Проверка статистических гипотез

Малов Сергей Васильевич

Санкт-Петербургский государственный электротехнический университет

24 октября 2020 г.

План

1. Постановка задачи проверки статистических гипотез
2. Статистические критерии
3. Построение статистического критерия

Статистическая гипотеза

Пусть $(\mathfrak{X}, \mathfrak{F}, \mathcal{P})$, $\{P_\theta : \theta \in \Theta\}$ – статистический эксперимент

Определение

Статистической гипотезой называется утверждение о параметре вида

$$H : \theta \in \Theta^* \subseteq \Theta.$$

- Одноточечная гипотеза $H_0 : \theta = \theta^*$, определяющая точное значение параметра, называется простой
    - $\theta^* \in \Theta$ – фиксированное значение параметра
- Другие гипотезы называются сложными

Постановка задачи

В классической постановке выдвигают:

(i) основную (или нулевую) гипотезу $H_0 : \theta \in \Theta_0$

(ii) альтернативную гипотезу (альтернативу) $H_A : \theta \in \Theta_A$: $\Theta \cap \Theta_A = \varnothing$.

Задача исследователя – по результатам наблюдений сделать выбор между основной гипотезой и альтернативой.

- С точки зрения модели удобно предполагать, что $\Theta_0 \cup \Theta_A = \Theta$
- На практике довольно часто данное условие ослабляют, предполагая наличие расстояния между $\mathrm{dist}(\Theta_0, \Theta_A) > 0$, чтобы обеспечить различимость H_0 и H_A.
- Основную гипотезу иногда называют нулевой
- Иногда выдвигают несколько взаимоисключающих гипотез, из которых требуется выбрать одну

Постановка задачи

Различные варианты истинного положения дел и решения исследователя:

	Принята H_0	Отвергнута H_0
H_0 верна	+	Ошибка I рода
H_0 не верна	Ошибка II рода	+

- Ошибка I рода – отвержение нулевой гипотезы при ее справедливости, т.е. принятие альтернативы
- Ошибка II рода – принятие нулевой гипотезы при справедливости альтернативы
- Если $\Theta_0 \cup \Theta_A \neq \Theta$, то возможно иное положение дел, при котором обе гипотезы не верны

Задача проверки значимости отклонения от основной гипотезы заключается лишь в нахождении несоответствия имеющихся статистических данных и основной гипотезы

- значимым результатом считается обнаружение несоответствия
- проверка значимости не подразумевает рассмотрения ошибки II рода

Типы задач проверки статистических гипотез

Тип задачи определяется

- имеющимися статистическими данных
- моделью эксперимента
- целями исследования

I. Проверка согласия

- Данные представляют собой однородный набор наблюдений
  - типичной моделью является выборка
- Гипотеза согласия ставится в терминах распределения отдельных наблюдений
  - простая: о согласии с некоторым фиксированным распределением
  - сложная: о принадлежности некоторому множеству распределений
- Соответствующая гипотеза называется гипотезой однородности

Типы задач проверки статистических гипотез

II. Проверка однородности

- Данные представляют собой
  - два или несколько однородных наборов наблюдений
  - неоднородный набор наблюдений распределения которых определяет ковариата
- Гипотеза устанавливает равенство распределений всех наблюдений
  - иными словами, утверждается однородность исходного набора наблюдений
  - при наличии нескольких выборок формулируется в терминах сравнений параметров
  - при наличии ковариаты устанавливает независимость распределения наблюдения от ковариаты
- Соответствующая гипотеза называется гипотезой однородности

Типы задач проверки статистических гипотез

III. Проверка независимости

- Данные представляют собой однородный набор многомерных наблюдений
  - каждое наблюдение представляет собой набор значений d признаков – вектор рамерности d
  - наиболее часто речь идет о выборке из d-мерного распределения, т.е. наблюдения независимы
- Гипотеза независимости утверждает независимость компонент d-мерного наблюдения

IV. Проверка случайности

- Данные представляют собой произвольный набор наблюдений одного типа
  - обычно наблюдения – случайные величины
- Гипотеза случайности устанавливает независимость и одинаковую распределенность наблюдений
  - иными словами, устанавливается, что исходный набор наблюдений - выборка

План

Статистический критерий

Определение

Статистическим критерием (тестом) называется статистика $\phi{:}\mathfrak{X}\to[0,1]$, определяющая вероятность отвергнуть основную гипотезу по результатам наблюдений.

