Математический институт им. В. А. Стеклова
Российской академии наук

Лекционные курсы НОЦ

Выпуск 14

Издание выходит с 2006 года

Д. М. Чибисов

Лекции по асимптотической теории ранговых критериев

МИАН

Москва
2009

УДК 519.23
ББК (В)22.17
Л43

Л43 **Лекционные курсы НОЦ**/ Математический институт им. В. А. Стеклова РАН (МИАН). – М.: МИАН, 2009. Вып. 14: Лекции по асимптотической теории ранговых критериев / Чибисов Д. М. – 176 с.

ISBN 5-98419-035-4

Серия “Лекционные курсы НОЦ” – рецензируемое продолжающееся издание Математического института им. В. А. Стеклова РАН. В серии “Лекционные курсы НОЦ” публикуются материалы специальных курсов, прочитанных в Математическом институте им. В. А. Стеклова Российской академии наук в рамках программы Научно-образовательный центр МИАН.

Настоящая брошюра содержит курс лекций “Лекции по асимптотической теории ранговых критериев” Д. М. Чибисова, прочитанный в 2005 году в Научно-образовательном центре МИАН.

ISBN 5-98419-035-4

Оглавление

Предисловие

Настоящие записки лекций написаны на основе курса, прочитанного в 2005 г. в рамках Научно-образовательного центра при Математическом институте РАН. Курс преследовал двоякую цель: во-первых, изложить основные результаты, касающиеся свойств ранговых критериев. В первую очередь – это предельные теоремы для распределений ранговых статистик и как следствие результаты об асимптотической мощности и эффективности ранговых критериев. Во-вторых, ознакомить слушателя с развитыми Л. Ле Камом в 1960-х гг. асимптотическими методами, которые и позволили в свое время построить асимптотическую теорию ранговых критериев и которые представляют значительный самостоятельный интерес, будучи применимы в разнообразных областях математической статистики.

При написании этого курса я старался сделать его максимально доступным для студента-старшекурсника в отношении требуемого уровня знаний с учетом разницы в объеме и уровне преподавания теории вероятностей и математической статистики в различных вузах. В качестве базового курса теории вероятностей я ориентировался на известный учебник Б. В. Гнеденко. Все используемые в лекциях результаты, не содержащиеся в этом учебнике, приводятся со ссылками на литературу, иногда с доказательствами. Приводятся также сведения из теории проверки гипотез, включая лемму Неймана–Пирсона с примерами ее применения. Эти статистические сведения, которые предположительно хорошо известны большинству читателей, так же как и упомянутые вспомогательные результаты, собраны в заключительной главе «Приложение». В связи с этим порядок изложения материала изменен по сравнению с курсом лекций.

Автор благодарен коллективу Математического института за творческую обстановку и поддержку, содействовавшие написанию данного курса. Особую признательность автор выражает О. В. Вискову и А. А. Лосеву за полезные обсуждения и замечания.

Глава 1. Введение

§ 1.1. Понятие о ранговых критериях

Типичной статистической задачей, где применяются ранговые критерии, является задача сравнения двух выборок. Предположим, что проведен эксперимент, имеющий целью выявить, дает ли эффект некоторое «воздействие». Объектами воздействия могут быть подопытные животные в биологии, пациенты в медицине, растения в сельском хозяйстве и т.п. Воздействие состоит, например, в применении новой методики лечения, агротехнического приема и пр. Эффект «воздействия» должен заключаться в увеличении (или уменьшении) некоторого числового показателя, которым характеризуются участвующие в эксперименте объекты (повышении урожайности, снижении содержания холестерина в крови, и т.п.). Будем считать для определенности, что целью воздействия является увеличение показателя. Объекты разделяются на две группы: экспериментальную и контрольную. К объектам первой группы применяется данное «воздействие», к объектам второй – нет. Предположим, что в результате эксперимента получены значения $X_1, \ldots, X_{n_1}$ показателя в экспериментальной группе и $X_{n_1+1}, \ldots, X_{n_1+n_2}$ в контрольной, n_1 и n_2 – размеры групп. Как проверить наличие эффекта «воздействия»?

Наиболее распространенным методом является применение критерия Стьюдента. Если предположить, что наблюдения X_1, $\ldots, X_N$, $N = n_1 + n_2$, являются реализациями независимых случайных величин, имеющих нормальные распределения $\mathcal{N}(\mu_1, \sigma^2)$ для первой группы (первых n_1 наблюдений) и $\mathcal{N}(\mu_2, \sigma^2)$ для второй (последних n_2 наблюдений), то вопрос о наличии эффекта «воздействия» формулируется как задача проверки статистической гипотезы $H_0\colon \mu_1 = \mu_2$ (или $H_0\colon \mu_1 \leqslant \mu_2$) против альтернативы $H_1\colon \mu_1 > \mu_2$. Принятие гипотезы H_0 трактуется как отсутствие положительного эффекта, а отклонение гипотезы, т.е. принятие H_1 – как свидетельство наличия эффекта. При сделанных предположениях для проверки гипотезы H_0 применяется известная статистика Стьюдента, имеющая при H_0 стандартное распределение Стьюдента с $N-2$ степенями свободы. Эта статистика представляет собой разность выборочных средних, нормированную множителем, включающим оценку неизвестной дисперсии σ^2.

Основным ограничением в использовании критерия Стьюдента является предположение о нормальности распределения наблюдений. При нарушении этого предположения статистика Стьюдента меняет свои свойства, и особенно «губительным» для нее является наличие у распределения «тяжелых хвостов».

Для наглядного пояснения этого понятия приведем игрушечный пример. Представим себе студенческую Спартакиаду по легкой атлетике, в рамках которой проводятся соревнования двух команд по бегу. Пусть команды равны по численности, и пусть наши выборки $X_1, \ldots, X_{n_1}$ и $X_{n_1+1}, \ldots, X_{n_1+n_2}$, $n_1 = n_2$, – это времена, показанные участниками первой и второй команд. Как определять команду-победительницу? Способ, идейно аналогичный критерию Стьюдента, состоит в сравнении команд по суммарному времени. Но представим себе картину подобных соревнований. Сначала финиширует небольшая плотная группа сильнейших, затем приходят к финишу «середняки», и потом с большим отставанием и большими разрывами между собой заканчивают дистанцию слабейшие. Понятно, что именно их времена дадут наибольший вклад в суммарное время команды. И понятно, что система, при которой первенство определяется показателями слабейших участников, несправедлива. Более разумным представляется сравнение по сумме занятых мест (что обычно и делается в спорте). Идея использования вместо числовых значений наблюдений их мест в общем вариационном ряду лежит в основе ранговых методов.

Более формально, «тяжелые хвосты» – это медленное убывание плотности распределения на бесконечности, при котором у распределения могут не существовать дисперсия или моменты еще более низких порядков. Хорошо известный пример такого распределения – распределение Коши, у которого не существует математического ожидания. Подобные распределения встречаются во многих приложениях, например, в биологии, финансовой математике и др. Применение критерия Стьюдента в таких случаях не оправдано не только потому, что распределение его статистики при гипотезе H_0 резко отличается от стандартного, но и потому, что при таких распределениях он становится неэффективным. В нормальном случае при росте объемов выборок дисперсии выборочных средних убывают, что повышает чувствительность критерия к различию средних. В случае же, например, распределения Коши, как известно, выборочное среднее распределено так же, как отдельное наблюдение, т.е. проведение повторных испытаний не увеличивает чувствительности критерия. Иными словами, если наблюдения подчиняются распределению Коши, то критерий Стьюдента не является состоятельным.

Подобных недостатков лишены ранговые критерии, в частности, самый простой и наиболее распространенный из них – критерий Уилкоксона–Манна–Уитни, или критерий суммы рангов. Статистика Уилкоксона [26] – это аналог суммы мест в вышеописанных соревнованиях. Для ее построения следует расположить наблюдения в общий вариационный ряд $X^{(1)} < \cdots < X^{(N)}$, т.е. выписать все $N = n_1 + n_2$ наблюдений в порядке возрастания, отметить среди них те, которые соответствуют элементам 1-й выборки, и сложить их номера («ранги») $R_1, \ldots, R_{n_1}$ в этом ряду. Понятно, что если «воздействие» действительно увеличивает исследуемый показатель, то элементы 1-й выборки будут иметь тенденцию к расположению на более далеких местах в вариационном ряду и тем самым сумма рангов будет иметь тенденцию к увеличению. Дальнейшие действия, конечно, отличаются от выявления команды-победительницы. Прежде всего надо найти распределение нашей статистики

$$W = \sum_{i=1}^{n_1} R_i \tag{1.1.1}$$

при гипотезе отсутствия эффекта «воздействия». Теперь (без предположения нормальности) это гипотеза $H_0\colon F_1 = F_2$, где F_1 и F_2 – функции распределения элементов 1-й и 2-й выборок соответственно. Об альтернативе H_1 поговорим чуть позже. Будем предполагать, что функции распределения F_1 и F_2 непрерывны, тогда все наблюдения с вероятностью 1 различны (чем и оправданы строгие знаки неравенства между членами вариационного ряда, выписанные выше)[1]. Ниже будет показано, что при гипоте-

[1]На практике часто возникают равные между собой наблюдения, например, вследствие округления численных данных (такие равенства называют *связями*, от англ. *ties*). Вопрос о применении ранговых критериев при наличии связей обсуждался в литературе. Соответствующие рекомендации можно найти, например, в книгах Гаека и Шидака [3] и Холлендера и Вулфа [13]. Последняя представляет собой руководство по непараметрической статисти-

зе H_0 распределение статистики W не зависит от распределения $F = F_1 = F_2$ (в предположении его непрерывности). Тем самым нулевое распределение W может быть раз и навсегда вычислено и табулировано. Это позволяет нам для заданного $\alpha > 0$, называемого уровнем значимости критерия, найти процентную точку C_α, для которой

$$P_0(W > C_\alpha) = \alpha$$

(индекс 0 означает, что вероятность вычисляется при выполнении гипотезы H_0; указанное равенство выполняется, вообще говоря, приближенно в силу дискретности распределения статистики W, эту проблему, общую для всех дискретных распределений, мы здесь не обсуждаем). Теперь, если фактически наблюдаемое значение W превышает C_α, то гипотеза H_0 отвергается на уровне значимости α и считается, что результаты эксперимента свидетельствуют о наличии эффекта «воздействия».

