Теория Вероятностей и Статистика

Теория проверки статистичских гипотез

Потанин Богдан Станиславович

старший преподаватель, кандидат экономических наук

2021-2022

- Часто у исследователей возникает мотивация проверить некоторую гипотезу, например, о том, что математическое ожидание зарплаты случайно взятого жителя составляет пятьдесят тысяч рублей.
- Интуиция подсказывает, что мы должны отвергать гипотезу, если наши данные плохо с ней согласуются.
- Например, если наша гипотеза заключается в том, что средняя зарплата по стране равняется пятидесяти тысячам рублей, но по результатам опроса большого числа респондентов их средняя зарплата оказалась равна ста тысячам рублей, то у нас может возникнуть сомнение в верности изначального предположения.
- Рассмотрим формальный статистический аппарат, унифицирующий принцип тестирования различного вида гипотез.

Статистический тест

Нулевая и альтернативная гипотезы

- Имеется выборка $X = (X_1, ..., X_n)$, из распределения $\mathcal{D}_X$.
- Рассмотрим множество распределений $\mathcal{D}$ такое, что $\mathcal{D}_X \in \mathcal{D}$.
- Требуется проверить соблюдение **нулевой гипотезы** о том, что $H_0 : \mathcal{D}_X \in \mathcal{D}_0$, где $\mathcal{D}_0 \subset \mathcal{D}$.
- В качестве **альтернативной гипотезы** предполагают, что $H_1 : \mathcal{D}_X \in \mathcal{D}_1$, где $\mathcal{D}_1 = \mathcal{D} - \mathcal{D}_0$.
- Если $\mathcal{D}_0$ состоит из одного элемента, то гипотеза H_0 называется **простой**, а в противном случае - **сложной**. По аналогии для альтернативной гипотезы H_1.
- Для наглядности множества H_0 и H_1 через ограничения на параметры распределений D, задающие множества D_0 и D_1. Такие гипотезы называются **параметрическими**.
  Пример:
- Пусть D состоит из всех нормальных распределений, что, пользуясь одинаковой распределенностью элементов выборки, можно сформулировать как $X_1 \sim \mathcal{N}(\mu, \sigma^2)$.
- Если $D_0 = \{N(0, 1)\}$, то нулевая гипотеза является простой, что можно сформулировать как $H_0 : X_1 \sim N(0, 1)$ или параметрически $H_0 : \mu = 0 \wedge \sigma^2 = 1$.
- Альтернативная гипотеза в таком случае будет сложной, поскольку D_1 включает все нормальные распределения, кроме стандартного нормального, что удобно сформулировать как параметрическую гипотезу $H_1 : \mu \neq 0 \vee \sigma^2 \neq 1$.
- Если D_0 состоит из всех нормальных распределений, у которых математическое ожидание равно дисперсии, то гипотеза является сложной и может быть сформулирована как $H_0 : X_1 \sim \mathcal{N}(\mu, \mu)$, или $H_0 : \mu = \sigma^2$, где $\mu > 0$. Альтернативная гипотеза в данном случае также будет сложной $H_1 : \mu \neq \sigma^2$.

- Принятие и отвержение нулевой гипотезы H_0 осуществляется в соответствии с правилом, именуемым **статистическим критерием**.
- Через x^s обозначим **выборочное пространство** – множество n-мерных векторов, состоящее из всех возможных реализаций выборки.
- Обозначим через $x^{(1)}$ **критическую область** – подмножество выборочного пространства, при попадании реализации выборки в которое нулевая гипотеза отвергается. То есть $x^{(1)} \in x^s$ и из $x \in x^{(1)}$ следует, что H_0 отвергается.
- Через $x^{(0)} = x^s - x^{(1)}$ обозначим **область принятия** нулевой гипотезы.
  Пример:
- Лаврентий подкидывает монетку три раза и проверяет простую гипотезу о том, что вероятности выпадения орла (обозначим как 1) и решки (обозначим как 0) совпадают. Лаврентий отвергает нулевую гипотезу, если все три раза монетка падает одной и той же стороной.
- У Лаврентия есть выборка $X = (X_1, X_2, X_3)$ из биномиального распределения $D = \{B(p) : p \in (0, 1)\}$, то есть $X_1 \sim B(p)$. Нулевая гипотеза является простой и ее удобно сформулировать параметрически $H_0 : p = 0.5$. Альтернативна гипотеза является сложной $H_1 : p \neq 0.5$.
- Критическая область имеет вид $x^{(1)} = \{(0, 0, 0), (1, 1, 1)\}$. В результате, например, если у Лаврентия выпадают три решки, то есть $x = (0, 0, 0)$, то реализация попадает в критическую область, вследствие чего нулевая гипотеза отвергается. Если же, например, сперва выпадают две решки, а затем – орел, то $x = (0, 0, 1)$ и нулевая гипотеза не отвергается.