- Критерий – правило, согласно которому принимается или отвергается основная гипотеза
  - доверительная область: $\{X\in\mathfrak{X}:\phi(X)=0\}$ – основная гипотеза принимается
  - критическая область: $\{X\in\mathfrak{X}:\phi(X)=1\}$ – основная гипотеза отвергается
  - область сомнений: $\{X\in\mathfrak{X}:\phi(X)=1\}$ – решение принимается в результате проведения испытания с вероятностью отвержения основная гипотезы $\phi(X)$ и вероятностью ее принятия $1-\phi(X)$.
- Критерий называется нерандомизованным, если результаты наблюдений однозначно определяют решение $\phi(\mathfrak{X})=\{0,1\}$
  - в остальных случаях критерий рандомизованный

Характеристики статистического критерия

- Статистический критерий определяет вероятности ошибок I и II рода
  - $P_\theta(\text{ош. I рода}) = \mathbb{E}_\theta\phi(X)$, $\theta \in \Theta_0$ – вероятность ошибки I рода
  - $P_\theta(\text{ош. II рода}) = 1 - \mathbb{E}_\theta\phi(X)$, $\theta \in \Theta_A$– вероятность ошибки II рода
- Мощность статистического критерия ϕ
  - $b_\phi(\theta) = 1 - P_\theta(\text{ош. II рода}) = \mathbb{E}_\theta\phi(X)$, $\theta \in \Theta_A$
- Уровень значимости статистического критерия ϕ
  - Классический подход Неймана–Пирсона заключается в том, чтобы ограничить ошибку I рода малым наперед заданным числом α
  - Значение α, ограничивающее вероятность ошибки I рода называется уровнем значимости критерия.
- Функция $b_\phi(\theta) = \mathbb{E}_\theta\phi(X)$ определяет
  - вероятность ошибки I рода при $\theta \in \Theta_0$
  - мощность критерия при $\theta \in \Theta_A$

Свойства статистического критерия

- Пусть ϕ – критерий уровня значимости α проверки основной гипотезы $H_0 : \theta \in \Theta_0$ при альтернативе $H_A : \theta \in \Theta_A$.
- Критерий несмещенный, если $\sup_{\theta \in \Theta_0} b_\phi(\theta) \leq \inf_{\theta \in \Theta_A} b_\phi(\theta)$
  - мощность не должна быть меньше вероятности ошибки I рода
- Качество статистического критерия при фиксированном уровне значимости определяется его мощностью
  - критерий ϕ называется равномерно наиболее мощным, если $b_\phi(\theta) \geq b_{\phi^*}(\theta)$ для любого $\theta \in \Theta_A$
  - равномерно-наиболее мощный критерий максимизирует мощность на множестве всех критериев при каждом значении параметра при справедливости альтернативы
  - наиболее мощного критерия обычно не существует в случае двухсторонней альтернативы
- Класс допустимых статистических критериев довольно часто сужают до несмещенных

Асимптотический подход

В асимптотической модели статистического эксперимента $(\mathfrak{X}_n, \mathfrak{F}_n, \mathcal{P}_n)$, $\mathcal{P}_n = \{P_{\theta,n}, \theta \in \Theta\}$ статистическим критерием называется по сути совокупность критериев $\phi = \{\phi_n\}_{n\in\mathbb{N}}$.

- Уровнем значимости асимптотического критерия называется число α: $$lim_{n\to\infty} \sup_{\theta\in\Theta_0} \mathbb{E}_\theta(\phi_n(X)) \geq 1 - \alpha$$
- Мощность асимптотического критерия также вычисляется асимптотически $$b_\phi(\theta) = \lim_{n\to\infty} \mathbb{E}_\theta(\phi_n(X)),\ \theta \in \Theta_A$$
- Асимптотический критерий несмещенный, если $\sup_{\theta\in\Theta_0} b_\phi(\theta) \leq \inf_{\theta\in\Theta_A} b_\phi(\theta)$
- Асимптотический критерий ϕ равномерно наиболее мощный, если для любого асимптотического критерия ϕ^*, если $b_\phi(\theta) \geq b_{\phi^*}(\theta)$ при каждом значении $\theta \in \Theta_A$.