Что касается чувствительности критерия к отклонениям от H_0, то Манном и Уитни [25] было показано, что предложенный ими в этой работе критерий (эквивалентный, как было выяснено, критерию Уилкоксона) состоятелен против альтернатив, при которых элементы одной выборки стохастически больше элементов другой. (Из двух случайных величин с функциями распределения $F_1(x)$ и $F_2(x)$ первая стохастически больше второй, если $F_1(x) < F_2(x)$ при всех x.) По содержательному смыслу это как раз то, что требуется в прикладных задачах типа описанной выше. Более детальное исследование асимптотической мощности ранговых критериев составляет одну из основных задач настоящего курса.

Статистика W относится к классу простых линейных ранговых статистик (ПЛРС), которые являются наиболее употребительными в практике. Общий вид ПЛРС[2] следующий: заданы два набора чисел $c_1, \dots, c_N$ и $a(1), \dots, a(N)$ и статистика имеет вид

$$S = \sum_{i=1}^{N} c_i a(R_i). \tag{1.1.2}$$

ке для прикладников. В 1999 г. вышло 2-е издание этой книги, значительно расширенное и дополненное. В настоящее время готовится перевод этого издания на русский язык.

[2] В данном курсе рассматривается только этот класс ранговых статистик, поэтому мы зачастую будем называть их просто ранговыми статистиками.

В рассмотренной выше статистике Уилкоксона $a(i) = i$ для $i = 1, \ldots, N$ и $c_i = 1$ при $i = 1, \ldots, n_1$, $c_i = 0$ при $i = n_1 + 1, \ldots, N$. Числа c_i называются *регрессионными коэффициентами*, а $a(i)$ – *метками*. Выбор меток позволяет настраивать статистику против тех или иных ожидаемых распределений F (этот вопрос будет рассмотрен в п. 6.2.4). Выбор коэффициентов c_i связан с видоизменениями самой постановки задачи. Для иллюстрации рассмотрим случай, когда предполагаемым отклонением от гипотезы H_0 одинаковой распределенности является наличие тренда, скажем, тенденции к возрастанию наблюдаемых значений с ростом индекса (который может пониматься как время). Простейшей ранговой статистикой для обнаружения линейного тренда может служить

$$S = \sum_{i=1}^{N} iR_i. \tag{1.1.3}$$

В зависимости от предполагаемого характера тренда могут использоваться и другие наборы c_i.

§ 1.2. Обзор содержания данного курса

В главе 2 изучаются основные свойства рангов и порядковых статистик при выполнении гипотезы H_0, используемые в дальнейшем.

В главе 3 доказывается теорема об асимптотической нормальности ранговых статистик (1.1.2). Как отмечалось выше, для построения критерия, основанного на какой-либо статистике (в нашем случае – на статистике вида (1.1.2)), требуется прежде всего знание распределения статистики при гипотезе H_0. Распределения наиболее употребительных ранговых статистик достаточно полно табулированы. Однако вычисление точных распределений ранговых статистик быстро усложняется с ростом числа наблюдений, поэтому при больших объемах выборок требуются аппроксимации для этих распределений. Основой для таких аппроксимаций служит теорема об асимптотической нормальности ПЛРС, доказываемая в гл. 3. Приводимое доказательство следует доказательству Я. Гаека [3], который заметил, что (при определенных условиях на константы c_i и $a(i)$) ранговая статистика вида (1.1.2) аппроксимируется суммой независимых (неодинаково распределенных) случайных величин, удовлетворяющих условию Линдеберга. Эта аппроксимация позволяет вывести центральную предельную теорему (ц.п.т.) для ранговой статистики из ц.п.т. для сумм независимых слагаемых.

В конце этой главы (§ 3.3) приводятся примеры наиболее употребительных ранговых статистик, демонстрирующие применение теоремы об асимптотической нормальности. Впоследствии эти статистики используются для иллюстрации получаемых асимптотических результатов.

Следующей задачей при исследовании любого класса критериев является изучение их мощности. Для этого требуется уметь находить распределения статистик критериев при альтернативах, т.е. при отклонениях от гипотезы H_0. Вычисление точных распределений ранговых статистик в этом случае гораздо сложнее, чем при нулевой гипотезе, и попытки таких вычислений при малых объемах выборок не принесли сколько-нибудь обозримых результатов. Поэтому здесь имеет смысл только нахождение асимптотической мощности при больших (стремящихся к бесконечности) объемах выборок. К этой задаче проявлялся значительный интерес в 50–60-х гг. прошлого века, но попытки ее решения оставались безуспешными, пока тот же Я. Гаек в конце 60-х гг. не получил ее исчерпывающего решения, применив развитые в начале 60-х гг. Л. Ле Камом[3] асимптотические методы, основанные на понятиях *контигуальности* и *локальной асимптотической нормальности.*

Эти понятия и основные результаты, связанные с ними, излагаются в настоящем курсе в главах 4 и 5. Здесь мы кратко поясним содержание этих глав. Для иллюстрации рассмотрим статистику Уилкоксона W_N (1.1.1) (имея в виду асимптотику при $N \to \infty$, мы снабжаем обозначение статистики индексом N). Будем предполагать, что объемы выборок n_1 и n_2 сопоставимы, т.е. их отношение стремится при $N \to \infty$ к положительной константе. Предположим, далее, что эффект «воздействия» носит аддитивный характер, т.е. элементы первой (экспериментальной) выборки распределены так, как элементы второй (контрольной) выборки, к каждому из которых прибавлено одно и то же положительное слагаемое θ. Иначе говоря, если $F(x)$ – функция распределения каждого из элементов второй выборки, то элементы

[3] Л. Ле Кам (L. Le Cam) – американский математик французского происхождения, внес значительный вклад в развитие асимптотических методов математической статистики.

первой выборки имеют функцию распределения $F(x-\theta)$. При фиксированном $\theta > 0$ элементы первой выборки стохастически больше элементов второй выборки (наглядно это видно из того, что они отличаются положительным слагаемым, а формально это следует из неравенства $F(x-\theta) < F(x)$). Согласно приведенному выше результату Манна и Уитни [25] критерий Уилкоксона состоятелен против таких альтернатив, т.е. его мощность стремится к 1 при каждом фиксированном $\theta > 0$. Свойство состоятельности является отправной точкой асимптотического анализа любой статистической процедуры: если с ростом числа наблюдений оценка параметра не сходится к его истинному значению или критерий не обнаруживает альтернативу интересующего нас вида с вероятностями ошибок 1-го и 2-го рода, стремящимися к нулю, то дальнейшее исследование асимптотических свойств такой процедуры вряд ли имеет смысл. Зная, однако, что, скажем, критерий Уилкоксона обладает приведенным выше свойством состоятельности, экспериментатор ставит более конкретный вопрос: если мы зададимся (малым) числом $\Delta > 0$, смысл которого заключается в том, что эффект «воздействия», меньший Δ, практически незначим, то сколько надо провести наблюдений, чтобы обнаружить эффект «воздействия» $\theta \geqslant \Delta$ с заданными вероятностями ошибок 1-го и 2-го рода? Или: при данной вероятности ошибки 1-го рода (уровне значимости) α и данном числе наблюдений $N = n_1 + n_2$ какова вероятность ошибки 2-го рода (или мощность критерия)[4] в зависимости от величины «воздействия» θ? Во всяком случае прикладник имеет дело с фиксированным конечным числом наблюдений (даже если оно было вычислено заранее) и асимптотика при $N \to \infty$ – это математический способ дать приближенное решение практической задачи. А именно, конкретная практическая задача вкладывается в последовательность подобных задач с растущим N, для которой оказывается возможным найти асимптотическое решение, применяемое как приближенное для задачи с конечным объемом выборки. Возможны различные асимптотические постановки задачи, в частности, известны подходы, при которых эффективность критерия характеризуется

[4] Мы напоминаем понятия, связанные с задачей проверки гипотез, в § 8.1 Приложения (глава 8). Укажем здесь лишь, что мощность $= 1 -$ (вероятность ошибки 2-го рода), и это соотношение позволяет пользоваться любым из этих понятий в зависимости от контекста. В дальнейшем мы чаще будем говорить о мощности критерия.

скоростью убывания вероятности ошибки 1-го или 2-го рода при фиксированной альтернативе. Обзор различных понятий эффективности можно найти в книге Я. Ю. Никитина [9]. Мы будем рассматривать асимптотику, когда при $N \to \infty$ вероятности ошибок 1-го и 2-го рода не стремятся к нулю, а ограничены снизу положительной константой. Как правило, вероятность ошибки 1-го рода будет предполагаться сходящейся к заданному $\alpha > 0$. Математически удобно рассматривать такие альтернативы, при которых вероятность ошибки 2-го рода также стремилась бы к константе, заключенной между 0 и 1. Мы видели, что этим свойством не обладают альтернативы, определяемые фиксированным значением $\theta > 0$, поскольку тогда вероятность ошибки 2-го рода стремится к нулю. То же будет происходить, если θ стремится к нулю слишком медленно. Сходимость этой вероятности (или мощности) к пределу, отличному от 0 и 1, будет иметь место, если при каждом N рассматривать эффект «воздействия» $\theta_N = t/\sqrt{N}$, где $t > 0$ – фиксированное число[5].

