

1 Критерии согласия или как перейти к параметрической модели

1.1 О критериях в целом

Напомним некоторую базовую информацию, которые вы проходили в прошлом семестре.

1.1.1 Ошибки первого и второго рода

Пусть у нас есть выборка $X_1, \dots, X_n \sim F$.

Определение 1. Критерием для проверки гипотезы $H_0 : F \in \mathcal{A}_0$ против альтернативы $H_1 : F \in \mathcal{A}_1$ мы называем правило, которое по выборке выдает число 0 (принять H_0 гипотезу) или число 1 (принять альтернативу H_1 , хотя более правильно говорить "отвергнуть гипотезу H_0 ").

Здесь $\mathcal{A}_0, \mathcal{A}_1$ — некоторые множества распределений.

Любой критерий задается критическим множеством D — таким множеством выборок, при попадании в которое критерий выдает 1.

Для критерия существует два вида ошибок: отвергнуть H_0 , когда она верна или принять H_0 , когда на самом деле верна H_1 .

Определение 2. Ошибка первого типа называется *ошибкой первого рода*, второго — *ошибкой второго рода*.

Чтобы не путаться в них, стоит держать в голове, что гипотеза и альтернатива асимметричны. Мы по умолчанию считаем гипотезу верной и пытаемся отвергнуть ее. Соответственно, ошибка отвержения верной гипотезы — ошибка первого рода — для нас особенно опасна. Ошибка второго рода — неприятна и мы пытаемся сделать ее поменьше, но насколько получится.

Например, врачи по умолчанию считают пациента больным от опасного вируса. Ошибка первого рода — упустить больного — недопустима и ее стараются сделать очень маленькой. Ошибка второго рода — задержать здорового — неприятна, но вторична.

Стоит помнить, что критерий принципиально асимметричен, он устроен так, чтобы сделать маленькой вероятность ошибки I рода, а за вероятностью ошибки II мы вынужденно следим куда слабее.

1.1.2 Фактический уровень значимости

В рамках курса математической статистики вы рассматривали критерии с заданным уровнем значимости α , то есть те, у которых вероятности всех возможных ошибок 1 рода равны (меньше или равны) α , т.е.

$$\mathbf{P}_F((X_1, \dots, X_n) \in D) = \alpha, \quad \forall F \in \mathcal{A}_0.$$

Более удобным являются следующий подход. Рассмотрим семейство критических множеств D_c с некоторым параметром, $D_{c_2} \subseteq D_{c_1}$, $c_1 \leq c_2$. Наиболее распространенным вариантом таких множеств являются $D_c = \{T(X_1, \dots, X_n) \geq c\}$, где T — некоторая статистика.

Определение 3. Статистика T называется статистикой критерия.

Найдем максимальное такое c , что D_c содержит нашу реализацию x_1, \dots, x_n .

Определение 4. Величину $\sup_{F \in \mathcal{A}_0} \mathbf{P}_F((X_1, \dots, X_n) \in D_c)$ называют *фактическим уровнем значимости* (p-value).

Это минимальный уровень значимости, при котором гипотеза H_0 при нашей выборке отвергается. Маленький фактический уровень значимости свидетельствует о том, что гипотеза крайне маловероятна и отвергается практически наверняка, большой — что данный критерий не отвергает нашу гипотезу.

Вопрос 1. Рассмотрим критерий $\bar{X} > c$ для проверки гипотезы $H_0 : \mu = 0$ с альтернативой $H_1 : \mu > 0$ для $X_i \sim \mathcal{N}(\mu, 1)$. Статистика \bar{x} приняла значение 3. Чему равен фактический уровень значимости?

Рассматривая критерии вида $\{T(X_1, \dots, X_n) \geq c\}$, где T — статистика с непрерывной функцией распределения, и считая их фактический уровень значимости, мы можем заметить, что при выполнении гипотезы он распределен равномерно на $[0, 1]$. Это простое следствие того, что $F(X) \sim R[0, 1]$ при $X \sim F$.

