Доверительные интервалы. Гипотезы

Центр биоэлектрических интерфейсов, 12 декабря 2018 г.

Денис Деркач, Влад Белавин

Оглавление

Понятие статистической гипотезы

Понятие статистической гипотезы

Определение

Статистическая гипотеза — определённое предположение о распределении вероятностей, лежащем в основе наблюдаемой выборки данных.

Выделяют 2 типа статистических гипотез:

- **Простая гипотеза** однозначно определяет функцию распределения на рассматриваемом множестве, например $\theta = \theta_0$ — простая гипотеза.
- **Сложная гипотеза** утверждает принадлежность распределения к некоторому множеству распределений на рассматриваемом множестве, например $\theta > \theta_0$ или $\theta < \theta_0$ — сложная гипотеза.

Понятие статистической гипотезы

Определение

Проверка статистической гипотезы — это процесс принятия решения о том, противоречит ли рассматриваемая статистическая гипотеза наблюдаемой выборке данных.

Всегда рассматривается задача проверки гипотезы H_0 против альтернативы H_1 (или же задача (H_0, H_1)).

Статистический критерий и его характеристики

Статистический критерий и его характеристики

Определение

Статистический критерий — строгое математическое правило, по которому принимается или отвергается статистическая гипотеза.

В зависимости от типа статистической гипотезы выделяют односторонние и двусторонние статистические критерии:

- Односторонний критерий

$$H_0 \ : \theta \leq \theta_0 \ \ vs \ \ H_1 \ : \theta > \theta_0.$$

- Двусторонний критерий

$$H_0 \ : \theta = \theta_0 \ \ vs \ \ H_1 \ : \theta \neq \theta_0.$$

Параметрический и непараметрический

Также возможно деление на параметрические и непараметрические критерии:

Определение

Параметрический критерий — критерий , предполагающий, что выборка порождена распределением из заданного параметрического семейства. В частности, существует много критериев, предназначенных для анализа выборок из нормального распределения.

Определение

Непараметрический критерий — критерий, не опирающийся на дополнительные предположения о распределении.

Статистический критерий и его характеристики

- Статистический критерий — это правило, которое для каждой реализации x выборки X должно приводить к одному из двух решений: принять гипотезу H_0 или отклонить ее (принять ее альтернативу H_1).
- В связи с этим каждому критерию соответствует некоторое разбиение выборочного пространства χ на два взаимно дополняющих множества χ_0 и χ_1.
- χ_0 состоит из тех реализаций выборки x, для которых H_0 принимается, а χ_1 из тех, для которых H_0 отвергается (принимается H_1).

Критическая область

Определение

В определениях предыдущего слайда

- χ_0 — область принятия гипотезы H_0,
- χ_1 — область отклонения H_0 (критическая область).

Таким образом любой критерий проверки гипотезы H_0 однозначно задается соответствующей критической областью χ_1.

Отклонение гипотезы

Выбор критической области в конкретной задаче делается на основе общего принципа принятия решения.

Определение

Общий принцип принятия решений состоит в следующем:

- Если в эксперименте наблюдается маловероятное при справедливости гипотезы H_0 событие, то считается, что гипотеза H_0 не согласуется с данными и в этом случае она отклоняется.
- В противном случае считается что данные согласуются с H_0 и H_0 принимается.

Отклонение H_0 или принятие H_1

В принципе, существует разница между "H_0 отклонена" и "H_1 принята". Лучше говорить, что данные отвергают H_0.

Уровень значимости

В соответвии с общим принципом принятия решения критическую область χ_1 выбирают так, чтобы была мала вероятность $P(X \in \chi_1 | H_0)$.

Определение

Говорят, что критерий **имеет уровень значимости** α, если

$$P(x \in \chi_1 | H_0) = \alpha.$$

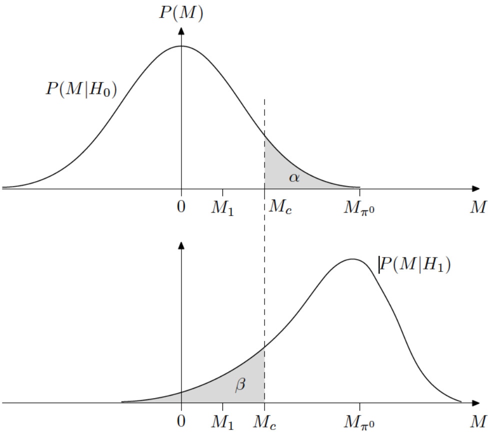
$$P(x \in X / \chi_1 | H_1) = \beta.$$

Статистический критерий и его характеристики

Следуя любому критерию мы можем принять правильное решение, либо совершить одну из двух ошибок — первого или второго рода. Для уровня значимости теста α :

Результат применения критерия	Верная гипотеза	
	H_0	H_1
$x \notin \chi_1$, Принятие H_0	OK $\mathbb{P} = 1 - \alpha$	ошибка 2-го рода $\mathbb{P} = \beta$
$x \in \chi_1$, Принятие H_1	ошибка 1-го рода $\mathbb{P} = \alpha$	OK $\mathbb{P} = 1 - \beta$

α и β



Обычно α фиксируют, β минимизируем с помощью выбора тестовой статистики.

Сравнение тестов

- Пусть критерий имеет критическую область χ_1, а $\mathcal{F} = F_0 \cup F_1$ — множество всех допустимых распределений выборки X.
- При этом F_0 — множество распределений, удовлетворяющих гипотезе H_0, а F_1 — соответственно H_1.