Отметим, что в асимптотической постановке рассматривается *последовательность статистических задач*, а именно, при каждом N имеется свое выборочное пространство (в данном случае $\mathbb{R}^N$), на котором заданы два семейства распределений $\mathcal{P}_N$ и $\mathcal{Q}_N$, отвечающих гипотезе и альтернативе. Результат наблюдений описывается точкой выборочного пространства (в данном случае $\mathbf{X}_N = (X_1, \ldots, X_N)$), по которой нужно решить, принадлежит ли ее распределение семейству $\mathcal{P}_N$ (т.е. верна гипотеза H_0) или семейству $\mathcal{Q}_N$ (т.е. верна альтернатива H_1). Упрощая задачу, мы задаем распределение F и тем самым (по предположению об одинаковой распределенности и независимости наблюдений при H_0) – распределение вектора наблюдений $\mathbf{X}_N$ при гипотезе H_0, а задавая кроме того набор параметров, характеризующих отклонение распределения при H_1 от распределения при H_0 (в данном случае – объемы выборок n_1 и n_2 и параметр сдвига θ), задаем распределение $\mathbf{X}_N$ при альтернативе H_1. Иначе говоря,

[5] Такая скорость убывания параметра имеет место при определенных условиях регулярности на функцию распределения $F(x)$, которые будут подробно рассматриваться в главе 7. Иначе, если F имеет какие-либо особенности, возможно обнаружение более близких альтернатив. Например, если $F(x)$ разрывна (скажем, в случае равномерного или экспоненциального распределений), а θ, как и выше, – параметр сдвига, то нетривиальная предельная мощность достигается при альтернативах вида $\theta_N = t/N$. Подобного рода «суперэффективность» в настоящем курсе рассматриваться не будет.

мы выделяем конкретные распределения $P_N \in \mathcal{P}_N$ и $Q_N \in \mathcal{Q}_N$ и рассматриваем задачу проверки простой гипотезы о том, что распределение $\mathbf{X}_N$ есть P_N против простой альтернативы о том, что это распределение есть Q_N. Таким образом, хотя ранговые критерии предназначены для проверки гипотез в ситуации отсутствия достоверных знаний о форме распределения F (в частности, при отказе от предположения нормальности), задание распределения F позволяет изучать мощностные свойства ранговых критериев в зависимости от того, какое распределение имеет место на самом деле.

Вернемся к вопросу о скорости сближения между гипотезой и альтернативой, при которой вероятности ошибок 1-го и 2-го рода остаются отделенными от 0 и 1. Мы исходили из свойства состоятельности конкретного критерия и говорили далее о его ошибках 1-го и 2-го рода. Но понятно, что желательное свойство последовательности пар распределений (P_N, Q_N) – невозможность их различения с вероятностями ошибок 1-го и 2-го рода, стремящимися к нулю, – не должно зависеть от выбора (возможно, неудачного) того или иного критерия, а должно быть присуще самой этой последовательности. В главе 4 приводится точное определение этого свойства (при его выполнении говорят, что последовательности P_N и Q_N взаимно *контигуальны*[6]) и доказывается общая теорема, дающая для него необходимые и достаточные условия в терминах распределений отношения правдоподобия P_N и Q_N. Отметим важное следствие свойства контигуальности P_N и Q_N: всякая последовательность случайных величин, сходящаяся к нулю по вероятности P_N, сходится к нулю и по вероятности Q_N. Это означает, в частности, что аппроксимация ранговой статистики суммой независимых случайных величин, построенная в гл. 3, остается в силе при контигуальных альтернативах, т.е. ее погрешность также стремится к нулю по вероятности.

В главе 4 не делается никаких предположений о структуре выборочных пространств, на которых определены вероятностные меры, и точек этих пространств, которые служат математическим выражением совокупности статистических данных. В следующей главе 5 рассматривается случай, соответствующий нашей статистической задаче, а именно, при каждом N в качестве выборочного пространства рассматривается пространство $\mathbb{R}^N$ точек

[6]Рассматривают также контигуальность одной последовательности относительно другой, см. гл. 4.

$\mathbf{X}_N = (X_1, \ldots, X_N)$, образующих вектор независимых случайных величин относительно каждого из распределений P_N и Q_N, одинаково распределенных в первом случае (P_N) и, вообще говоря, неодинаково распределенных во втором (Q_N). Мы предполагаем, что меры P_N и Q_N абсолютно непрерывны по мере Лебега на $\mathbb{R}^N$, т.е. имеют плотности распределения p_N и q_N, такие, что для любого борелевского множества $A \in \mathbb{R}^N$

$$P_N(A) = \int_A p_N(\mathbf{x})\, d\mathbf{x}, \quad Q_N(A) = \int_A q_N(\mathbf{x})\, d\mathbf{x},$$

где интегралы берутся по мере Лебега. Предполагается, что каждое отдельное наблюдение X_i, $i = 1, \ldots, N$, есть случайная величина с плотностью распределения $p(x, \theta)$, $\theta \in \Theta \subset \mathbb{R}$, по мере Лебега на прямой, причем распределение P_N образовано плотностями с одним и тем же значением $\theta = \theta_0$. Для простоты обозначений будем считать, что $\theta_0 = 0$ и опускать этот аргумент в обозначении плотности, т.е. писать $p(x) = p(x, \theta_0)$. Тогда совместная плотность p_N распределения P_N равна

$$p_N(\mathbf{x}) = \prod_{i=1}^{N} p(x_i), \quad \mathbf{x} = (x_1, \ldots, x_N).$$

Для описания альтернативного распределения Q_N, имея в виду отклонения от одинаковой распределенности более общего вида, чем введенные ранее в связи с критерием Уилкоксона, будем предполагать, что при каждом N задан набор чисел $c_{N,1}, \ldots, c_{N,N}$, подчиненных условиям нормировки $\sum_{i=1}^N c_{N,i}^2 = 1$ и равномерной малости $\max_i |c_{N,i}| \to 0$ при $N \to \infty$. Тогда в качестве альтернативы к распределению P_N будут рассматриваться распределения $Q_N = P_{N,t}$, $t \in \mathbb{R}$, с плотностью

$$p_{N,t}(\mathbf{x}) = \prod_{i=1}^{N} p(x_i, c_{N,i}t), \quad \mathbf{x} = (x_1, \ldots, x_N).$$

Таким образом, мы рассматриваем *семейство* альтернативных распределений, зависящее от параметра t, граничной точкой которого при $t = 0$ является распределение P_N, отвечающее нулевой гипотезе. Указанная форма введения малых нормирующих множителей в виде набора $\{c_{N,i}\}$ охватывает как случай, когда наблюдения одинаково распределены и при гипотезе и при альтернативе (при $c_{N,i} = 1/\sqrt{N}$, $i = 1, \ldots, N$), рассматриваемый здесь безотносительно к ранговым критериям, так и различные виды разнораспределенности, к которым применимы ранговые критерии.

Для простоты изложения в настоящем обзоре будем предполагать, что множества $\{x\colon p(x,\theta)>0\}$ не зависят от θ. При этом предположении распределения, отвечающие различным значениям θ, взаимно абсолютно непрерывны и при этом отношение правдоподобия $p(x,\theta)/p(x,0)$ определено на указанном выше множестве, имеющем вероятность 1 относительно любого из этих распределений.

Рассмотрим отношение правдоподобия (ОП) совместных распределений

$$L_{N,t}(\mathbf{X}_N)=\frac{p_{N,t}(\mathbf{X}_N)}{p_N(\mathbf{X}_N)}=\prod_{i=1}^{N}\frac{p(X_i,c_{N,i}t)}{p_N(X_i)},$$

аргументом которого служит случайная точка $\mathbf{X}_N=(X_1,\dots,X_N)$ выборочного пространства, имеющая одно из распределений $P_{N,t}$ или P_N. Обозначим для краткости[7] $l(x,\theta)=\log p(x,\theta)$, $l(x)=\log p(x)=\log p(x,0)$. Тогда логарифм ОП $\Lambda_{N,t}(\mathbf{X}_N)=\log L_{N,t}(\mathbf{X}_N)$ имеет вид

$$\Lambda_{N,t}(\mathbf{X}_N)=\sum_{i=1}^{N}[l(X_i,c_{N,i}t)-l(X_i)]. \tag{1.2.1}$$

Обозначим $l^{(j)}(x,\theta)=\frac{\partial^j}{\partial\theta^j}l(x,\theta)$, $l^{(j)}(x)=l^{(j)}(x,0)$, $j=1,2,\dots$. Разложим каждую разность под знаком суммы по формуле Тейлора по второму аргументу:

$$l(X_i,c_{N,i}t)-l(X_i)=c_{N,i}tl^{(1)}(X_i)+\frac{1}{2}c_{N,i}^2t^2l^{(2)}(X_i)+\cdots. \tag{1.2.2}$$

Обозначим

$$Z_N=\sum_{i=1}^{N}c_{N,i}l^{(1)}(X_i),$$

$$I(\theta)=E_\theta(l^{(1)}(X_1,\theta))^2=\int_{\mathbb{R}}(l^{(1)}(x,\theta))^2\,p(x,\theta)\,dx.$$

Величину $I(\theta)$ называют *фишеровской информацией* (о параметре θ, содержащейся в семействе $p(x,\theta)$). Следуя нашему соглашению, будем писать $I=I(0)$. Вспомним равенства $E\,l^{(1)}(X_1)=0$,

[7] В настоящем курсе log обозначает натуральный логарифм.