- Часто формулировать $x^{(1)}$ в явном виде бывает неудобно, поскольку критическая область состоит из n-мерных векторов.
- В качестве альтернативы можно найти такую статистику $T(X)$ и такое подмножество ее носителя $\mathcal{T}$, что $x \in x^{(s)}$ тогда и только тогда, когда $T(x) \in \mathcal{T}$.
- Статистика $T(X)$ именуется **статистикой критерия**, а $\mathcal{T}$ – **ее критической областью**.
  Пример:
- Вернемся к примеру с Лаврентием, трижды подбрасывающим монетку и проверяющим, что она является правильной.
- Зададим тот же самый статистический критерий через тестовую статистику $T(X) = X_1 + X_2 + X_3$. Для того, чтобы гарантировать $x^{(1)} = \{(0,0,0),(1,1,1)\}$ будем отвергать нулевую гипотезу, если $T(x) = 3$ или $T(x) = 0$, откуда $\mathcal{T} = \{0,3\}$.
- Например, если у Лаврентия выпало три орла, то $T(x) = 1 + 1 + 1 = 3$ и нулевая гипотеза отвергается. Если же выпал ровно один орел, то $T(x) = 1$ и нулевая гипотеза не отвергается.

Статистический тест

Ошибки первого и второго рода

- **Ошибка первого рода** возникает, когда (ошибочно) отвергается верная нулевая гипотеза. Ее вероятность рассчитывается как:
$$\alpha = P(X \in x^{(1)}|H_0) = P(T(X) \in \mathcal{T}|H_0)$$
- **Ошибка второго рода** возникает, когда (ошибочно) отвергается верная альтернативная гипотеза. Ее вероятность рассчитывается как:
$$\beta = P(X \notin x^{(1)}|H_1) = P(T(X) \notin \mathcal{T}|H_1)$$

Пример:
- Лаврентий трижды подкидывает монетку и тестирует простую гипотезу $H_0 : p = 0.5$ против простой альтернативы $H_1 : p = 0.6$, где p - вероятность выпадения орла.
- Будем использовать ту же тестовую статистику, обращая внимание, что $T(X)|H_0 \sim B(0.5)$ и $T(X)|H_1 \sim B(0.6)$.
- Ошибка первого рода произойдет, если монетка правильная, то есть $p = 0.5$, но она трижды выпадет одной стороной, вероятность чего составляет:
$$\alpha = P(T(X) \in \{0, 3\}|H_0) = P(T(X) = 0|H_0) + P(T(X) = 3|H_0) = 0.5^3 + 0.5^3 = 0.25$$
- Ошибка второго рода произойдет, если монетка неправильная, то есть $p = 0.6$, но она хоть раз выпадет другой стороной, что произойдет со следующей вероятностью:
$$\beta = P(T(X) \notin \{0, 3\}|H_1) = 1 - P(T(X) = 0|H_1) - P(T(X) = 3|H_1) = 1 - 0.6^3 - 0.6^3 = 0.568$$

- **Мощность критерия** равняется вероятности не совершить ошибку второго рода $1-\beta$.
- **Уровень значимости** критерия совпадает с вероятностью ошибки первого рода α.
- Как правило, уменьшать вероятность ошибки одного рода приходится за счет увеличения вероятности ошибки второго рода. Поэтому, исследователи пытаются найти такой критерий, что при фиксированном уровне значимости его мощность будет максимальной.
  Пример:
- Лаврентий трижды подкидывает монетку и тестирует простую гипотезу $H_0 : p = 0.5$ против простой альтернативы $H_1 : p = 0.6$, где p – вероятность выпадения орла.
- Уровень значимости критерия совпадает с посчитанной ранее вероятностью ошибки первого рода $\alpha = 0.25$
- Мощность критерия нетрудно вычислить используя посчитанную ранее вероятность ошибки второго рода $1 - \beta = 1 - 0.568 = 0.432$.