Определение

$W(F) = W(F; \chi_1) = P(X \in \chi_1 | F), F \in \mathcal{F}$ называется **функцией мощности критерия**.

Другими словами мощность критерия показывает вероятность попадания значения выборки X в критическую область χ_1, когда F — ее истинное распределение.

Статистический критерий и его характеристики

Через функцию мощности легко выразить вероятности обоих типов ошибок, свойственных нашему критерию.

Определение

$W(F)$ — **вероятность ошибки первого рода** при $F \in F_0$.
$1 - W(F)$ — **вероятность ошибки второго рода** при $F \in F_1$.

Статистический критерий и его характеристики

Определение

Размер критерия:

$$\alpha = sup_{F \in F_0} W(F).$$

Отсюда легко видеть что если размер критерия не превосходит α, то его уровень равен α.

Статистический критерий и его характеристики

- Логично стремление построить критерий так, чтобы свести к минимуму вероятности ошибок обоих типов.
- Однако при фиксированном объеме выборки сумма вероятностей ошибок обоих типов не может быть сделана сколь угодно малой.
- Поэтому руководствуются рациональным принципом выбора критической области: Из всех критических областей удовлетворяющих заданному уровню значимости выбирается та, для которой вероятность ошибки 2-го рода минимальна.

Статистический критерий. Пример

- Пусть $X_1, X_2, \ldots, X_n \sim N(\mu, \sigma^2)$, где σ — известно.
- Рассмотрим 2 гипотезы $H_0 : \mu \leq 0 \quad H_1 : \mu > 0$.
- Рассмотрим критерий – H_0 отклоняется если $T = \bar{X} > c$.
- Тогда критическая область $\chi_1 = \{(x_1, ...x_n) : T(x_1, ..., x_n) > c\}$.
- Значит

$$W(\mu) = P_\mu(\bar{X} > c) = P_\mu\left(\frac{\sqrt{n}(\bar{X} - \mu)}{\sigma} > \frac{\sqrt{n}(c - \mu)}{\sigma}\right)$$

$$= P\left(Z > \frac{\sqrt{n}(c - \mu)}{\sigma}\right) = 1 - \Phi\left(\frac{\sqrt{n}(c - \mu)}{\sigma}\right),$$

где Z — случайная величина со стандартным нормальным распределением.

Статистический критерий. Пример

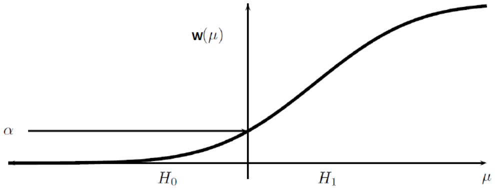


Рис.: $W(\mu)$

Пример

Отсюда легко видеть, что размер критерия равен
$W(0) = 1 - \Phi(\frac{\sqrt{n}c}{\sigma})$

Статистический критерий. Пример

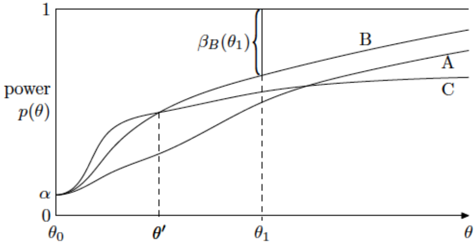
Пример

Чтобы размер критерия равнялся α необходимо, чтобы $c = \frac{\sigma\Phi^{-1}(1-\alpha)}{\sqrt{n}}$.

Определение

Наиболее мощный критерий — критерий, который имеет максимальную мощность относительно гипотезы H_1 (т.е максимальную $1 - W(F)$ при $F \in F_1$) среди всех критериев размера α.

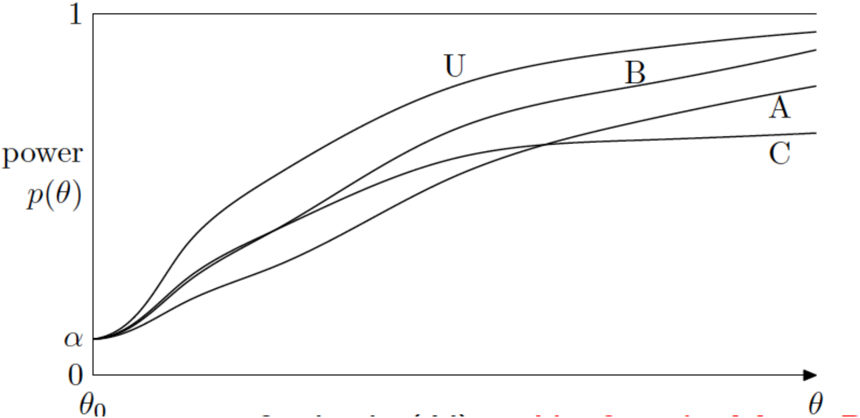
Наиболее мощный критерий



Power functions of tests A, B, and C at significance level α.

Of these tests, B is the best for $\theta > \theta'$. For smaller values of θ, C is better.

Равномерно наиболее мощный критерий



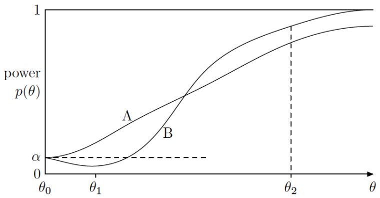
Мощность критериев

- В конкретных задачах наиболее мощный критерий не всегда достижим, поэтому в реальных задачах часто приходится ограничиваться более умеренными требованиями.
- Минимальным таким требованием являются требования несмещенности и состоятельности.