$E\,l^{(2)}(X_1) = -E(l^{(1)}(X_1))^2 = -I$, которые возникают в курсе математической статистики, например, при доказательстве неравенства Рао–Крамера. Подставляя (1.2.2) в (1.2.1), получаем для $\Lambda_{N,t}$ представление

$$\Lambda_{N,t}(\mathbf{X}_N) = tZ_N - \frac{1}{2}t^2 I + \eta_{N,t}, \tag{1.2.3}$$

где $P_N(|\eta_{N,t}| > \varepsilon) \to 0$ при каждом t для любого $\varepsilon > 0$ (иначе говоря, $\eta_{N,t} \to 0$ по P_N-вероятности). Поясним здесь вывод соотношения (1.2.3) в случае, когда $c_{N,i} = 1/\sqrt{N}$, $i = 1, \ldots, N$ (общий случай будет рассмотрен в главах 5–7). В этом случае соотношение (1.2.3) аналогично соотношениям, из которых в курсах математической статистики выводится асимптотическая нормальность оценки максимального правдоподобия и может быть доказано аналогичным методом. Обычно в курсах статистики предполагается, что существует 3-я производная $l^{(3)}(x,\theta)$, мажорированная по θ из некоторой окрестности нуля Δ функцией, имеющей конечное математическое ожидание, т.е. предполагается, что существует функция $M(x)$, $EM(x) < \infty$, такая, что $|l^{(3)}(x,\theta)| \leqslant M(x)$ при $\theta \in \Delta$. Тогда при подстановке (1.2.2) в (1.2.1) сумма первых слагаемых в (1.2.2) дает tZ_N просто по обозначению Z_N, сумма вторых слагаемых содержит сумму $\frac{1}{n}\sum_{i=1}^N l^{(2)}(X_i) \xrightarrow{p} E\,l^{(2)}(X_1) = -I$ (по закону больших чисел) с общим множителем $\frac{1}{2}t^2$, а остаточный член $\eta_{N,t}$ включает разность между этой суммой и ее пределом $-I$ и сумму не выписанных явно остаточных членов в формуле (1.2.2), содержащих $l^{(3)}(X_i,\theta^*)$ в промежуточной точке θ^* с множителем $N^{-3/2}$, обеспечивающим сходимость этой суммы к нулю. Указанным способом (с небольшими видоизменениями) представление (1.2.3) может быть доказано для многих употребительных в статистике семейств распределений $p(x,\theta)$, поскольку они, как правило, допускают дифференцирование по параметру нужных порядков и обладают необходимым числом моментов. Предлагаем читателю в качестве упражнения провести такие доказательства для различных семейств.

Тем не менее можно привести простые примеры плотностей, у которых существует весьма ограниченное число производных по параметру. Рассмотрим, например, гамма-распределение с плотностью

$$p_\alpha(x) = \frac{1}{\Gamma(\alpha)} x^{\alpha-1} e^{-x} \text{ при } x > 0, \quad p_\alpha(x) = 0 \text{ при } x \leqslant 0,$$

и образуем семейство $p_\alpha(x-\theta)$, $\theta \in \mathbb{R}$, сдвигов этого распределения. Функция $p_\alpha(x-\theta)$ дифференцируема по θ при $\alpha > 2$, дважды дифференцируема по θ при $\alpha > 3$, и т.д. Приведенный выше способ доказательства требует как минимум двукратной дифференцируемости, т.е. условия $\alpha > 3$. В главе 7 представление (1.2.3) доказывается для альтернатив, задаваемых описанными выше наборами $\{c_{N,i}\}$, по существу, при минимальных условиях на $p(x,\theta)$. А именно, требуется дифференцируемость (однократная) функции $l(x,\theta)$ по θ и конечность и непрерывность фишеровской информации $I(\theta)$ в некоторой окрестности точки $\theta = 0$. В случае, когда θ – параметр сдвига, $I(\theta) = I$ не зависит от θ и единственным условием является конечность фишеровской информации $I < \infty$. В приведенном выше примере это означает, что (1.2.3) выполняется при $\alpha > 2$. (Отметим, что иногда для простоты изложения мы вводим условие, что носитель плотности $p(x,\theta)$ не зависит от θ. В общем случае это требование излишне и результаты главы 7 получены без этого предположения, что позволяет применить их к примеру о сдвигах гамма-распределения.)

Вернемся теперь к главе 5. В п. 5.2.2 на основе соображений, подобных приведенным выше, проводится несколько более подробный вывод соотношения (1.2.3), а в п. 5.2.3 формулируется теорема, устанавливающая это соотношение при упомянутых выше минимальных условиях. Доказательство этой теоремы, как было сказано, отложено до главы 7, а сама глава 5 (а также глава 6) посвящены выводу статистических следствий из соотношения (1.2.3). Доказательства в гл. 7 относительно сложны, и читатель, не заинтересованный в технике этих доказательств, может опустить эту главу, а соотношение (1.2.3) доказать самостоятельно указанным выше методом, требующим более сильных условий, которые тем не менее выполнены для большинства семейств распределений, встречающихся в приложениях.

Семейства $P_{N,t}$, для которых выполнено соотношение (1.2.3), называются *локально асимптотически нормальными*. Это название связано с тем, что для таких семейств выполнены асимптотически важнейшие свойства, которыми в точном смысле обладает семейство нормальных распределений, зависящих от параметра сдвига. Соотношение (1.2.3) играет ключевую роль в получении результатов о распределениях ранговых критериев при локальных альтернативах и об их асимптотической эффективности. Прежде всего, из него следует, что введенные выше совместные распределения P_N и $P_{N,t}$ вектора наблюдений взаимно контигуальны. Поэтому, как уже отмечалось, аппроксимация ранговой статистики суммой независимых случайных величин, установленная при гипотезе H_0 в главе 3, остается в силе и при альтернативах $P_{N,t}$. Соотношение (1.2.3) лежит в основе так называемой *третьей леммы Ле Кама*, которая, в частности, утверждает, что если сумма независимых слагаемых асимптотически нормальна относительно распределения P_N (т.е. при гипотезе H_0), то она асимптотически нормальна и относительно $P_{N,t}$, причем предельные нормальные распределения отличаются только сдвигом, для которого дается явная формула. Отсюда непосредственно выводится асимптотическая нормальность ранговых статистик при локальных альтернативах.

Знание асимптотических распределений ранговых статистик при альтернативах позволяет не только находить предельную мощность соответствующих критериев, но и сравнивать ранговые критерии между собой и с «параметрическими» критериями по их асимптотической эффективности. Кратко можно сказать, что асимптотическая относительная эффективность (АОЭ) двух критериев измеряется обратным отношением объемов выборок, требуемых тому и другому критерию для достижения одной и той же мощности при альтернативах, приближающихся к гипотезе[8].

Первоначально ранговые критерии рассматривались как простые в вычислительном отношении процедуры для предварительной обработки данных. Предполагалось, что отбрасывание численных значений наблюдений и сохранение только информации об их относительном порядке расположения по величине должно резко снижать эффективность таких критериев. Впоследствии

[8] «Обратное отношение» означает, что АОЭ 1-го критерия относительно 2-го определяется как $e_{1,2} = n_2/n_1$ (а не n_1/n_2), где n_1 и n_2 – упомянутые необходимые объемы выборок, с тем, чтобы при $n_1 > n_2$ было $e_{1,2} < 1$, т.е. эффективность критерия была бы тем меньше, чем больше наблюдений требуется для достижения той же мощности. Точнее, следовало бы говорить, что это АОЭ по Питману, или питмановская АОЭ, поскольку существуют и другие определения относительной эффективности. В настоящем курсе другие понятия эффективности рассматриваться не будут, поэтому это уточнение, как правило, будет опускаться.

выяснилось, что ранговые критерии обладают высокой эффективностью. В частности, для заданного параметрического семейства распределений можно построить асимптотически эффективные ранговые критерии, т.е. критерии, имеющие АОЭ, равную 1, по отношению к наилучшим «параметрическим» критериям для данного семейства. При этом, если принятое параметрическое семейство не соответствует истинному распределению статистических данных, то, как правило, ранговые критерии меньше теряют в эффективности, чем соответствующие «параметрические» критерии. Результаты о вычислении АОЭ ранговых критериев и о построении асимптотически эффективных ранговых критериев будут даны в § 6.2. Там же приводится краткий обзор имеющихся в литературе результатов об асимптотической эффективности ранговых критериев, вывод которых не включен в данный курс лекций.

Заключительная глава – Приложение. Она содержит некоторые сведения из математической статистики и вспомогательные результаты (теоремы о сходимости распределений и интегралов, специального вида центральная предельная теорема и закон больших чисел), используемые в основном тексте.

Глава 2. Ранги и их свойства при гипотезе H_0

§ 2.1. Ранги и порядковые статистики

Мы будем обозначать заглавными буквами случайные величины (векторы) и малыми буквами – неслучайные переменные. Операция упорядочения по величине и введение рангов могут быть определены для произвольного (неслучайного) вектора с несовпадающими компонентами. Рассмотрим вектор $\mathbf{x} = (x_1, \dots, x_N)$, все координаты которого различны. Обозначим

$$\mathbf{x}^{(\cdot)} = (x^{(1)} < \cdots < x^{(N)})$$

вектор, полученный из $\mathbf{x}$ упорядочением координат по возрастанию. Тогда каждой компоненте вектора $\mathbf{x}$ соответствует номер равной ей компоненты вектора $\mathbf{x}^{(\cdot)}$, который и будет называться ее рангом. Ранг компоненты x_i будет обозначаться r_i. Это значит, что

$$x_1 = x^{(r_1)}, \quad x_2 = x^{(r_2)}, \quad \dots, \quad x_N = x^{(r_N)}. \tag{2.1.1}$$

Это равенство и является определением рангов.

Для наглядности напомним спортивную аналогию: перед началом соревнований участникам присваиваются стартовые номера, а по окончании забега регистрируются результаты участников $x_1, \dots, x_N$ в порядке их стартовых номеров. После этого составляется финишный протокол, в котором результаты располагаются по возрастанию, $x^{(1)} < \cdots < x^{(N)}$, и порядковый номер в этом протоколе означает занятое спортсменом место. Обозначим r_i место, занятое i-м участником. Таким образом, результат 1-го участника, x_1, стоит на r_1-м месте в финишном протоколе, т.е. $x_1 = x^{(r_1)}$, и т.д., что и выражается равенством (2.1.1).

Следует помнить, что вектор $\mathbf{x}^{(\cdot)}$ и вектор рангов $\mathbf{r} = (r_1, \dots, r_N)$ являются функциями исходного вектора $\mathbf{x}$, т.е. $\mathbf{x}^{(\cdot)} = \mathbf{x}^{(\cdot)}(\mathbf{x})$, $\mathbf{r} = \mathbf{r}(\mathbf{x})$, что, как правило, не будет указываться явно. Отметим, что вектор $\mathbf{r}$ представляет собой перестановку чисел $1, \dots, N$. Множество всевозможных таких перестановок обозначим $\mathcal{R}$. Как известно, количество элементов этого множества (число перестановок) равно $N!$.