Таким образом, если у нас в доступе есть большое число выборок, мы можем по каждой найти фактический уровень значимости и посмотреть на распределение этих уровней. При выполнении гипотезы оно должно быть близко к равномерному. При альтернативах, по возможности, величина должна быть маленькой.

Фактический уровень значимости позволяет перевести статистику критерия в удобную для восприятия шкалу. Сама статистика критерия имеет при гипотезе какое-то распределение, а вот статистика критерия — равномерна. Равномерную величину легче представить и понять много это или мало.

1.2 Критерии согласия

1.2.1 Общая постановка

В рамках сегодняшнего занятия мы будем рассматривать критерии согласия и принадлежности параметрическому семейству.

В качестве H_0 будем рассматривать $F \in F_\theta$, т.е. принадлежность F какому-то параметрическому семейству. Альтернативой H_1 мы будем считать все остальные распределения. В англоязычной среде такие тесты называют goodness of fit.

Для построения критерия уровня α достаточно найти некоторое свойство, которые бы выполнялось для всех распределений нашего класса достаточно вероятно (с вероятностью не менее $1 - \alpha$).

При этом сколько-то удовлетворительно мажорировать вероятность ошибки второго рода не удастся, поскольку вне нашего параметрического семейства есть сколь угодно похожие на наши распределения. Но по-крайней мере, можно искать критерий, от которого мы ожидаем, что при альтернативе он чаще попадает в критическое множество.

Пример 1. Теорема Пирсона утверждает, что если вектор $\vec{X} = (X_1, \dots, X_k)$ принимает значения $(1, 0, \dots, 0), \dots, (0, 0, \dots, 1)$ с вероятностями p_1, \dots, p_k , то при n -кратном розыгрыше н.о.р. таких векторов, частоты $(\nu_1, \dots, \nu_k) = \vec{X}_1 + \dots + \vec{X}_n$ появления различных исходов удовлетворяют соотношению

$$T = \sum_{i=1}^k \frac{(\nu_i - np_i)^2}{np_i} \xrightarrow{d} Y \sim \chi_{k-1}^2,$$

где χ_{k-1}^2 — распределение хи-квадрат с $k - 1$ степенью свободы.

Пусть мы хотим проверить гипотезу о том, что вероятности — заданные p_1^0, \dots, p_k^0 и используем

$$T = \sum_{i=1}^k \frac{(\nu_i - np_i^0)^2}{np_i^0} \xrightarrow{d} Y \sim \chi_{k-1}^2.$$

Пусть y_p — квантиль уровня p (то есть $F^{-1}(p)$, где F — функция распределения) для χ_{k-1}^2 . Тогда любое из критических множеств $\{T < y_\alpha\}$, $\{T > y_{1-\alpha}\}$, $\{T < y_{\alpha/2}\} \cup \{T > y_{1-\alpha/2}\}$ задает критерий с уровнем значимости $1 - \alpha$. При этом при альтернативе мы ожидаем, что статистика T будет принимать большие значения, поскольку ν_i/n будут стремиться не к p_i^0 , а к каким-то другим вероятностям. Следовательно, критерий с первым критическим множеством будет крайне неудачным. Он действительно редко (с вероятностью α) будет попадать в критическое множество при верной гипотезе, но еще реже будет попадать туда при альтернативе. Из этих соображений понятно, что наиболее разумный критерий имеет второй вид. Это и есть критерий хи-квадрат, который мы изучали в рамках математической статистики.

Итак, чтобы построить критерий нам нужно провести следующую процедуру:

1. Найти статистику критерия T , которая при гипотезе должна принимать меньшие значения чем при альтернативе;

2. Доказать теорему из теории вероятностей о том, что T имеет заданную функцию распределения (или предельную функцию распределения) $F_{T,0}$ при верной гипотезе.
3. Найти мое значение t статистики и подсчитать p-value $1 - F_{T,0}(t)$.