Параметрический и непараметрический подходы

Различают параметрические и непараметрические гипотезы

- Гипотеза параметрическая, если она допускает представление в виде $H : \theta \in \Theta^* \subseteq \mathbb{R}^d$
- В остальных случаях гипотеза непараметрическая
- Параметрическая гипотеза может быть сформулирована и в непараметрической модели
  - формально, непараметрическая модель искусственно адаптируется к семипараметрической
- Тип критерия определяется типом основной гипотезы и альтернативы
  - критерий параметрический, если основная и альтернативная гипотезы параметрические
  - критерий непараметрический, если хотя бы одна из гипотез непараметрическая

План

1. Постановка задачи проверки статистических гипотез
2. Статистические критерии
3. Построение статистического критерия

Построение статистического критерия

Обычно в основе статистичекого критерия лежит статистика критерия.

- Статистика критерия T удовлетворяет следующим условиям
  - распределение $T(X)$ при основной гипотезе не зависит от параметра: $P_\theta(T(X) \in A) = P_0(T(X) \in A)$, для любого A, $\theta \in \Theta_0$.
  - это распределение изучено (существуют таблицы)
  - распределение $T(X)$ при альтернативе $\theta \in \Theta_A$ отличается от ее распределения при нулевой гипотезе
  - для асимптотического критерия все условия формулируют в терминах асимптотических распределений
- Выбираем набор множеств $\mathcal{I}$ и находим
$$I_\alpha \in \mathcal{I}\colon P_0(T \in I_\alpha) \geq (=) 1 - \alpha$$
- Получаем нерандомизованный критерий
$$\phi(X) = \begin{cases} 0, & T \in I_\alpha \\ 1, & T \notin I_\alpha \end{cases}$$
- Уровень значимости построенного критерия равен α

Построение статистического критерия

Для статиститческого критерия на базе статистики T

- P-значение определяется как наименьшее α, такое что $T \notin I_\alpha$.
- При положительном распределении T наиболее часто используется $I_\alpha = [0, x_\alpha]$
  - x_α удовлетворяет условию $P_{\theta_0}(T > x_\alpha) = 1 - F_T(x_\alpha) = \alpha$
    - F_T функция распределения T при нулевой гипотезе
  - обычно такой критерий получается несмещенным
  - В этом случае P-значение равно $PV = 1 - F_T(T)$
  - если F_T непрерывная функция, то P-значение $PV \sim U(0,1)$ имеет равномерное распределение при нулевой гипотезе
    - преобразование Смирнова
- Аналогичное свойство выполнено и для других тестов, построенных с использованием статистик, имеющих при нулевой гипотезе непрерывное распределение.

Распределение P-значения

Утверждение (обобщение преобразования Смирнова).

Пусть G статистика критерия для проверки статистической гипотезы H_0, $\{I_\alpha\}_{\alpha\in[0,1]}$ — семейство вложеных замкнутых множеств $I_\alpha \subseteq I_{\alpha_1}$ при любых $\alpha > \alpha_1$ и

$$P(T \in I_\alpha) = 1 - \alpha \quad \text{при всех} \quad \alpha \in [0,1],$$

P-значение — $PV = \inf\{\alpha : T \notin I_\alpha\}$. Тогда, $PV \sim U(0,1)$.

Доказательство.

Отметим, что для любого $\epsilon > 0$: $\alpha + \epsilon \leq 1$,

$$P(PV \in [\alpha, \alpha+\epsilon)) = P(T \in I_\alpha \backslash I_{\alpha+\epsilon}) = P(T \in I_\alpha) - P(T \in I_{\alpha+\epsilon}) = \epsilon.$$

Следовательно, распределение PV – аюсолютно непрпрерывно и имеет $U(0,1)$ распределение.

Распределение P-значения

Утверждение (обобщение преобразования Смирнова).

Пусть G статистика критерия для проверки статистической гипотезы H_0, $\{I_\alpha\}_{\alpha\in[0,1]}$ — семейство вложеных замкнутых множеств $I_\alpha \subseteq I_{\alpha_1}$ при любых $\alpha > \alpha_1$ и

$$P(T \in I_\alpha) = 1 - \alpha \quad \text{при всех} \quad \alpha \in [0, 1],$$

P-значение — $PV = \inf\{\alpha : T \notin I_\alpha\}$. Тогда, $PV \sim U(0, 1)$.

Доказательство.

Отметим, что для любого $\epsilon > 0$: $\alpha + \epsilon \leq 1$,

$$P(PV \in [\alpha, \alpha + \epsilon)) = P(T \in I_\alpha \backslash I_{\alpha+\epsilon}) = P(T \in I_\alpha) - P(T \in I_{\alpha+\epsilon}) = \epsilon.$$

Следовательно, распределение PV – аюсолютно непрпрерывно и имеет $U(0, 1)$ распределение.