Переупорядочение координат вектора является линейной операцией. Рассмотрим произвольную перестановку $\mathbf{r} = (r_1, \ldots, r_N)$ чисел $1, \ldots, N$. Пользуясь стандартным обозначением $\mathbf{e}_i = (0, \ldots, 1, \ldots, 0)$ для i-го координатного вектора (т.е. вектора (строки) с единственной ненулевой координатой 1 на i-м месте), введем матрицу

$$A_{\mathbf{r}} = \begin{pmatrix} \mathbf{e}_{r_1} \\ \mathbf{e}_{r_2} \\ \ldots \\ \mathbf{e}_{r_N} \end{pmatrix}.$$

(A – квадратная матрица $(N \times N)$, в i-й строке которой стоит 1 в позиции r_i, $i = 1, \ldots, N$, а остальные элементы равны нулю.) Тогда связь между $\mathbf{x}$ и соответствующим ему $\mathbf{x}^{(\,\cdot\,)}$ (записанными как векторы-столбцы) выражается формулой

$$\mathbf{x} = A_{\mathbf{r}(\mathbf{x})}\mathbf{x}^{(\,\cdot\,)}.$$

Поясним это соотношение. Пространство $\mathbb{R}^N$ (за исключением множества точек, у которых какие-либо координаты равны между собой) распадается на $N!$ областей $\mathbb{X}_{\mathbf{r}}^{(\,\cdot\,)}$, отвечающих различным упорядочениям координат по величине. А именно, $\mathbb{X}_{\mathbf{r}}^{(\,\cdot\,)} = \{\mathbf{x}\colon \mathbf{r}(\mathbf{x}) = \mathbf{r}\}$. Обозначим $\mathbb{X}^{(\,\cdot\,)}$ область, соответствующую тождественной перестановке $\mathbf{r} = (1, \ldots, N)$, т.е. $\mathbb{X}^{(\,\cdot\,)} = \{\mathbf{x} = (x_1, \ldots, x_n)\colon x_1 < \cdots < x_n\}$. При упорядочении координат точка $\mathbf{x}$ переходит в точку $\mathbf{x}^{(\,\cdot\,)}$, лежащую в $\mathbb{X}^{(\,\cdot\,)}$. Матрица $A_{\mathbf{r}(\mathbf{x})}$ переводит $\mathbf{x}^{(\,\cdot\,)}$ в $\mathbf{x}$, а область $\mathbb{X}^{(\,\cdot\,)}$ в область $\mathbb{X}_{\mathbf{r}(\mathbf{x})}^{(\,\cdot\,)}$, состоящую из точек, координаты которых упорядочены по величине так же, как координаты точки $\mathbf{x}$. Отметим, что все $N!$ областей $\mathbb{X}_{\mathbf{r}}^{(\,\cdot\,)}$ получаются из области $\mathbb{X}^{(\,\cdot\,)}$ применением к ней преобразований $A_{\mathbf{r}}$, отвечающих всевозможным $\mathbf{r} \in \mathcal{R}$. Предлагаем читателю выписать явно введенные выше матрицы, области и преобразования для $N = 2$ и 3.

Сказанное выше применимо к вектору наблюдений $\mathbf{X} = (X_1, \ldots, X_N)$ и вектору порядковых статистик[1] $\mathbf{X}^{(\,\cdot\,)} = (X^{(1)}, \ldots, X^{(N)})$. Как и раньше, по вектору $\mathbf{X}$ определяется вектор рангов $\mathbf{R} = \mathbf{R}(\mathbf{X})$ и справедлива формула

$$\mathbf{X} = A_{\mathbf{R}}\mathbf{X}^{(\,\cdot\,)}. \tag{2.1.2}$$

Заметим, что матрицы $A_{\mathbf{r}}$ – ортогональные, поэтому обратные к ним матрицы получаются транспонированием, и

$$\mathbf{X}^{(\,\cdot\,)} = A_{\mathbf{R}}^T\mathbf{X}.$$

Установим свойства распределений вектора порядковых статистик $\mathbf{X}^{(\,\cdot\,)}$ и вектора рангов $\mathbf{R}$ в предположении, что выполнена гипотеза H_0, согласно которой $X_1, \ldots, X_N$ – независимые одинаково распределенные случайные величины, имеющие одну и ту же плотность $p(x)$ по мере Лебега. Тогда совместная плотность распределения вектора $\mathbf{X}$ равна

$$p_N(\mathbf{x}) = \prod_{i=1}^{N} p(x_i).$$

Отметим очевидное свойство: $p_N(\mathbf{x})$ симметрична относительно своих аргументов, т.е. не меняется при их перестановках.

Лемма 2.1.1. *При гипотезе H_0 плотность распределения $p_N^{(\,\cdot\,)}(\mathbf{x})$ вектора порядковых статистик $\mathbf{X}^{(\,\cdot\,)}$ равна*

$$p_N^{(\,\cdot\,)}(\mathbf{x}) = N!\,p_N(\mathbf{x}) \ \ \textit{при} \ \ \mathbf{x} \in \mathbb{X}^{(\,\cdot\,)}$$

и $p_N^{(\,\cdot\,)}(\mathbf{x}) = 0$ вне $\mathbb{X}^{(\,\cdot\,)}$.

Доказательство. Последнее утверждение следует из определения $\mathbf{X}^{(\,\cdot\,)}$. Возьмем произвольное борелевское множество $B \subset \mathbb{X}^{(\,\cdot\,)}$. Событие $\mathbf{X}^{(\,\cdot\,)} \in B$ происходит тогда и только тогда, когда $\mathbf{X}$ принадлежит одному из множеств $A_{\mathbf{r}}B$, $\mathbf{r} \in \mathcal{R}$. Поэтому

$$P(\mathbf{X}^{(\,\cdot\,)} \in B) = \int_B p_N^{(\,\cdot\,)}(\mathbf{x})\,d\mathbf{x} = \sum_{\mathbf{r}} \int_{A_{\mathbf{r}}B} p_N(\mathbf{x})\,d\mathbf{x}. \tag{2.1.3}$$

---

[1] Вектор порядковых статистик, т.е. вектор наблюдений, расположенных в порядке возрастания, называют в русскоязычной литературе *вариационным рядом* (этот термин встречается, например, в курсе Б. В. Гнеденко [4]), а сами порядковые статистики называют тогда членами вариационного ряда. Название *порядковая статистика* пришло из англоязычной терминологии (order statistic).

Заметим, что $|\det(A_{\mathbf{r}})| = 1$, $\mathbf{r} \in \mathcal{R}$. В каждом из интегралов под знаком суммы сделаем замену $\mathbf{x} = A_{\mathbf{r}}\mathbf{y}$. Тогда этот интеграл будет равен

$$\int_B p_N(A_{\mathbf{r}}\mathbf{y})\,d\mathbf{y} = \int_B p_N(\mathbf{y})\,d\mathbf{y}$$

в силу симметрии p_N. Следовательно, переобозначая переменную интегрирования, получаем

$$\int_B p_N^{(\,\cdot\,)}(\mathbf{x})\,d\mathbf{x} = N! \int_B p_N(\mathbf{x})\,d\mathbf{x},$$

что доказывает лемму. □

Лемма 2.1.2. *При гипотезе H_0 вектор рангов $\mathbf{R}$ и вектор порядковых статистик $\mathbf{X}^{(\,\cdot\,)}$ стохастически независимы и*

$$P(\mathbf{R} = \mathbf{r}) = \frac{1}{N!}, \quad \mathbf{r} \in \mathcal{R}.$$

Доказательство. Для произвольных $\mathbf{r} \in \mathcal{R}$ и борелевского множества $B \subset \mathbb{X}^{(\,\cdot\,)}$ рассмотрим вероятность

$$P(\mathbf{R} = \mathbf{r}, \mathbf{X}^{(\,\cdot\,)} \in B) = P(\mathbf{X} \in A_{\mathbf{r}}B).$$

Эта вероятность равна одному из слагаемых суммы в (2.1.3), соответствующему выбранному $\mathbf{r}$. Преобразуя его, как в доказательстве леммы 2.1.1, получаем

$$P(\mathbf{R} = \mathbf{r}, \mathbf{X}^{(\,\cdot\,)} \in B) = \int_B p_N(\mathbf{x})\,d\mathbf{x} = \frac{1}{N!}P(\mathbf{X}^{(\,\cdot\,)} \in B).$$

Отсюда следует утверждение леммы. □

Нам далее потребуются два простых следствия, описывающих условные распределения вектора $\mathbf{X}$ при фиксированном $\mathbf{X}^{(\,\cdot\,)}$ и фиксированном $\mathbf{R}$.

Следствие 2.1.1. *Условное распределение вектора наблюдений $\mathbf{X}$ при фиксированном векторе порядковых статистик $\mathbf{X}^{(\,\cdot\,)} = \mathbf{x} = (x_1, \ldots, x_N)$ есть распределение вектора*

$$(x_{R_1}, x_{R_2}, \ldots, x_{R_N}),$$

где $\mathbf{R} = (R_1, \ldots, R_N)$ – случайная перестановка, принимающая любое значение из $\mathcal{R}$ с равными вероятностями $1/N!$.

Доказательство. Для векторов $\mathbf{X}$ и $\mathbf{X}^{(\cdot)}$ справедлив аналог соотношения (2.1.1). Подставляя в него $\mathbf{X}^{(\cdot)} = \mathbf{x}$ и учитывая, что условие на $\mathbf{X}^{(\cdot)}$ не влияет на распределение $\mathbf{R}$ в силу их независимости, получаем следствие. □

Второе следствие будет частным случаем того очевидного факта, что условное распределение вектора $\mathbf{X}$ при фиксированном $\mathbf{R} = \mathbf{r}$ в силу (2.1.2) есть распределение вектора порядковых статистик с соответствующей перестановкой компонент. А именно, рассмотрим распределение одной компоненты X_1 при фиксированном значении ее ранга R_1.