Несмещённость критерия

Определение

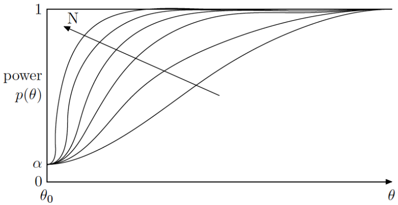
Статистический критерий называется несмещенным если при любом альтернативном распределении данных мы должны попадать в критическую область с большей вероятностью нежели при нулевой гипотезе.



Состоятельные критерии

Определение

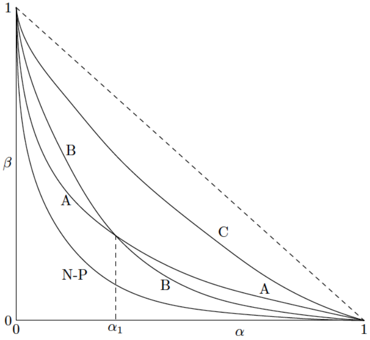
Статистический критерий называется состоятельным если в случае истинности альтернативной гипотезы при большом числе наблюдений мы будем попадать в критическую область с вероятностью близкой к 1 (т.е., отклоняя нулевую гипотезу, мы будем принимать правильное решение).



Критерий Неймана-Пирсона

Мотивация

Логично было бы построить другие характеристики критериев, например, в плоскости $(\alpha; \beta)$.
Все тесты находятся между пунктирной линией и кривой N-P.



Критерий Неймана-Пирсона для случая двух простых гипотез

Лемма (Неймана-Пирсона)

$H_0 : \theta = \theta_0 \quad vs. \quad H_1 : \theta = \theta_1$

Статистика Неймана-Пирсона:

$$T = \frac{\mathfrak{L}(\theta_1)}{\mathfrak{L}(\theta_0)} = \frac{\prod_{i=1}^n f(x_i;\theta_1)}{\prod_{i=1}^n f(x_i;\theta_0)}. \tag{1}$$

Допустим, что H_0 отвергается при $T > k$. Выберем k так, что $\mathbb{P}_{\theta_0}(T > k) = \alpha$.

Тогда, критерий Неймана-Пирсона (на основе статистики (1)) будет иметь наибольшую мощность $W(\theta_1)$ среди всех критериев размера α.

Проблемы статистики Неймана-Пирсона

- статистика должна быть полностью известна для любых x;
- мы можем оценивать только простые гипотезы.

В реальной жизни, обычно мы оцениваем многомерные с известными или неизвестными мешающими параметрами (сложные гиптезы). В общем случае, мы не можем построить равномерно наиболее мощный критерий, но для экспоненциального семейства распределений мы можем построить достаточные статистики.

Приблизительное решение

Вообще, в случае отстутвия возможности строить глобальный критерий, мы можем построить локальный:

- $H_0 : \theta = \theta_0;$
- $H_1 : \theta = \theta_1 = \theta_0 + \Delta.$

В этом случае, мы можем разложить лог-правдоподобие в ряд Тейлора:

$$\ln L(X;\theta_1) = \ln L(X;\theta_0) + \Delta \frac{\partial L(X;\theta)}{\partial \theta}\Big|_{\theta=\theta_0} + \dots.$$

Но по лемме Неймана-Пирсона:

$$\ln L(X;\theta_1) - \ln L(X;\theta_0) \leq c_\alpha$$

Таким образом, мы построили локальный критерий, который зависит только от производной правдоподобия: $\frac{\partial L(X;\theta)}{\partial \theta}\Big|_{\theta=\theta_0} > k_\alpha$

Локальные тесты

Можно вспомнить, что:

$$\mathbb{E}\left[\frac{\partial L(X;\theta)}{\partial\theta}|_{\theta=\theta_0}\right] = 0;$$

$$\mathbb{E}\left[\frac{\partial L(X;\theta)}{\partial\theta}\right] = +NI,$$

где N - количество событий, а I - информационная матрица Фишера.

Вспомним об ассимптотической нормальности правдоподобия, можем переписать локальный критерий так:

$$\frac{\partial L(X;\theta)}{\partial\theta}|_{\theta=\theta_0} \geq \lambda_\alpha\sqrt{NI},$$

Это будет самым мощным локальным критерием.

Байесовский подход

Нам необходимо найти:

$$\mathbb{P}(hyp|data) = \frac{\mathbb{P}(data|hyp)\mathbb{P}(hyp)}{\mathbb{P}(data)}$$

Нормализацию можно найти, интегрируя по всем возможным значениям параметров, что довольно сложно для некоторых типов гипотез. Мы можем изучать:

$$R = \frac{\mathbb{P}(data|H_0)\mathbb{P}(H_0)}{\mathbb{P}(data|H_1)\mathbb{P}(H_1)}$$

Полученное соотношение можно рассматривать как шансы на успех при ставке H_0 против H_1. При этом соотношение всё равно будет зависеть от априорного знания.

Построение статистических критериев

Алгоритм решения задачи проверки гипотез

1. Сформулировать (математически) нулевую гипотезу $\mathbb{H}_0$.
2. Сформулировать (математически) альтернативу.
3. Выбрать уровень значимости α.
4. Получить выборку $\mathbf{X}^\ell$.
5. Подсчитать значение статистики $T = T(\mathbf{X}^\ell)$.
6. Построить критическую область $\mathcal{R}_\alpha$.
7. На основе шагов 5—6 сделать вывод **о согласии $\mathbb{H}_0$ с данными $\mathbf{X}^\ell$.**

Основным вопросом является выбор T, который желательно осуществлять до сбора данных.

