его исправить?

²Спортивная метрология : учебник для среднего проф. образования / В. В. Афанасьев, И. А. Осетров, А. В. Муравьев, П. В. Михайлов ; отв. ред. В. В. Афанасьев. 2-е изд., испр. и доп. М. : Юрайт, 2017.

Проведено три эксперимента при уровне значимости $\alpha = 0.05$:

- в первом эксперименте найдено $p\text{-value} = 0.02$, нулевая гипотеза H_0 отвергнута;
- во втором эксперименте найдено $p\text{-value} = 0.049$, нулевая гипотеза H_0 отвергнута;
- в третьем эксперименте найдено $p\text{-value} = 0.15$, нулевая гипотеза H_0 принята.

В каком из этих трех случаев вероятность верного решения наибольшая?

Предположим, журнал будет публиковать только те результаты, которые статистически значимы на уровне 0,05.

Какой процент публикуемых им статей содержит ошибки первого рода?

Алекс разрабатывает новые лекарства, но все его лекарства неэффективны. Однако он тщательный статистик, поэтому случайным образом делит своих пациентов на контрольную и лечебную группы. Контрольная группа получает плацебо, а экспериментальная группа получает экспериментальное лечение. Его нулевая гипотеза H_0 состоит в том, что лечение не лучше плацебо. Он использует уровень значимости $\alpha = 0,05$. Если его р-значение меньше α , он публикует статью, в которой утверждается, что лечение значительно лучше, чем плацебо.

(а) Поскольку его методы лечения никогда не бывают эффективными, какой процент его экспериментов приводит к опубликованным статьям?

(б) Какой процент его опубликованных статей содержит ошибки первого рода? то есть описать лечение, которое не лучше, чем плацебо?

Боб — гениальный фармацевт, поэтому все предложенные им лекарства эффективны. Он также внимательный ученый и статистик. так что он тоже проводит двойные слепые, плацебо-контролируемые, рандомизированные исследования. Его нулевая гипотеза состоит в том, что новое лечение не лучше плацебо. Он также использует уровень значимости $\alpha = 0,05$ и публикует статью, если $p < \alpha$.

- (a) Какой процент его экспериментов приводит к публикациям?
- (b) Какой процент его опубликованных статей содержит ошибки типа I, т. е. описывает лечение, которое не лучше плацебо?

Две монеты: вероятность выпадения орла 0,5 для C_1 и 0,6 для C_2 .

Мы выбираем одну наугад, переворачиваем ее 8 раз и получаем 6 решек.

1. $H_0 = \text{'Монета } C_1'}$ $H_0 = \text{'Монета } C_2$.

Вы отвергаете H_0 на уровне значимости $\alpha = 0,05$?

2. $H_0 = \text{'Монета } C_2'}$ $H_1 = \text{'Монета } C_1'$.

Вы отвергаете H_0 на уровне значимости $\alpha = 0,05$?

3. Ваши ответы на (1) и (2) кажутся парадоксальными?

k	0	1	2	3	4	5	6	7	8
$p = 0.5$.004	.031	.109	.219	.273	.219	.109	.031	.004
$p = 0.6$.001	.008	.041	.124	.232	.279	.209	.090	.017