Следствие 2.1.2. *Условное распределение X_1 при условии $R_1 = j$, $j = 1, \ldots, N$, есть распределение j-й порядковой статистики $X^{(j)}$.*

Следствие очевидно. (Если мы знаем, что 1-й участник занял j-е место, то его результат занимает j-ю строчку финишного протокола.)

В лемме 2.1.1 мы получили совместное распределение вектора порядковых статистик, но нам потребуется также распределение отдельной порядковой статистики. Это распределение хорошо известно, мы напомним его для специального случая, когда выборка получена из распределения, равномерного на $[0, 1]$. В таком случае в обозначениях наблюдений, порядковых статистик и соответствующих переменных (принимающих значения в $[0, 1]$) будем использовать буквы U, u вместо X, x, например, вектор наблюдений будет $\mathbf{U} = (U_1, \ldots, U_N)$, вектор порядковых статистик будет $\mathbf{U}^{(\cdot)}$, и т.п. Чтобы указать, что порядковые статистики соответствуют выборке объема N, мы будем снабжать их индексом N, т.е. писать $\mathbf{U}_N^{(\cdot)} = (U_N^{(1)}, \ldots, U_N^{(N)})$.

Известно (см., например, [3], гл. 2, § 1), что $U_N^{(j)}$ имеет плотность распределения

$$f_N^{(j)}(u) = N C_{N-1}^{j-1} u^{j-1} (1-u)^{N-j}, \tag{2.1.4}$$

а математическое ожидание и дисперсия этого распределения равны

$$EU_N^{(j)} = \frac{j}{N+1}, \quad \operatorname{var} U_N^{(j)} = \frac{j(N-j+1)}{(N+1)^2(N+2)}. \tag{2.1.5}$$

Обратим внимание, что $\max_j \operatorname{var} U_N^{(j)} = O(1/N)$ при $N \to \infty$, т.е. при росте N распределения порядковых статистик концентрируются около своих средних значений.

§ 2.2. Свойство свободы от распределения

Напомним, что результаты предыдущего параграфа получены в предположении, что выполнена гипотеза H_0, согласно которой $X_1, \ldots, X_N$ – независимые одинаково распределенные случайные величины, имеющие распределение F с плотностью $p(x)$ по мере Лебега. По лемме 2.1.2 распределение вероятностей статистики

$$S = \sum_{i=1}^{N} c_i a(R_i)$$

(см. (1.1.2)) полностью определяется тем фактом, что ранги $R_1, \ldots, R_N$ образуют случайную перестановку, принимающую любое значение из $\mathcal{R}$ с равными вероятностями $1/N!$. (На самом деле лемма 2.1.2 справедлива и без условия существования плотности: достаточно, чтобы функция распределения $F(x)$ была непрерывной.) Тем самым распределение S не зависит от распределения F исходных случайных величин $X_1, \ldots, X_N$. Статистики с таким свойством называют *свободными от распределения*. Этот термин уточняет, в каком смысле ранговые статистики являются «непараметрическими».

Выше уже говорилось, что «параметрические» критерии не обладают подобным свойством. Например, статистика Стьюдента, имеющая известное распределение Стьюдента, когда наблюдения X_i распределены нормально, будет иметь распределение, отличное от «табличного», при ином распределении F.

Точные распределения наиболее употребительных ранговых статистик достаточно полно табулированы. Кроме того, вычисление их распределений заложено в большинство статистических пакетов программ.

Однако вычисления точных распределений ранговых статистик, проводимые комбинаторными методами, значительно усложняются при больших объемах выборок. Результат следующей главы позволяет пользоваться в таких случаях нормальной аппроксимацией.

Глава 3. Асимптотическая нормальность простых линейных ранговых статистик при гипотезе H_0

§ 3.1. ПЛРС и аппроксимирующая сумма независимых случайных величин

Мы рассматриваем статистики вида (1.1.2), снабжая статистики и входящие в них константы индексом N. Таким образом, мы имеем две последовательности наборов констант

$$(c_{N,1}, \ldots, c_{N,N}), \quad (a_N(1), \ldots, a_N(N))$$

и рассматриваем распределение статистики

$$S_N = \sum_{i=1}^{N} c_{N,i} a_N(R_{N,i}), \tag{3.1.1}$$

где $\mathbf{R}_N = (R_{N,1}, \ldots, R_{N,N})$ – вектор, распределенный при гипотезе H_0 равномерно на $\mathcal{R}_N$. Иначе говоря, S_N – это сумма произведений $c_{N,i}$ на константы $a_N(i)$, расположенные в случайном порядке. Приведем элементарно проверяемые формулы для математического ожидания и дисперсии S_N. Обозначим

$$\begin{gathered} \bar{c}_N = \frac{1}{N} \sum_{i=1}^{N} c_{N,i}, \quad \bar{a}_N = \frac{1}{N} \sum_{i=1}^{N} a_N(i), \\ \sigma_a^2 = \frac{1}{N-1} \sum_{i=1}^{N} (a_N(i) - \bar{a}_N)^2. \end{gathered} \tag{3.1.2}$$

Тогда при гипотезе H_0

$$ES_N = \bar{a}_N \sum_{i=1}^{N} c_{N,i} = N \bar{a}_N \bar{c}_N, \quad \operatorname{var} S_N = \sigma_a^2 \sum_{i=1}^{N} (c_{N,i} - \bar{c}_N)^2. \tag{3.1.3}$$

Целью настоящей главы является доказательство асимптотической нормальности статистик (3.1.1) при гипотезе H_0, т.е. утверждения, что

$$\frac{S_N - ES_N}{\sqrt{\operatorname{var} S_N}} \xrightarrow{d} \mathcal{N}(0,1) \text{ при } N \to \infty.$$

Теорема об асимптотической нормальности статистик вида (3.1.1) известна в вероятностной литературе как комбинаторная предельная теорема и первые доказательства ее появились вне связи с ранговыми критериями. Она доказывалась неоднократно различными методами и при различных условиях на константы $c_{N,i}$ и $a_N(i)$ (одно из доказательств принадлежит автору этих строк). Здесь мы приведем доказательство Я. Гаека [3], предполагающее определенную структуру меток $a_N(i)$ (которая будет описана в ближайших абзацах). Эта структура дает возможность аппроксимировать S_N суммой независимых случайных величин, к которой применима центральная предельная теорема, откуда непосредственно вытекает асимптотическая нормальность S_N при гипотезе H_0. Более того, с помощью этой аппроксимации в дальнейшем (§ 6.2) доказывается асимптотическая нормальность S_N при локальных альтернативах, которая используется для получения асимптотической мощности ранговых критериев.

Суммы, аппроксимирующие S_N, имеют вид

$$T_N = \sum_{i=1}^{N} c_{N,i}\varphi(U_i), \tag{3.1.4}$$

где U_i, $i = 1, \ldots, N$, – равномерно распределенные на $[0,1]$ случайные величины, а φ – некоторая функция[1], определенная на $[0,1]$. Подход Я. Гаека состоит в построении меток $a_N(i)$ так, чтобы они в определенном смысле аппроксимировали функцию φ, и доказательстве того, что S_N с такими метками близка к T_N.

[1]Буква φ употребляется в настоящем курсе для обозначения двух разных функций: с одной стороны, это только что введенная функция, входящая в статистики вида (3.1.4), с другой стороны, $\varphi(x)$ обозначает плотность нормального распределения. Мы сохраняем это двойное значение буквы φ, поскольку то и другое обозначение являются общепринятыми. Как правило, смысл понятен из контекста. Там, где возможно неправильное понимание, будут делаться соответствующие примечания, а в одном пункте, где постоянно встречаются обе функции, плотность нормального распределения обозначается $\psi(x)$.

Перейдем к точным формулировкам. С выборкой $(U_1, \ldots, U_N)$, как и выше, связан вектор рангов $\mathbf{R} = (R_{N,1}, \ldots, R_{N,N})$ и вектор порядковых статистик $\mathbf{U}^{(\cdot)} = (U_N^{(1)}, \ldots, U_N^{(N)})$. Пусть функция $\varphi(u)$, $u \in [0,1]$, удовлетворяет условию

$$E\varphi^2(U_1) = \int_0^1 \varphi^2(u)\,du < \infty. \tag{3.1.5}$$

Определим метки $a_N^{\varphi}(j)$, $j = 1, \ldots, N$, равенством

$$a_N^{\varphi}(j) = E\varphi(U_N^{(j)}), \quad j = 1, \ldots, N. \tag{3.1.6}$$

Поясним это соотношение. Мы хотим аппроксимировать случайную величину $\varphi(U)$ дискретной случайной величиной, принимающей N значений. Если φ – гладкая функция, то это можно сделать, разбив отрезок $[0,1]$ на N равных отрезков и построив ступенчатую функцию, принимающую на этих отрезках постоянные значения, близкие к φ. В качестве таких значений можно взять, например, значения φ в серединах отрезков разбиения; другой способ – взять какое-либо усреднение φ, например, положить

$$a_N(j) = N \int_{\Delta_j} \varphi(u)\,du, \tag{3.1.7}$$

где $\Delta_j = [(j-1)/N, j/N]$. Это выражение можно рассматривать как математическое ожидание $\varphi(U^{[j]})$, где $U^{[j]}$ – случайная величина, распределенная равномерно на Δ_j. Формула (3.1.6) также представляет усреднение функции φ, но по распределению j-й порядковой статистики (2.1.4):

$$a_N^{\varphi}(j) = \int \varphi(u) f_N^{(j)}(u)\,du, \tag{3.1.8}$$

сконцентрированному вблизи точки $j/(N+1)$ (см. формулу (2.1.5) и комментарий к ней). Оказывается, что такое определение меток позволяет обойтись без всяких условий гладкости на φ, как показывает следующая лемма.