Критерий Вальда (Z-тест)

Пусть:

- θ — скалярный параметр;
- $\hat{\theta}$ — его оценка;
- $\hat{se}$ — оценка стандартной ошибки оценки $\hat{\theta}$.

Гипотеза:

$$H_0 : \theta = \theta_0 \;\; vs. \;\; H_1 : \theta \neq \theta_0.$$

Определение (Критерий Вальда размера α)

Если $\hat{\theta}$ — асимптотически нормальная оценка параметра θ, т.е.

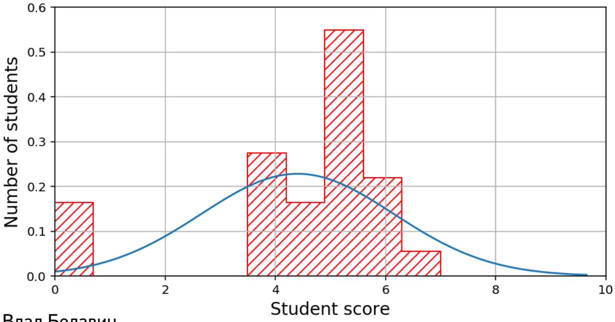
$$W = \frac{\hat{\theta} - \theta_0}{\hat{se}} \to \mathbb{N}(0,1), n \to \infty,$$

то гипотеза H_0 отклоняется, если $|W| > z_{\alpha/2}$.

Критерий Вальда (Z-test)

Рассмотрим выборку:

```
group1 = [
    6, 4, 4, 0, 5, 5, 7, 5, 4.5, 5,
    4, 4, 4.5, 0, 5, 5, 5, 5, 4, 5,
    0, 6, 5, 6, 4.5, 6]
```



Критерий Вальда (Z-test)

› Пусть $X_i \sim \mathcal{N}(\theta, \sigma^2)$. Проверим гипотезу

$$\mathbb{H}_0 : \theta = 5 \quad \text{vs.} \quad \mathbb{H}_1 : \theta \neq 5.$$

Критерий Вальда (Z-test)

- Пусть $X_i \sim \mathcal{N}(\theta, \sigma^2)$. Проверим гипотезу

$$\mathbb{H}_0 : \theta = 5 \quad \text{vs.} \quad \mathbb{H}_1 : \theta \neq 5.$$

- Как получить $\widehat{\theta}$ — оценку среднего?

Критерий Вальда (Z-test)

- Пусть $X_i \sim \mathcal{N}(\theta, \sigma^2)$. Проверим гипотезу
$$\mathbb{H}_0 : \theta = 5 \quad \text{vs.} \quad \mathbb{H}_1 : \theta \neq 5.$$
- Как получить $\widehat{\theta}$ — оценку среднего?
  $\widehat{\theta} = 4.4$
- Как получить $\widehat{se}$ — оценку стандартной ошибки оценки $\widehat{\theta}$?

Критерий Вальда (Z-test)

- Пусть $X_i \sim \mathcal{N}(\theta, \sigma^2)$. Проверим гипотезу
$$\mathbb{H}_0 : \theta = 5 \quad \text{vs.} \quad \mathbb{H}_1 : \theta \neq 5.$$
- Как получить $\widehat{\theta}$ — оценку среднего?
$\widehat{\theta} = 4.4$
- Как получить $\widehat{se}$ — оценку стандартной ошибки оценки $\widehat{\theta}$?
$\widehat{se} = 1.748/\sqrt{26} = 0.342$
- Как получить $\widehat{se}$, если — оценка стандартной ошибки оценки $\widehat{\theta}$.

Критерий Вальда (Z-test)

- Пусть $X_i \sim \mathcal{N}(\theta, \sigma^2)$. Проверим гипотезу
$$\mathbb{H}_0 : \theta = 5 \quad \text{vs.} \quad \mathbb{H}_1 : \theta \neq 5.$$
- Как получить $\widehat{\theta}$ — оценку среднего?
$\widehat{\theta} = 4.4$
- Как получить $\widehat{se}$ — оценку стандартной ошибки оценки $\widehat{\theta}$?
$\widehat{se} = 1.748/\sqrt{26} = 0.342$
- Как получить $\widehat{se}$, если — оценка стандартной ошибки оценки $\widehat{\theta}$. Например, бутстреп!
- При $\alpha = 0.05$, поскольку $z_{\alpha/2} = 1.96$, имеем
$$W = \frac{\widehat{\theta} - \theta}{\widehat{se}} = \frac{4.4 - 5}{0.342} = -1.738$$

Теорема

Асимптотически размер критерия Вальда равен α, то есть

$$W(F) = \mathbb{P}(|W| > z_{\alpha/2}|f) \to \alpha, \;\; n \to \infty, \;\; f \in F_0.$$

Доказательство.