В некоторых случаях при разных точках из гипотезы T может иметь различные распределения, в таком случае мы ищем $\max(1 - F_{T,0}(t))$ среди всех моих ф.р. из гипотезы.

1.3 О простой гипотезе и сложной альтернативе

Рассмотрим сперва проверку гипотезы для случая, когда гипотеза H_0 простая — $F = F_0$. Один такой критерий мы уже разобрали

1. Критерий хи-квадрат.

Разобьем область значений нашей величины на непересекающиеся диапазоны $\Delta_1, \dots, \Delta_k$, подсчитаем $p_i = \mathbf{P}(X \in \Delta_i)$, $X \sim F_0$. Тогда если H_0 верна, то ν_i — количества X_1, \dots, X_n , попавших в Δ_i , удовлетворяют теореме Пирсона

$$T = \sum_{i=1}^k \frac{(\nu_i - np_i)^2}{np_i} \xrightarrow{d} Y \sim \chi_{k-1}^2.$$

Критерий χ^2 предлагает взять критическое множество $D = \{T > y_{1-\alpha}\}$.

На практике k обычно берут около $\log_2 n$.

Этот критерий а) асимптотический и в целом не гарантирует никаких особенных свойств при конечных n б) подменяет гипотезу на дискретизированную, из-за чего может принять мое распределение за другое, но достаточно разумен и служит основой множества подходов.

Для дискретных данных с конечным числом значений критерий состоятелен.

В Python хи-квадрат задается `chisquare` из `scipy.stats`. Этому критерию нужно указать вектор настоящих частот и предполагаемых частот (не вероятностей p_i , а именно частот np_i).

2. Критерий Колмогорова.

Этот критерий функционирует только для непрерывных F_0 . В силу теоремы Колмогорова в этом случае

$$\sqrt{n}D_n = \sqrt{n} \sup_x |\hat{F}_n - F_0| \xrightarrow{d} K \sim K(x),$$

где $K(x)$ — распределение Колмогорова, \hat{F}_n — ЭФР. Критерий Колмогорова предлагает использовать критическое множество $\{\sqrt{n}D_n > k_{1-\alpha}\}$, где $k_{1-\alpha}$ — квантиль распределения Колмогорова.

Этот критерий также асимптотический, хотя для непрерывных F можно рассматривать и точный критерий при фиксированном n . Здесь ключевым является то, что распределение D_n при верной гипотезе не зависит от F_0 и потому может быть найдено явно.

Вопрос 2. Доказать состоятельность критерия Колмогорова.

Критерии омега-квадрат. Подобно критерию Колмогорова они работают лишь в непрерывном случае, опираются на $|\hat{F}_n(x) - F(x)|$, но пользуются соотношениями

$$\omega_i^2 = n \int_{\mathbb{R}} (\hat{F}_n(x) - F(x))^2 g_i(F(x)) dF(x) \xrightarrow{d} W_i \sim W_i(x), i = 1, 2$$

где $g_1(x) = 1$, $g_2(x) = 1/(x(1-x))$, а $W_1(x)$, $W_2(x)$ — некоторые распределения. Соответственно, первый критерий (он называется Крамера-Мизеса) предлагает критическое множество $\{\omega_1^2 > w_{1-\alpha,1}\}$, второй (Андерсона-Дарлинга) — $\{\omega_2^2 > w_{1-\alpha,2}\}$, где $w_{1-\alpha,i}$ — квантили W_i , $i = 1, 2$.

Отметим, что подсчет статистик по указанным формулам неудобен (требуется численно брать супремум или интеграл), однако дискретность ЭФР позволяет упростить каждую из них. Приведу как пример

вид статистики Андерсона-Дарлинга:

$$\omega_1^2 = -n - \sum_{i=1}^n \left(\frac{2i-1}{n} \ln F(X_{(i)}) + \left(2 - \frac{2i-1}{n} \right) \ln(1 - F(X_{(i+1)})) \right).$$

Обратите внимание на то, что здесь берется $F(X_{(i)})$ – то есть функция применяется к вариационному ряду, а значит выборку сперва нужно упорядочить.