Статистический тест

Дополнительный пример

Время прочтения книги (в минутах) ученым котом является экспоненциальной случайной величиной. Ученый кот тестирует гипотезу о том, что математическое ожидание соответствующего времени равняется 10 минутам, против альтернативы о том, что оно равно 20 минутам. Ученый кот замерил время, понадобившееся ему для прочтения каждой из 10 книг. Он отвергает нулевую гипотезу, если чтение по крайней мере одной из книг заняло более получаса.

- Дайте параметрическую формулировку нулевой и альтернативной гипотез.
- Формализуйте статистический критерий ученого кота (вспомните экстремальные порядковые статистики).
- Найдите уровень значимости критерия ученого кота.
- Рассчитайте мощность критерия ученого кота.

Решение:

- Поскольку $E(X_1) = 1/\lambda$, то при верной нулевой гипотезе $E(X_1) = 20$, откуда $\lambda = 1/10 = 0.1$, то есть $H_0 : \lambda = 1/10 = 0.1$. По аналогии $H_1 : \lambda = 1/20 = 0.05$. Обе гипотезы – простые.
- Используется тестовая статистика $T(X) = \max(X_1, ..., X_n)$ с критической областью $\mathcal{T} = (30, \infty)$.
- Уровень значимости совпадает с вероятностью ошибки первого рода:

$$\alpha = P(\max(X_1, ..., X_{10}) > 30|H_0) = 1 - P(\max(X_1, ..., X_{10}) \leq 30|H_0) = 1 - \left(1 - e^{-0.1\times 30}\right)^{10} \approx 0.4$$

- Вычислим мощность критерия с помощью вероятности ошибки второго рода:

$$1 - \beta = 1 - P(\max(X_1, ..., X_{10}) \leq 30|H_1) = 1 - \left(1 - e^{-0.05\times 30}\right)^{10} \approx 0.92$$

- Будем называть **p-value** наименьший уровень значимости, при котором нулевая гипотеза не отвергается с учетом того, что реализация тестовой статистики оказалась равна $T(x)$.
- Обозначим через $\mathcal{T}_\alpha$ критическую область тестовой статистики, при которой уровень значимости теста равняется α, откуда:

$$\text{p-value} = \min_{\alpha : T(x) \notin \mathcal{T}_\alpha} \alpha$$

- Чем меньше p-value, тем уверенней можно отвергать нулевую гипотезу.
  Пример:
- Имеется выборка $X = (X_1, X_2)$ из Хи-квадрат распределения. Тестируется гипотеза $H_0 : X_1 \sim \chi^2(10)$. Тестовая статистика имеет вид $T(X) = X_1 + X_2$. Нулевая гипотеза отвергается, если $T(X) < k_\alpha$, то есть $\mathcal{T}_\alpha = (0, k_\alpha)$, где k_α зависит от уровня значимости α, который равняется:

$$\alpha = P(T(X) < k_\alpha | H_0) = P(X_1 + X_2 < k_\alpha | H_0) = F_{\chi^2(10\times 2)}(k_\alpha) = F_{\chi^2(20)}(k_\alpha)$$

  Из полученного выражения следует, что k_α является квантилью уровня α Хи-квадрат случайной величины с 20-ю степенями свободы. Пусть сумма наблюдений в выборке оказалась равна 15, то есть $T(x) = 15$. Тогда, учитывая, что k_α строго возрастает по α, получаем:

$$\text{p-value} = \min_{\alpha : 15 \notin (0, k_\alpha)} \alpha = \min_{\alpha : 15 \in [k_\alpha, \infty)} \alpha = \min_{k_\alpha : 15 \in [k_\alpha, \infty)} F_{\chi^2(20)}(k_\alpha) = \min_{k_\alpha = 15} F_{\chi^2(20)}(k_\alpha) = F_{\chi^2(20)}(15) \approx 0.22$$

Статистический тест

Графическая иллюстрация p-value

Выборка: $X = (X_1, X_2)$

Реализация выборки: $x = (1, 2)$

Тестовая статистика: $T(X) = X_1 + X_2$

Реализация тестовой статистики: $T(x) = 3$

Нулевая гипотеза: $H_0 : X_1 \sim \chi^2(1)$

Распределение статистики: $T(X)|H_0 \sim \chi^2(2)$

Критическая область: $\mathcal{T}_{p-value} = (3, \infty)$

p-value$=1 - F_{\chi^2(2)}(3) \approx 0.22$