Лемма 3.1.1. *При условии* (3.1.5)

$$\lim_{N\to\infty} E[a_N^{\varphi}(R_{N,1}) - \varphi(U_1)]^2 = 0. \tag{3.1.9}$$

Прежде чем доказывать это соотношение, попробуем понять его наглядно. Как мы видели, $a_N^{\varphi}(j)$ можно рассматривать как усреднение функции φ по малой окрестности точки $j/(N+1)$. Предположим, что функция φ достаточно гладкая. Тогда $a_N^{\varphi}(j) \approx \varphi(j/(N+1))$ и формула (3.1.9) говорит нам, что отдельное наблюдение U_1 близко к своему «относительному рангу» $R_{N,1}/(N+1)$. Попробуем пояснить, почему этого следует ожидать, на наглядном примере. Представьте себе, что вы проходите тест (экзамен, конкурс) в составе группы из 100 человек и оценка производится по 100-балльной системе, причем (в отличие от общепринятого) наивысшая оценка – это 1, а самая низкая – это 100. Допустим (хотя это и не вполне реалистично), что результат каждого участника – случайная величина, равномерно распределенная на числах от 1 до 100. И вот вы знаете свой результат, но вам важно знать занятое место, например, вы хотите оказаться в первой двадцатке. Можно ли спрогнозировать занятое место по результату? С известной долей приближения – можно. Если вы набрали 50 баллов, то ясно, что вы где-то в середине и шансов попасть в двадцатку почти нет. А если у вас 20 баллов, то вполне может быть, а если 10 баллов, то шансы очень высокие. Иначе говоря, следует ожидать, что занятое место будет близко к полученной оценке. Чтобы сопоставить это с нашими общими формулами, переведем оценку в проценты, т.е., скажем, 80 баллов будем считать как 0.8. Тогда оценка (приблизительно) равномерно распределена на $[0,1]$ (а точнее – на решетке с шагом 0.01). А переходя от численности 100 к общему N, получаем, что Ваше «относительное место» $R_{N,1}/N$ близко к Вашей оценке U_1. Перейдем теперь к формальному доказательству леммы.

Доказательство леммы. Предположим сначала, что φ имеет ограниченную производную. Тогда и сама φ ограничена, так что мы предполагаем, что $\varphi(u) \leqslant C$, $\varphi'(u) \leqslant C$ для некоторого $C < \infty$. Чтобы оценить математическое ожидание в (3.1.9), оценим сначала условное математическое ожидание того же выражения при фиксированном $R_{N,1} = j$, а потом применим формулу полного математического ожидания, т.е. получим безусловное математическое ожидание как сумму произведений условных математических ожиданий на вероятности условий.

При условии $R_{N,1} = j$ имеем $a_N^{\varphi}(R_{N,1}) = a_N^{\varphi}(j)$, а $\varphi(U_1) = \varphi(U_N^{(j)})$ согласно следствию 2.1.2. Вспоминая (3.1.6) (и меняя местами слагаемые в квадратных скобках), получаем

$$E\big([a_N^{\varphi}(R_{N,1}) - \varphi(U_1)]^2 \mid R_{N,1} = j\big)$$
$$= E[\varphi(U_N^{(j)}) - E\varphi(U_N^{(j)})]^2 = \operatorname{var}\varphi(U_N^{(j)}). \qquad (3.1.10)$$

Обозначим

$$D = \varphi(U_N^{(j)}) - \varphi\left(\frac{j}{N+1}\right).$$

Тогда $\operatorname{var}\varphi(U_N^{(j)}) \leqslant ED^2$. Мы предполагаем, что $\varphi'(u) \leqslant C$, поэтому $|D| \leqslant C|U_N^{(j)} - \frac{j}{N+1}|$ и по формулам (2.1.5) имеем

$$ED^2 \leqslant \frac{C_1}{N},$$

где константа C_1 не зависит от j. Тем самым мы получили равномерную по j оценку для условного математического ожидания (3.1.10), следовательно, применяя формулу полного математического ожидания, мы получим такую же оценку C_1/N для искомого безусловного математического ожидания.

Для завершения доказательства остается освободиться от предположения гладкости φ. Заметим, что по лемме 2.1.2 равенство (3.1.6) можно переписать в виде

$$a_N^{\varphi}(j) = E[\varphi(U_1) \mid R_{N,1} = j]. \qquad (3.1.11)$$

Напомним, что для произвольных случайных величин X и Y

$$E(E(X \mid Y))^2 \leqslant E(E(X^2 \mid Y)) = EX^2. \qquad (3.1.12)$$

Возьмем произвольное $\varepsilon > 0$ и найдем функцию $\tilde\varphi$, имеющую ограниченную производную, так, чтобы

$$\int_0^1 (\varphi - \tilde\varphi)^2\, du < \varepsilon.$$

Например, в качестве $\tilde\varphi$ можно взять частичную сумму ряда Фурье функции φ. По доказанному, соотношение (3.1.9) выполняется для функции $\tilde\varphi$ и меток $a_N^{\tilde\varphi}$. Покажем, что математическое ожидание в этом соотношении мало отличается от математического ожидания, соответствующего функции φ. Для этого установим близость соответствующих слагаемых. Имеем

$$E(\tilde\varphi(U_1) - \varphi(U_1))^2 < \varepsilon$$

по определению $\tilde{\varphi}$. В силу (3.1.12) имеем

$$E(a_N^{\varphi}(R_{N,1}))^2 \leqslant E\varphi^2(U_1) = \int_0^1 \varphi^2\,du. \qquad (3.1.13)$$

По линейности, $a_N^{\varphi} - a_N^{\tilde{\varphi}} = a_N^{\varphi-\tilde{\varphi}}$. Применяя (3.1.13) к $\varphi - \tilde{\varphi}$, получаем

$$E(a_N^{\varphi}(R_{N,1}) - a_N^{\tilde{\varphi}}(R_{N,1}))^2 < \varepsilon.$$

Из этих оценок, применяя неравенство $(a+b)^2 \leqslant 2(a^2+b^2)$, выводим требуемое соотношение для функции φ. $\square$

§ 3.2. Теорема об аппроксимации и асимптотическая нормальность ПЛРС

3.2.1. Вспомогательный результат. Здесь мы приводим лемму, позволяющую устанавливать асимптотическую нормальность возникающих у нас сумм независимых случайных величин.

Рассмотрим последовательность случайных величин следующего вида: пусть $Y_1, Y_2, \ldots$ – независимые одинаково распределенные случайные величины с $EY_1^2 < \infty$ и пусть при каждом N задан набор констант $c_{N,i}$, $i = 1, \ldots, N$. Мы рассматриваем последовательность сумм $Z_N = \sum_{i=1}^N c_{N,i} Y_i$. Предположим, что $c_{N,i}$ удовлетворяют условию:

$$\frac{\max_i c_{N,i}^2}{\sum_{i=1}^N c_{N,i}^2} \to 0 \ \text{ при } \ N \to \infty. \qquad (3.2.1)$$

Легко подсчитать, что

$$EZ_N = N\bar{c}_N EY_1, \quad \operatorname{var} Z_N = \sum_{i=1}^N c_{N,i}^2 \operatorname{var} Y_1,$$

где $\bar{c}_N = N^{-1} \sum_{i=1}^N c_{N,i}$.

Лемма 3.2.1. *При указанных условиях на $Y_1, Y_2, \ldots$ и константы $c_{N,i}$*

$$\frac{Z_N - EZ_N}{\sqrt{\operatorname{var} Z_N}} \xrightarrow{d} \mathcal{N}(0,1).$$

Эта лемма с доказательством, состоящим в проверке условия Линдеберга, приводится как теорема 8.3.1 в § 8.3.1 Приложения.

3.2.2. Теорема об асимптотической нормальности ПЛРС. Мы рассматриваем ПЛРС

$$S_N = \sum_{i=1}^{N} c_{N,i} a_N^{\varphi}(R_{N,i}), \tag{3.2.2}$$

где метки $a_N^{\varphi}(j)$ определены формулой (3.1.6) (или, что то же, (3.1.8)) по некоторой функции $\varphi(u)$, $0 < u < 1$. Основным результатом этого пункта будет теорема 3.2.1 об асимптотической нормальности S_N, выводимая как следствие теоремы 3.2.2 об аппроксимации статистики S_N суммой вида T_N (3.1.4).

Напомним, что функция φ предполагается удовлетворяющей условию

$$E\varphi^2(U_1) = \int_0^1 \varphi^2(u)\,du < \infty \tag{3.2.3}$$

(это условие (3.1.5), повторяемое здесь для удобства читателя). Обозначим

$$\bar{\varphi} = \int_0^1 \varphi(u)\,du.$$

В дополнение к условию (3.1.5) (или (3.2.3)) предположим, что

$$\int_0^1 (\varphi(u) - \bar{\varphi})^2\,du > 0. \tag{3.2.4}$$

(Это условие означает, что φ не вырождается в постоянную.)

Относительно констант $c_{N,i}$ будет предполагаться условие равномерной малости

$$\frac{\max_i (c_{N,i} - \bar{c}_N)^2}{\sum_{i=1}^{N} (c_{N,i} - \bar{c}_N)^2} \to 0 \text{ при } N \to \infty, \tag{3.2.5}$$

где $\bar{c}_N$ определены в (3.1.2). Обозначим

$$\mu_N = \bar{c}_N \sum_{i=1}^{N} a_N^{\varphi}(i), \quad \sigma_N^2 = \sum_{i=1}^{N} (c_{N,i} - \bar{c}_N)^2 \int_0^1 (\varphi(u) - \bar{\varphi})^2\,du. \tag{3.2.6}$$

Теорема 3.2.1. *Пусть выполнены условия* (3.2.3), (3.2.4) *и* (3.2.5). *Тогда S_N асимптотически нормальна с параметрами μ_N, σ_N^2, т.е.*

$$\frac{S_N - \mu_N}{\sigma_N} \xrightarrow{d} \mathcal{N}(0,1). \tag{3.2.7}$$

Замечание 3.2.1. Сравнивая (3.2.6) и формулу для дисперсии S_N (см. (3.1.2) и (3.1.3)), видим, что, вообще говоря, $\sigma_N^2 \neq \operatorname{var} S_N$. Однако $\sigma_N^2/\operatorname{var} S_N \to 1$ и утверждение (3.2.7) верно также с $\sqrt{\operatorname{var} S_N}$ вместо σ_N в знаменателе. Доказательство этого утверждения мы опускаем.