Зачем нужны эти критерии и чем они отличаются от критерия Колмогорова?

Идейно разница между ними такова:

- Критерий Колмогорова улавливает наибольшее отклонение между ЭФР и ф.р. Неважно, как долго встречалось это отклонение, возникло ли оно на узком диапазоне или было при большом количестве x , критична здесь только величина максимального перепада.
- Критерий Крамера-Мизеса лучше реагирует на продолжительные по времени отклонения.
- Критерий Андерсона-Дарлинга фокусируется на отклонении при тех значениях, которые редки для предполагаемой ф.р.

На практике зачастую критерий Андерсона-Дарлинга выглядит заметно сильнее обоих конкурентов.

В Python критерий Колмогорова задан функция `kstest` из библиотеки `numpy` пакета `stats`. Критерий Крамера-Мизеса есть в `scipy.stats.cramervonmises`, а вот пакет `anderson` в том же пакете – это другой, сложный критерий (о таких пойдет речь ниже). Квантили распределения Андерсона-Дарлинга есть в приложенном файле `ADQuantiles.txt`.

1.3.1 Визуальный способ проверки

Нетрудно заметить, что проверку гипотезы $F = F_0$ для непрерывной F_0 можно свести к проверке гипотезы равномерности для непрерывных F_0 , просто применив F_0 к элементам выборки. Это позволяет осуществлять и визуальную проверку, близость к равномерному распределению вполне успешно идентифицируется графически.

Такого рода механизм предлагает так называемый `quantile-quantile plot`. Для его построения по оси абсцисс откладываются квантили распределения F_0 , а по оси ординат – упорядоченная выборка. В случае выполнения гипотезы график должен быть близок к прямой $y=x$.

В Python можно реализовать это с помощью `stats.probplot(массиве, dist=распределение)`

2 Проверка принадлежности параметрическим семействам

2.1 Сложный критерий хи-квадрат и критерий отношения правдоподобий

1. Критерий хи-квадрат

Критерий хи-квадрат в этом случае удастся модернизировать, если вместо неизвестных параметров подставить оценки ОМП для них.

Более конкретно, рассмотрим вероятности $p_i(\theta) = \mathbf{P}_\theta(\Delta_i)$. Подсчитаем количества попаданий ν_i в Δ_i . Найдем ОМП для θ по функции правдоподобия

$$L(x_1, \dots, x_n; \theta) = \prod_{i=1}^n p_i(\theta)^{\nu_i}.$$

Подставив в $p_i(\theta)$ полученную оценку для θ , можно найти новую статистику хи-квадрат. В этом случае можно показать, что полученная статистика будет иметь распределение χ_{k-l-1}^2 , где l – размерность параметра θ , k – число Δ_i .

Отметим, что важным является то, что оценивание проводится именно на основе группированных данных, то есть правдоподобие строится именно по ним. В противном случае теорема о предельном распределении неверна.

В случае сложного критерия мы не сможем выбирать интервалы Δ_i так, чтобы вероятности попадания в них были одинаковы, поскольку эти вероятности зависят от неизвестного параметра. Зачастую на практике выбирают интервалы так, чтобы в них попадали близкие количества наблюдений. Более правильный (и более эффективный метод группирования) называется асимптотическое оптимальное группирование. В этом случае группировка осуществляется так, чтобы информация Фишера (информационная матрица Фишера) по сгруппированным данным была как можно ближе к информации исходных данных. В случае скалярного параметра эта задача сводится к максимизации количества информации Фишера о параметре по группированной выборке

$$\max_{x_0 < x_1 < \dots < x_k} \sum_{i=1}^k \left(\frac{\partial \ln p_i(\theta)}{\partial \theta} \right)^2 p_i(\theta),$$

где $\Delta_i = [x_{i-1}, x_i)$, а в случае многомерного — максимизации определителя информационной матрицы Фишера. Здесь вместо неизвестных параметров используются их ОМП. Граничные точки разбиения для различных распределений представлены здесь.