При условии, что $\theta = \theta_0$, в силу асимптотической нормальности оценки выполнено $\frac{\hat{\theta}-\theta_0}{\hat{se}} \rightsquigarrow \mathbb{N}(0,1)$. Следовательно, вероятность отклонить основную гипотезу, когда она на самом деле верна, равняется:

$$\mathbb{P}(|W| > z_{\alpha/2}|f) = \mathbb{P}(\frac{|\hat{\theta} - \theta_0|}{\hat{se}} > z_{\alpha/2}|f) \to$$
$$\to \mathbb{P}(|Z| > z_{\alpha/2}) = \alpha.$$

□

Пример (Сравнение средних значений)

- Пусть $X_1, ..., X_m$ и $Y_1, ..., Y_n$ - две независимые выборки из генеральных совокупностей.
- Средние значения которых равны μ_1 и μ_2 соответсвенно.
- s_1^2 и s_2^2 — выборочные дисперсии.
- Положим $\delta = \mu_1 - \mu_2$.

$$H_0 : \delta = 0 \;\; vs. \;\; H_1 : \delta \neq 0; \quad \hat{\delta} = \bar{X} - \bar{Y}; \quad \hat{se} = \sqrt{\frac{s_1^2}{m} + \frac{s_2^2}{n}}.$$

Гипотеза H_0 отвергается, если $|W| > z_{\alpha/2}$, где

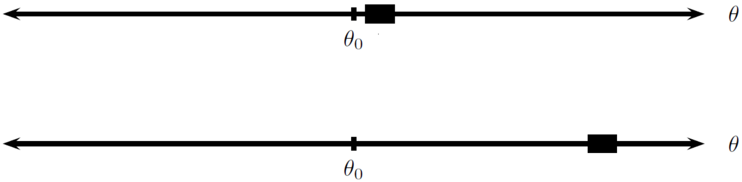
$$W = \frac{\hat{\delta} - 0}{\hat{se}} = \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{s_1^2}{m} + \frac{s_2^2}{n}}}.$$

Теорема

Критерий Вальда размера α отклоняет гипотезу $H_0 : \theta = \theta_0$ в пользу $H_1 : \theta \neq \theta_0$ если и только если $\theta_0 \notin C$, где

$$C = (\hat{\theta} - \hat{se} z_{\alpha/2}, \hat{\theta} + \hat{se} z_{\alpha/2})$$

Таким образом, тестирование гипотезы эквивалентно проверке, попало ли значение Θ_0 в доверительный интервал.

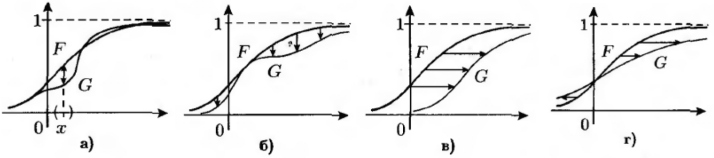


Мотивация непараметрических методов

Часто, бывает непонятно какой параметр стоит оценивать и вообще надо ли оценивать параметры. Например, для сравнения двух выборок можно использовать непараметрические тесты.

U-тест Манна-Уитни

- Имеем две выборки $\mathbf{X}^n \sim F(x)$ и $\mathbf{Y}^m \sim G(x)$
- Гипотеза однородности $\mathbb{H}_0 : F(x) = G(x), x \in \mathbb{R}$



- Бывает важно уловить отклонения от $\mathbb{H}_0$ только определенного типа (наличие прироста Y_j по сравнению с X_i)
- Сужение типов альтернатив $\implies$ более эффективные критерии
- Проверяем гипотезу однородности против альтернативы доминирования $\mathbb{H}_1$ (варианты б) и в))

U-тест Манна-Уитни

- Построим вариационный ряд из объединенной выборки $(X_1, \ldots, X_n, Y_1, \ldots, X_m)$
  - Верна $\mathbb{H}_0 \implies$ значения Y_j рассеяны по всему ряду
  - Иначе средний ранг значений Y_j относительно большой
- Обозначим S_j ранг порядковой статистики $Y_{(j)}$ в этом ряду
- Положим $V = S_1 + \cdots + S_m$
- Критическая область: $V \geqslant c$, где $c = \text{const}$
- **Большие выборки:** (Mann-Whitney-Wilcoxon, MWW)

$$U = \sum_{i=1}^{n} \sum_{j=1}^{m} I_{X_i < Y_j} \to \mathcal{N}\left(\frac{nm}{2}, \frac{nm(n+m+1)}{12}\right)$$

Алгоритмические проверки

Для таблицы ошибок первого и второго типа, мы можем получить характерные количества событий.

Результат применения критерия	Верная гипотеза	
	H_0	H_1
$x \notin \chi_1$, Принятие H_0	True Positive	False Positive
$x \in \chi_1$, Принятие H_1	False Negative	True Negative

Необходимо построить такие характеристики, которые хорошо не сильно зависели бы от размеров выборки.

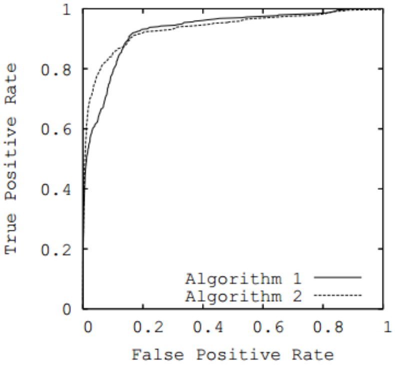
Receiver operating characteristic

Введём две величины:

$$TPR = \frac{TP}{TP + FN};$$
$$FPR = \frac{FP}{FP + TN}.$$

И построим график, зависимости TPR от FPR для разных уровней принятия решений.

Будем характеризовать качество характеристики площадью под кривой ROC (Area Under Curve), ROC AUC. Чем больше AUC, тем лучше, при этом $0.5 < AUC < 1$



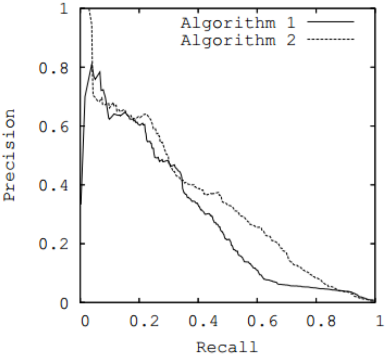
Precision-Recall Characteristic

Введём ещё две величины:

$$Precision = \frac{TP}{TP + FP};$$

$$Recall = \frac{TP}{TP + FN}.$$

Будем характеризовать качество характеристики площадью под кривой PR (Area Under Curve), PR AUC.