Доверительные множества и критерии

Использование доверительных интервалов для построения статистических критериев

- Пусть
  - $H_0 : \theta = \theta_0$ – простая гипотеза
  - $G(X, \theta)$ – генератор доверительного множества
  - $\hat{\Theta} = \{\theta : G(X, \theta) \in I_\alpha\}$ – доверительное множество уровня доверия $1 - \alpha$
- Дополнительно предположим, что распределение $G(X, \theta)$ отличается от распределения $G(X, \theta_0)$ при любом $\theta \in \Theta_A$.
- Статистический критерий для проверки H_0

$$\phi(X) = \begin{cases} 0, & G(X, \theta_0) \in I_\alpha \\ 1, & G(X, \theta_0) \notin I_\alpha \end{cases} \quad \text{или} \quad \phi(X) = \begin{cases} 0, & \theta_0 \in \hat{\Theta} \\ 1, & \theta_0 \notin \hat{\Theta} \end{cases}$$

  - уровень значимости данного критерия равен α
- При построении критериев для сложной гипотезы $H_0 : \theta \in \Theta_0$
  - требуется распределение статистики критерия при $\theta \in \Theta_0$
  - I_α находятся из условия, что наибольшее значение вероятности при $\theta \in \Theta_0$ не менее $1 - \alpha$

Проверка значимости

Наиболее распространенным методом в медицинских исследованиях является проверка значимости отклонений от основной гипотезы.

- Существенным результатом является отвержение основной гипотезы – выявление значимых отклонений.
- Реальные значения отклонений от основной гипотезы не изучаются.
- Наличие богатой статистической информации позволяет выявлять даже несущественные с практической точки зрения различия.
- Если проверяется несколько статистических гипотез, то необходима поправка.

Задача

Пусть $X_1, \ldots, X_n$ – выборка из нормального распределения $\mathcal{N}(a, \sigma^2)$. Построить критерий значимости отклонений от гипотезы согласия $H_0 : a = 0$ при альтернативе $H_A : a \neq 0$.

Решение. В качестве статистики критерия используем статистику Стьюдента
$$T(X) = \sqrt{n-1}\frac{\overline{X}}{s}$$
По лемме Фишера п. 4 статистика T имеет при H_0 распределение Стьюдента с $n-1$ степенью свободы. Выбираем $I_\alpha = [-x_\alpha, x_\alpha]$: $S_{n-1}(x_\alpha) = 1 - \alpha/2$. Получаем нерандомизованный статистический критерий
$$\phi(X) = \begin{cases} 0, & T(X, \theta_0) \in [-x_\alpha, x_\alpha] \\ 1, & T(X, \theta_0) \notin [-x_\alpha, x_\alpha] \end{cases}$$
Вероятность ошибки I рода равна $\mathbb{E}_{(0,\sigma^2)}\phi(X) = P_{(0,\sigma^2)}(\phi(X) = 1) = \alpha$. Построенный критерий имеет уровень значимости α. ∎

- При односторонней альтернативе $H_A : \theta > 0$ следует использовать ту же статистику T с $I_\alpha = (-\infty, x_\alpha)$, $x_\alpha : \Phi(x_\alpha) = 1 - \alpha$.

Задача

Пусть $X_1, \dots, X_n$ и $Y_1, \dots, Y_k$ – независимые выборки из нормальных распределений $\mathcal{N}(a, \sigma^2)$ и $\mathcal{N}(b, \sigma^2)$. Построить критерий значимости отклонений от гипотезы однородности $H_0 : a = b$ при альтернативе $H_A : a \neq b$.

Решение. В качестве статистики критерия используем статистику Стьюдента для двух выборок

$$T(X) = \sqrt{\frac{nk(n+k-2)}{k+n}} \frac{\overline{X} + \overline{Y}}{\sqrt{ns_x^2 + ms_y^2}}$$

Ранее было показано (см. лекцию «доверительное оценивание»), что при основной гипотезе $T \sim S_{n+k-2}$. Выбираем $I_\alpha = [-x_\alpha, x_\alpha]$: $S_{n-1}(x_\alpha) = 1 - \alpha/2$. Тогда статистический критерий

$$\phi(X) = \begin{cases} 0, & T(X) \in [-x_\alpha, x_\alpha] \\ 1, & T(X) \notin [-x_\alpha, x_\alpha] \end{cases}$$

имеет уровень значимости α. ■

- Предположение о равенстве дисперсий достаточно ограничительно и может требовать проверки.