В случае нормального распределения критерий хи-квадрат реализован, например, в пакете `nortest` в функции `pearson.test`.

2. Критерий отношения правдоподобий

Это более общий метод, позволяющий работать со сложными параметрическими моделями. Критерий хи-квадрат, в действительности, является аппроксимацией этого критерия.

Можно представлять этот критерий как некоторое обобщение критерия Неймана-Пирсона. Пусть $\theta = (\theta_1, \dots, \theta_r)$ и основная гипотеза

$$H_0 : (\theta_{q+1}, \dots, \theta_r) = (\theta_{0,q}, \dots, \theta_{0,r}),$$

то есть часть параметров фиксирована, а остальные произвольны. Тогда найдем статистику отношения правдоподобия

$$T = 2 \ln \frac{L(x_1, \dots, x_n, \hat{\theta})}{L(x_1, \dots, x_n, \hat{\theta}_0)},$$

где $\hat{\theta}_0$ — ОМП при H_0 , $\hat{\theta}$ — ОМП в общей параметрической модели. Оказывается, при H_0 в так называемых сильно регулярных моделях справедливо соотношение $T \xrightarrow{d} Y \sim \chi^2_{r-q}$, $n \rightarrow \infty$, откуда $T > y_{1-\alpha}$ задает асимптотический критерий уровня α . Этот результат называется теоремой Уилкса. Не будем останавливаться подробно на формулировке условий сильной регулярности, отметим лишь ключевые: правдоподобие должно достаточно гладко зависеть от параметра и область изменения параметра является открытым множеством

В частности, критерий работает для выборок из дискретного распределения, где параметричность модели уже не требуется.

Вопрос 3. Как будет выглядеть критерий отношения правдоподобий для дискретных выборок с k возможными значениями?

2.2 Критерии Колмогорова и омега-квадрат при сложной гипотезе

- Критерий Колмогорова-Смирнова также применим к сложной гипотезе при подстановке состоятельных оценок (например, ОМП). Однако в этом случае предельное распределение уже не будет колмогоровским, а будет своим для каждого класса распределений. Это замечание крайне важно и зачастую игнорируется прикладными исследователями.

Так для нормальных распределений при этом получится так называемое распределение Лиллиефорса (соответствующий критерий называют критерием Лиллиефорса).

В общем случае можно определить критическое множество с помощью метода Монте-Карло.

- Критерии Андерсона-Дарлинга и Крамера-Мизеса будут верны и в случае параметрических семейств, но опять-таки предельное распределение станет иным и будет зависеть от распределения

выборки.

В `scipy.stats.anderson` реализовано несколько критериев такого рода. Критерий Лиллиефорса реализован в `statsmodels.stats.lilliefors`.

3 О некоторых параметрических семействах

Для некоторых семейств существуют довольно мощные специализированные критерии.

1. Для нормальных распределений.

Начать можно с построения QQplot с помощью `statmodels qqplot`. Изучив отклонения от нормальности графически, мы можем применить методы. Опишем наиболее популярные.

(a) Метод Шапиро-Уилка, задающийся статистикой

$$\frac{(\sum_{i=1}^n a_i X_{(i)})^2}{\sum_{i=1}^n (X_i - \bar{X})^2},$$

где a_i — некоторые константы. В Python он задан `shapiro` из все того же пакета `stats`. Этот тест показывает наиболее хорошие результаты при общей альтернативе даже при небольших выборках. Его статистика достаточно громоздка.

(b) Как мы уже писали выше критерий Андерсона-Дарлинга для нормального распределения есть тест `anderson()` из `stats` в `scipy`.