ROC vs PR

- максимизация AUC эквивалентна в обоих случаях;
- в случае сильно несбалансированных выборок PR предпочтительнее.

P-значение

Определение

Пусть для каждого $\alpha \in (0,1)$ имеется критерий размера α для некоторой статистики (функции от выборки) $T(X^n)$ с критической областью R_α. Тогда

$$p - value = inf\{\alpha : T(X^n) \in R_\alpha\}.$$

- Таким образом, $p - value$ - это наименьший уровень значимости, на котором еще можно отклонить H_0.
- Чем меньше $p - value$ - тем вероятнее, что H_0 надо отклонить.

Типичные значения для $p - value$:

- $< 0.01 \to H_0$ - заведомо не верна
- $0.01 - 0.05 \to H_0$ - не верна
- $0.05 - 0.10 \to H_0$ - скорее не верна
- $> 0.1 \to$ ничего определенного о гипотезе H_0 сказать нельзя.

Большое p-value не является подтверждением гипотезы H_0.
Большое $p - value$ появляется, если:

- H_0 - верна.
- H_0 - неверна, но мощность критерия недостаточна.

Теорема

Пусть критерий размера α, построенный для статистики $T(X^n)$, имеет вид: H_0 отвергается, если $T(X^n) > c_\alpha$

Гипотезе H_0 соответствует семейство распределений F_0.

Тогда

$$p - value = \sup_{f \in F_0} \mathbb{P}(T(X^n) \geq T(x^n)|f),$$

где x^n - реализация выборки X^n. Если $F_0 = f$, то

$$p - value = \mathbb{P}(T(X^n) \geq T(x^n)|f).$$

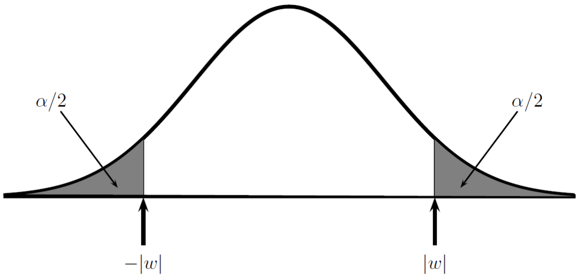
То есть $p - value$ - это вероятность (при выполнении гипотезы H_0) того, что статистика $T(X^n)$ примет значение больше либо равное тому, которое реализовалось в опыте (реализация x^n)

Теорема

Пусть $w = \frac{(\hat{\theta}-\theta_0)}{\hat{se}}$ - наблюдаемое значение статистики Вальда W. Тогда:

$$p - value = \mathbb{P}(|W| > |w||f) \simeq \mathbb{P}(|Z| > |w|) = 2\Phi(-|w|),$$

где $Z \sim N(0,1), \quad f \in F_0$.



Пример (Равенство значений холестерина в крови)

Здесь как и в примере про сравнение средних значений:

$$W = \frac{\hat{\delta} - 0}{\hat{se}} = \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{s_1^2}{m} + \frac{s_2^2}{n}}} = \frac{216.2 - 195.3}{\sqrt{5^2 + 2.4^2}} = 3.78.$$

Пусть $Z \sim N(0, 1)$, тогда

$$p - value = \mathbb{P}(|Z| > 3.78) = 2 \cdot \mathbb{P}(Z < -3.78) = 0.0002$$

.

Распределение хи-квадрат (χ^2)

Определение (распределение хи-квадрат(χ^2))

Пусть $Z_1, ..., Z_k$ - независимые стандартно нормально распределенные случайные величины. $V = \sum_{i=1}^{k} Z_i^2$, тогда $V \sim \chi_k^2$ - хи-квадрат с k степенями свободы

$$F(V) = \frac{v^{\frac{k}{2}-1} e^{-\frac{V}{2}}}{2^{\frac{k}{2}} \Gamma(\frac{k}{2})}, \quad \mathbb{E}(V) = k, \quad \mathbb{V}(V) = 2k$$

$\chi_{k,\alpha}^2 = F^{-1}(1-\alpha)$ - верхняя квантиль, F - функция распределения, т.е.

$$\mathbb{P}(\chi_k^2 > \chi_{k,\alpha}^2) = \alpha.$$

Критерий на основе отношения правдоподобия

Определение

Рассмотрим две конкурирующие гипотезы

$$H_0 : f \in F_0 \quad vs. \quad H_1 : f \in F_1$$

Пусть $\hat{f}$ - ОМП и $\hat{f}_0$ - ОМП при $f \in F_0$ Статистика отношения правдоподобия имеет вид:

$$\lambda = 2 \log \frac{\sup_{f \in \mathcal{F}} \mathfrak{L}(f)}{\sup_{f \in F_0} \mathfrak{L}(f)} = 2 \log \frac{\mathfrak{L}(\hat{f})}{\mathfrak{L}(\hat{f}_0)}.$$

Теорема

Допустим, что $f = (\theta_1, \dots, \theta_q, \theta_{q+1}, ..., \theta_r)$. Пусть $F_0 = \{f : (\theta_{q+1}, \dots, \theta_r) = (\theta_{0,q+1}, ..., \theta_{0,r})\}$.
Пусть λ - критерий на основе отношения правдоподобия. При гипотезе $H_0 : f \in F_0$

$$\lambda(x^n) \rightsquigarrow \chi^2_{q,\alpha},$$

где q - размерность F за вычетом размерности F_0. $p-value$ для критерия равно $\mathbb{P}(\chi^2_q > \lambda)$.