Задача

Пусть $X_1, \dots, X_n$ и $Y_1, \dots, Y_k$ – независимые выборки из нормальных распределений $\mathcal{N}(a, \sigma_1^2)$ и $\mathcal{N}(b, \sigma_2^2)$. Построить критерий значимости отклонений от гипотезы равенства дисперсий $H_0 : \sigma_1^2 = \sigma_2^2$ при альтернативе $H_A :, \sigma_1^2 \neq \sigma_2^2$.

Решение. Используем статистику критерия

$$V(X) = \frac{ns_x^2}{ks_y^2}$$

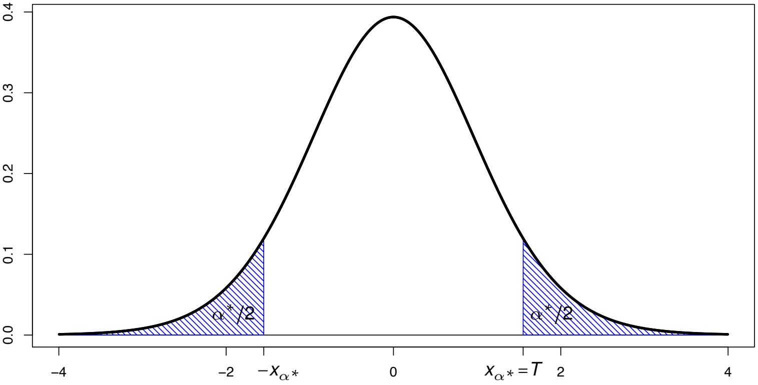
Известно (следствие к лемме Фишера), что при основной гипотезе $S \sim F_{n-1,k-1}$ (распределение Фишера–Снедекора). Выбираем $I_\alpha = [x_{1,\alpha}, x_{2,\alpha}]$: $F_{n-1,k-1}(x_{2,\alpha}) = 1 - \alpha/2$, $F_{n-1,k-1}(x_{1,\alpha}) = \alpha/2$ и получаем статистический критерий

$$\phi(X) = \begin{cases} 0, & V(X) \in [-x_\alpha, x_\alpha] \\ 1, & V(X) \notin [-x_\alpha, x_\alpha] \end{cases}$$

уровеня значимости α. ■

- Предположение о равенстве дисперсий достаточно ограничительно и может требовать проверки.

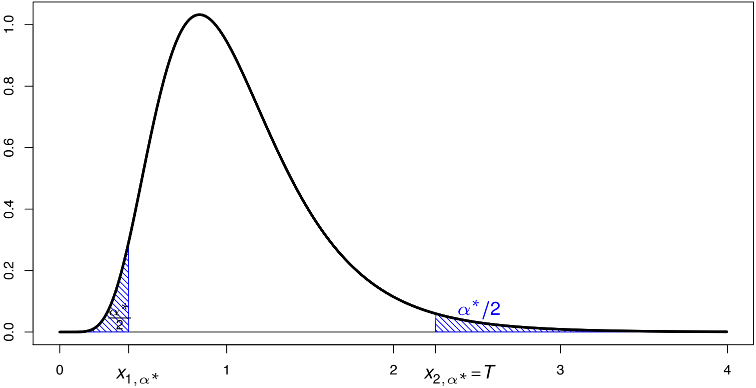
Вычисление P-значений для двухсторонних критериев Стьюдента проверки согласия и однородности производится аналогично



- $PV = \alpha^* = 2(1 - S_m(T)) = 1 - F_{1,m}(T)$
  - $m = n - 1$ для критерия согласия Стьюдента
  - $m = n + k - 2$ для критерия однородности Стьюдента
  - $F_{1,m}$ – функция распределения Фишера–Снедекора с параметрами 1 и m

Вычисление P-значений

Вычисление P-значений критерия проверки равенства дисперсий



- $PV = \alpha^* = 2\,min(1 - F_{n-1,m-1}(T), F_{n-1,m-1}(T))$
  - $m = n - 1$ для критерия согласия Стьюдента
  - $m = n + k - 2$ для критерия однородности Стьюдента
  - $F_{1,m}$ – функция распределения Фишера–Снедекора с параметрами 1 и m