(c) Хорошими критериями являются критерии, основанные на коэффициентах асимметрии Sk и эксцесса C . Они подменяют гипотезу о нормальности на гипотезы о том, что эксцесс и асимметрия распределения такая же, как и у нормального. Асимметрией и эксцессом называют

$$\frac{\mu_3}{\mu_2^{3/2}}, \quad \frac{\mu_4}{\mu_2^2} - 3,$$

где μ_i — i -й центрированный момент. Для их оценки естественно использовать аналогичные формулы с $\hat{\mu}_i = \overline{(X - \bar{X})^i}$:

$$Sk = \frac{\hat{\mu}_3}{\hat{\mu}_2^{3/2}}, \quad K = \frac{\hat{\mu}_4}{\hat{\mu}_2^2} - 3.$$

- Критерий Харке-Бера `scipy.stats.jarque_bera` использует статистику

$$JB = n \left(\frac{Sk^2}{6} + \frac{K^2}{24} \right).$$

- Критерием K^2 Д'Агостино (не путать с D-критерием Д'Агостино) называют критерий, основанный на асимметрии, использующий статистику

$$T_1 = C_1 \operatorname{asinh}(C_2 g_1),$$

где C_1, C_2 — некоторые параметры, зависящие от размера выборки. Эта величина асимптотически нормальна при выполнении основной гипотезы и смещена влево или вправо, если распределение асимметрично.

- Вторая часть критерия Д'Агостино связана с эксцессом (иногда эту часть называют критерием Анскомба-Глинна, поскольку эти авторы получили нормальную аппроксимацию для статистики) и использует асимптотическую нормальность статистики

$$T_2 = C_3 \left(C_4 - \sqrt[3]{\frac{C_5}{C_6 + K}} \right),$$

где C_3, C_4, C_5, C_6 — некоторые параметры, зависящие от размера выборки. Мы опускаем явный вид $C_1 - C_6$, поскольку эти параметры достаточно громоздки. Эта величина отлавливает отклонения от нормальности в пользу распределений с другим эксцессом.

- Можно использовать комбинацию $T_1^2 + T_2^2$. При выполнении гипотезы статистика критерия имеет предельное распределение χ^2_2 . В Python критерий Д’Агостино реализован в `normaltest` в `scipy.stats`.

2. Для экспоненциальных распределений.

Для экспоненциального распределения также существует ряд удачных критериев, в частности, адаптация критерия Шапиро-Уилка. Однако, в Python из известных мне реализован только тест Андерсона-Дарлинга (все той же функцией, только с другим параметром).

Материал ниже не рассказывался на лекции и не нужен для решения задач семинара. Однако он может пригодиться в вашем проекте.

4 Проверка множества гипотез

Предположим, что у нас есть цепочка гипотез $H_{0,i}$ против $H_{1,i}$, $i \leq k$. Если мы хотим получить итоговую вероятность ошибки I рода не больше α , то как организовать процесс?

Конечно, если статистики критериев независимы, то мы можем просто проверять каждую из гипотез на уровне значимости $1 - \sqrt[k]{1 - \alpha}$.

В общем случае возможны несколько методов:

4.1 Метод Бонферрони

Каждую из гипотез проверять на уровне $1 - \alpha/k$. Этот метод даст вероятность не более α того, что мы отвергнем хотя бы одну верную гипотезу.

Соответственно, фактический уровень значимости проверки всех наших гипотез мы оцениваем суммой фактических уровней значимости каждой.

4.2 Метод Беньямини-Хохберга

Если у нас нет необходимости не ошибаться, то мы можем наблюдать за долей ошибочно отвергнутых гипотез $H_{i,0}$. Предположим, что m_0 из гипотез H_0 верны, а остальные $m - m_0$ — нет. Пусть N — число отвергнутых гипотез H_0 , N_1 — число ошибочно отвергнутых H_0 . Тогда N_1/N называют FDP (False Discovery Proportion), где в случае $N = 0$ FDP=0.

Назовем FDR (False Discovery Risk) $\mathbf{E}(N_1/N)$ — среднее число ошибочно отвергнутых H_0 .

Тогда разумно рассматривать систему критериев, таких что при любом $m_0 \leq m$ $FDR \leq \alpha$, т.е. среднее число отвергнутых гипотез не больше α .