Пример

Пусть $f = (\theta_1, \dots, \theta_5)$, необходимо проверить, что $\theta_4 = \theta_5 = 0$. Тогда у предельного распределения имеется 3 степени свободы.

Пример (Горох Менделя $1/3$)

Пример. Горох Менделя. Два типа: круглые желтые зерна и сморщенные зеленые зерна. Имеется 4 типа потомков: круглые желтые, сморщенные желтые, круглые зеленые и сморщенные зеленые. Количество потомков каждого типа образуют мультиномиальное распределение с вероятностью $p = (p_1, p_2, p_3, p_4)$.

Из теории следует, что

$$p = (\frac{9}{16}, \frac{3}{16}, \frac{3}{16}, \frac{1}{16}).$$

В опыте получено, что $n = 556, X = (315, 101, 108, 32)$.

Пример (Горох Менделя $2/3$)

Статистика отношений правдоподобия для

$$H_0 : p = p_0 \quad vs. \quad H_1 : p \neq p_0$$

принимает вид:

$$\lambda = 2\log\frac{\mathfrak{L}(\hat{p})}{\mathfrak{L}(\hat{p_0})} = 2\sum_{j=1}^{4} X_j \log\frac{\mathfrak{L}(\hat{p})}{\mathfrak{L}(\hat{p_0})} = 2\cdot(315\log(\frac{315/556}{9/16})+$$

$$+101\log(\frac{101/556}{3/16}) + 108\log(\frac{108/556}{3/16}) + 32\log(\frac{32/556}{1/16})) = 0.48$$

Пример (Горох Менделя $3/3$)

При гипотезе H_1 4 параметра. Так как сумма параметров должна равняться 1, то размерность пространства параметров равна 3. При гипотезе H_0 свободных параметров нет, значит количество степеней свободы равно 3 и χ_3^3 является предельным распределением.

$$p - value = \mathbb{P}(\chi_3^2 > 0.48) = 0.92$$

Замечание

Как правило, и критерий χ^2, и критерий отношения правдоподобий дают примерно одинаковые результаты при условии, что размер выборки достаточно большой.

Критерий согласия

Необходимо проверить согласие параметрической модели распределения и полученных в экспериментальных данных Пусть $\mathfrak{M} = \{f(x; f) : f \in F\}$ - параметрическая модель, $f = (\theta_1, ..., \theta_s)$.

Допустим, что область значений выборки $(-\infty, \infty)$. Разделим $(-\infty, \infty)$ на k непересекающихся интервалов $I_1, ..., I_k$. Для $j = 1, ..., k$ обозначим через

$$p_j(f) = \int_{I_j} f(x, F) dx$$

вероятность попадания наблюдения в интервал I_j при условии, что модель верна. Пусть N_j - количество элементов выборки, попавших в I_j

Можно показать, что функция правдоподобия для параметра θ на основе наблюдений $N_1, \dots, N_k$ (в предположении, что эти наблюдения имеют мультиноминальное распределение) имеет вид:

$$Q(f) = \prod_{j=1}^{k} p_i(f)^{N_j}$$

Максимизируя $Q(f)$ получаем оценку $\tilde{f} = (\tilde{\theta}_1, \dots, \tilde{\theta}_s)$.

Определение (Статистика критерия согласия)

$$Q = \sum_{j=1}^{k} \frac{(N_j - np_j(\tilde{f}))^2}{np_j(\tilde{f})}$$

Теорема (Сходимость к χ^2)

Пусть H_0: распределение выборки $\{X_1, \dots, X_n\}$ принадлежит параметрическому классу $\mathfrak{M} = \{f(x; f) : f \in F\}$.

Тогда статистика критерия согласия $Q \; : \; Q \xrightarrow{P} \chi^2_{k-1-s}$

И $p - value = \mathbb{P}(\chi^2_{k-1-s} > q)$, где q- наблюдаемое в эксперименте значение статистики Q.

Критерий согласия не позволяет проверить правильность модели:

- Если H_0 не была отклонена, то это не значит, что модель верна.
- Модель могла быть не отклонена из-за низкой мощности критерия.
- Если H_0 была отклонена, то модель заведомо плохая.

Лучше использовать непараметрические модели.

t-критерий

Определение

Случайная величина имеет распределение Стьюдента (t-распределение) с k степенями свободы, если:

$$f(t) = \frac{\Gamma(\frac{k+1}{2})}{\sqrt{k\pi}\Gamma(\frac{k}{2})(1+\frac{t^2}{k})^{\frac{k+1}{2}}}$$

При $k \to \infty$ t-распределение стремится к стандартному нормальному распределению. При $k = 1$ t-распределение совпадает с распределением Коши.

t-критерий используют, когда распределение данных близко к нормальному, а размер выборки невелик.

Теорема (t-критерий)

Пусть $X_1, ..., X_n \sim N(\mu, \sigma^2)$, где параметры (μ, σ^2) неизвестны.

$$H_0 : \mu = \mu_0 \;\; vs. \;\; H_0 : \mu \neq \mu_0$$

Обозначим через S_n^2 выборочную дисперсию. Тогда статистика критерия:

$$T = \frac{\sqrt{n}(\bar{X}_n - \mu_0)}{S_n}$$

Основная гипотеза отвергается, если $|T| > t_{n-1,\alpha/2}$, где $t_{n-1,\alpha/2}$– квантиль распределения Стьюдента с **n-1** степенями свободы.