Метод Беньямини-Хохберга строит такую процедуру отвержения/принятия. Упорядочим фактические уровни значимости имеющихся критериев $p_{(1)} \leq p_{(2)} \leq \dots \leq p_{(k)}$. Положим $l_i = i\alpha/(kC_k)$, $C_k = \sum_{i=1}^k i^{-1}$ в случае зависимых критериев и $C_k = 1$ иначе. Тогда положим $R = \max\{i : p_{(i)} < l_i\}$, $P = p_{(R)}$. Метод предлагает отвергать те из $H_{0,i}$, для которых $p_i < P$.

Пример 2. Пусть p-value 10 критериев приняли значения 0.00017, 0.00448, 0.00671, 0.00907, 0.01220, 0.33626, 0.39341, 0.53882, 0.58125, 0.98617.

Мы хотим проверить эту совокупность на уровне значимости 0.05. Тогда в случае метода Бонферрони мы отвергнем те, у которых p-value меньше 0.005, то есть первые 2.

В случае метода Беньямини-Хохберга в общем случае мы считаем $C_k = 2.92$, $l_i = i\alpha/29.2$, сравнивая $p_{(i)}$ и l_i , убеждаемся, что гипотеза отвергается лишь в первом случае.

В случае независимости в том методе $l_i = i\alpha/k$ и гипотеза отвергнется в первых 5 ситуациях.

Вопрос 4. После операции у многих людей появляются неприятные ощущения (nausea), которые можно снять некоторыми лекарствами. Известно, что плацебо помогает в 55% случаев. Проверить эффективность лекарств на уровне 0.05, пользуясь а) методом Бонферрони, б) методом Беньямини-Хохберга

	Number of Patients	Incidence of Nausea
Chlorpromazine	75	26
Dimenhydrinate	85	52
Pentobarbital (100 mg)	67	35
Pentobarbital (150 mg)	85	37

Ответы на вопросы

Ответ 1. Фактический уровень значимости равен $\mathbf{P}_{H_0}(\bar{X} > 3) = 1 - \Phi(3) \approx 0.0013$.

Ответ 2. Предположим, что верна альтернатива, то есть $F = F_1 \neq F_0$. Тогда \hat{F}_n сходится к F_1 , а значит $\sup |\hat{F}_n(x) - F_0(x)|$ сходится к $d = \sup |F_1(x) - F_0(x)| > 0$. Значит $\sqrt{n}D_n \rightarrow \infty$, $\mathbf{P}_{F_1}(\sqrt{n}D_n > c) \rightarrow 0$ при всех c .

Ответ 3. Наиболее очевидным контрпримером является случай, когда наша подвыборка имеет тот же размер, что вся выборка. Тогда набор p-value будет состоять из одного и того же числа. В случае, если подвыборки будут меньше размера, этот эффект будет не так заметен, но все же p-value будут зависимы через то, что они считаются по одной и той же выборки. Тем не менее, если исходная выборка будет большого размера, а подвыборки не столь большого, то мы добьемся нужного эффекта.

Ответ 4. Если лекарства эффективны, то для бернуллиевских выборок $X_i \sim \text{Bern}(p)$, представленных в таблице, верно $H_1 : p > 0.55$, а если нет, то $H_0 : p = p_0 = 0.55$. Возьмем в качестве критерия $\bar{X} > c$, при гипотезе в силу ЦПТ $\mathbf{P}(\bar{X} > c)$ сходится к $\Phi\left(\sqrt{n}\frac{c-p_0}{\sqrt{p_0(1-p_0)}}\right)$. Подсчет p-value дает нам 0.03, 0.99, 0.88, 0.39. Следовательно, в методе Бонферрони мы сравниваем их с $\alpha/4 = 0.0125$ и не отвергаем. В методе Беньямини-Хохберга мы получим $C_k \approx 2$, $l_i = i\alpha/8$, что меньше любого из заданных p-value. Увы, ни про одно из лекарств мы не можем уверенно утверждать, что оно работает.