При больших n выполняется $T \sim N(0, 1)$, то есть при больших n t-критерий эквивалентен критерию Вальда.

Множественные тесты

Множественные тесты

Рассмотрим m различных случаев проверки гипотез:

$$H_{0i} \quad vs. \quad H_{1i}, \quad i = 1, \ldots, m.$$

Обозначим через $P_1, ..., P_m$ величины $p - values$ этих проверок.

Определение (Метод Бонферрони)

Для заданных $p - values$ $P_1, ..., P_m$ основная гипотеза H_{0i} отклоняется, если

$$P_i < \frac{\alpha}{m}.$$

Теорема

При применении метода Бонферрони вероятность неправильно отклонить любую из основных гипотез меньше либо равна α.

Доказательство.

Обозначим через

- R событие, состоящее в том, что по крайней мере одна из основных гипотез была ложно отклонена;
- R_i — что ложно отклонена была i-ая основная гипотеза.

Тогда

$$\mathrm{P}(R) = \mathrm{P}\Big(\bigcup_{i=1}^{m} R_i\Big) \leq \sum_{i=1}^{m} \mathrm{P}(R_i) = \sum_{i=1}^{m} \frac{\alpha}{m} = \alpha.$$

□

False Discovery Rate, FDR

	$\mathbb{H}_0$ не отклонена	$\mathbb{H}_0$ отклонена	$\sum$
$\mathbb{H}_0$ верна	U	V	m_0
$\mathbb{H}_0$ неверна	T	S	m_1
$\sum$	$m - R$	R	m

Определение

Исходя из таблицы, доля (FDP) и частота (FDR) появления ложных отклонений (среди отклонений вообще):

$$FDP = \begin{cases} \frac{V}{R}, & R > 0, \\ 0, & R = 0, \end{cases}$$

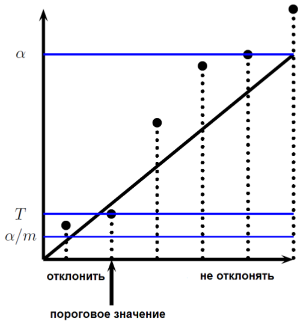
$$FDR = \mathbb{E}(FDP).$$

Метод $Benjamini - Hochberg$

1. Пусть $P_{(1)} < \cdots < P_{(m)}$ -величины p-value, отсортированные по возрастанию.
2. Пусть $C_m = 1$ в случае, если $P_{(1)}, ..., P_{(m)}$ независимы, в противном случае $C_m = \sum_{i=1}^{m} \frac{1}{i}$.
3. Определим: $I_i = \frac{i\alpha}{C_m m}$, $R = \max\{i : P_{(i)} < I_i\}$.
4. $T = P_{(R)}$ - пороговое значение метода.
5. Отклоняются такие H_{0i}, для которых $P_i \leq T$.

Теорема

$$FDR = \mathbb{E}(FDP) \leq \frac{m_0}{m}\alpha \leq \alpha.$$



Критерий перестановок

Определение (Постановка задачи)

Критерий перестановок применяется для проверки того, отличаются ли распределения.
Пусть $X_1, ...X_m \sim F_X$ и $Y_1, ..., Y_n \sim F_Y$ - две независимые выборки. Требуется решить:

$$\mathbb{H}_0 : F_X = F_Y \quad vs. \quad \mathbb{H}_1 : F_X \neq F_Y$$

Критерий перестановок - "точный" в том смысле, что он не использует предположения об асимптотической сходимости к нормальному распределению.

Критерий перестановок:

1. Обозначим через $T(x_1, ..., x_m, y_1, ..., y_n)$ некоторую тестовую статистику, например, $T(X_1, ..., X_m, Y_1, ..., Y_n) = |\overline{X}_m - \overline{Y}_n|$.
2. Положим $N = m + n$ и рассмотрим все $N!$ перестановок объединенной выборки $X_1, ..., X_m, Y_1, ..., Y_n$.
3. Для каждой из перестановок подсчитаем значение статистики T.
4. Обозначим эти значения $T_1, ..., T_{N!}$.

Теорема (Критерий перестановок)

Если $\mathbb{H}_0$ верна, то при фиксированных упорядоченных значениях $\{X_1, ..., X_m, Y_1, ..., Y_n\}$ значение статистики T распределены равномерно на множестве $T_1, ..., T_{N!}$.

Теорема

Обозначим как перестановочное распределение статистики T такое, согласно которому:

$$P_0(T = T_i) = \frac{1}{N!}, \quad i = 1, ..., N!$$

Пусть t_{obs} - значение статистики, которое было получено в опыте. Тогда:

$$p - value = \mathrm{P}(T > t_{obs}|f) = \frac{1}{N!}\sum_{j=1}^{N!}\mathbb{I}(T_j > t_{obs}), \quad f \in \mathcal{F}_0$$

Пример

Допустим, что $(X_1, X_2, Y_1) = (1, 9, 3)$. Пусть $T(X_1, X_2, Y_1) = |\overline{X} - \overline{Y}| = 2$, тогда

Перестановка	Значение T	Вероятность
(1,9,3)	2	1/6
(9,1,3)	2	1/6
(1,3,9)	7	1/6
(3,1,9)	7	1/6
(3,9,1)	5	1/6
(9,3,1)	5	1/6

$$p - value = \mathrm{P}(T > 2) = 4/6$$