Замечание 3.2.2. В сформулированной теореме 3.2.1 и последующей теореме 3.2.2 рассматриваются ранговые статистики с метками (3.1.6), но используется не конкретный вид этих меток, а свойство (3.1.9). Поэтому эти теоремы остаются в силе для любых других меток, обладающих тем же свойством. Приведем, со ссылкой на книгу [3], два вида «приближенных меток», для которых выполняется (3.1.9). Во-первых, это метки (3.1.7), которые удовлетворяют (3.1.9) при том же условии квадратичной интегрируемости (3.1.5), что и метки (3.1.6) (см. [3], с. 212). Во-вторых, это метки

$$a_N(j) = \varphi\left(\frac{j}{N+1}\right), \quad j = 1, \ldots, N, \tag{3.2.8}$$

фактически использовавшиеся при доказательстве леммы 3.1.1. Эти метки обладают свойством (3.1.9), если функция $\varphi(u)$, $0 < u < 1$, представима в виде конечной суммы монотонных функций, интегрируемых в квадрате (см. [3], с. 211).

Доказательство теоремы 3.2.1. Заметим, что центрированная статистика $S_N - \mu_N$ имеет вид ПЛРС

$$\widetilde{S}_N = S_N - \mu_N = \sum_{i=1}^{N} \tilde{c}_{N,i} a_N^{\varphi}(R_{N,i}) \tag{3.2.9}$$

с константами

$$\tilde{c}_{N,i} = c_{N,i} - \bar{c}_N, \quad \text{такими, что} \quad \sum_{i=1}^{N} \tilde{c}_{N,i} = 0. \tag{3.2.10}$$

Положим

$$T_N = \sum_{i=1}^{N} \tilde{c}_{N,i} \varphi(U_i). \tag{3.2.11}$$

Это сумма независимых (неодинаково распределенных) случайных величин. Нетрудно подсчитать, что

$$ET_N = 0, \quad \operatorname{var} T_N = \sigma_N^2.$$

В силу условия (3.2.5) константы $\tilde{c}_{N,i}$ удовлетворяют условию (3.2.1), при котором выполнена лемма 3.2.1, согласно которой T_N асимптотически нормальна с параметрами $(0, \sigma_N^2)$.

В следующей теореме, представляющей и самостоятельный интерес, будет показано, что

$$E\left(\frac{\widetilde{S}_N - T_N}{\sigma_N}\right)^2 \to 0. \tag{3.2.12}$$

Тогда из полученной асимптотической нормальности T_N будет следовать утверждение теоремы. □

Замечание 3.2.3. Отметим, что ранговые статистики S_N (см. (3.1.1)) и суммы (3.1.4) по-разному «реагируют» на изменение констант $c_{N,i}$. Замена в статистике $S_N = \sum_i c_{N,i} a_N(R_{N,i})$ констант $c_{N,i}$ на $c_{N,i} + c$, $c \in \mathbb{R}$, добавляет к S_N *неслучайное* слагаемое $c \sum_i a_N(i)$ (что и было использовано выше при центрировании S_N). Стоит заметить, что такое преобразование не влияет на свойства S_N как статистики критерия: статистика, отличающаяся неслучайным слагаемым, определяет тот же самый критерий. Статистика же T_N при таком же изменении коэффициентов $c_{N,i}$ приобретает *случайную* добавку $c \sum_i \varphi(U_i)$. Следующая ниже теорема показывает, что T_N может аппроксимировать ранговую статистику S_N только при условии, что сумма входящих в нее коэффициентов $c_{N,i}$ равна нулю.

Теорема 3.2.2. *Пусть выполнены условия* (3.2.3), (3.2.4), (3.2.5) *и* (3.2.10). *Тогда имеет место сходимость* (3.2.12).

Доказательство. Чтобы упростить обозначения, предположим, что изначально $\sum_i c_{N,i} = 0$. Тогда $\bar{c}_N = 0$, $\tilde{c}_{N,i} = c_{N,i}$ и $\widetilde{S}_N = S_N$.

Рассмотрим условное математическое ожидание $(S_N - T_N)^2$ при фиксированном $\mathbf{U}^{(\cdot)} = \mathbf{u}$. В силу независимости $\mathbf{U}^{(\cdot)}$ и $\mathbf{R}$ (лемма 2.1.2) это условие не повлияет на $a_N^{\varphi}(R_{N,i})$, входящие в S_N, а U_i при таком условии перейдут в $u_{R_{N,i}}$ (следствие 2.1.1). Таким образом,

$$E[(S_N - T_N)^2 \mid \mathbf{U}^{(\cdot)} = \mathbf{u}] = E\Big[\sum c_{N,i}(a_N^{\varphi}(R_{N,i}) - \varphi(u_{R_{N,i}}))\Big]^2.$$

Выражение в квадратных скобках представляет собой ПЛРС вида (3.1.1) с $a_N(R_{N,i}) = a_N^{\varphi}(R_{N,i}) - \varphi(u_{R_{N,i}})$. В силу предположения $\bar{c}_N = 0$ ее математическое ожидание равно 0, следовательно, математическое ожидание ее квадрата есть ее дисперсия, которая вычисляется по формулам (3.1.2), (3.1.3). Если в формуле для σ_a^2 в (3.1.2) отбросить вычитаемое $\bar{a}_N$, то выражение увеличится и использование так полученной величины вместо σ_a^2 даст оценку сверху для $\operatorname{var} S_N$. Поэтому

$$E[(S_N - T_N)^2 \mid \mathbf{U}^{(\,\cdot\,)} = \mathbf{u}] \leqslant \frac{1}{N-1} \sum_i c_{N,i}^2 \sum_j [a_N^\varphi(j) - \varphi(u_j)]^2. \tag{3.2.13}$$

Представим вторую сумму в правой части этого неравенства также в виде условного математического ожидания при условии $\mathbf{U}^{(\,\cdot\,)} = \mathbf{u}$. Пользуясь снова следствием 2.1.1, получаем, что разность $a_N^\varphi(R_{N,1}) - \varphi(U_1)$ при условии $\mathbf{U}^{(\,\cdot\,)} = \mathbf{u}$ распределена, как $a_N^\varphi(R_{N,1}) - \varphi(u_{R_{N,1}})$. Поскольку $R_{N,1}$ принимает значения $1, \ldots, N$ с вероятностями $1/N$, а векторы $\mathbf{U}^{(\,\cdot\,)}$ и $\mathbf{R}$ независимы (лемма 2.1.2), имеем

$$E\bigl([a_N^\varphi(R_{N,1}) - \varphi(u_{R_{N,1}})]^2 \mid \mathbf{U}^{(\,\cdot\,)} = \mathbf{u}\bigr) = \frac{1}{N} \sum [a_N^\varphi(j) - \varphi(u_j)]^2.$$

Применяя это равенство (читаемое «справа налево») в (3.2.13), получаем

$$\begin{aligned} &E[(S_N - T_N)^2 \mid \mathbf{U}^{(\,\cdot\,)} = \mathbf{u}] \\ &\qquad \leqslant \frac{N}{N-1} \sum c_{N,i}^2 E\bigl([a_N^\varphi(R_{N,1}) - \varphi(u_{R_{N,1}})]^2 \mid \mathbf{U}^{(\,\cdot\,)} = \mathbf{u}\bigr). \end{aligned}$$

Поскольку это неравенство выполняется для любого $\mathbf{u}$, оно сохранится и при взятии математических ожиданий от обеих частей и будет справедливо для безусловных математических ожиданий, т.е.

$$E(S_N - T_N)^2 \leqslant \frac{N}{N-1} \sum c_{N,i}^2 E[a_N^\varphi(R_{N,1}) - \varphi(u_{R_{N,1}})]^2.$$

Используя формулу (3.2.6) для σ_N^2, получаем

$$E\biggl(\frac{S_N - T_N}{\sigma_N}\biggr)^2 \leqslant \frac{N}{N-1} \frac{1}{\int (\varphi - \bar{\varphi})^2 \, du} E[a_N^\varphi(R_{N,1}) - \varphi(U_1)]^2 \to 0$$

по лемме 3.1.1. Теорема доказана. □

§ 3.3. Примеры асимптотически нормальных ПЛРС

В этом параграфе мы продемонстрируем применение теоремы 3.2.1 на примере нескольких конкретных ранговых статистик вида (3.1.1). Первые три статистики предназначены для задачи сравнения двух выборок и имеют вид

$$S_N = \sum_{i=1}^{n_1} a_N(R_{N,i}). \tag{3.3.1}$$

Затем, в п. 3.3.4, мы рассмотрим ранговые статистики для проверки гипотезы об отсутствии тренда в модели линейной регрессии, или, иначе говоря, о равенстве нулю угла наклона линии регрессии.

Согласно (3.3.1) коэффициенты $c_{N,i}$ в двухвыборочных статистиках имеют вид

$$c_{N,i} = \begin{cases} 1 & \text{при } \ i = 1, \dots, n_1, \\ 0 & \text{при } \ i = n_1 + 1, \dots, N. \end{cases} \tag{3.3.2}$$

Проверим, когда такой набор констант $c_{N,i}$ удовлетворяет условию (3.2.5). Будем считать для определенности, что $n_1 \leqslant n_2$ и обозначим $\lambda = n_1/N$. Тогда $\bar{c}_N = \lambda$ и, используя обозначение (3.2.10),

$$\tilde{c}_{N,i} = c_{N,i} - \bar{c}_N = \begin{cases} 1-\lambda & \text{при } \ i = 1, \dots, n_1, \\ -\lambda & \text{при } \ i = n_1 + 1, \dots, N. \end{cases} \tag{3.3.3}$$

Нетрудно подсчитать, что

$$\sum_1^N \tilde{c}_{N,i}^2 = N\lambda(1-\lambda). \tag{3.3.4}$$

При нашем предположении $n_1 \leqslant n_2$ имеем $\max_i \tilde{c}_{N,i}^2 = (1-\lambda)^2$, следовательно, отношение в левой части (3.2.5) равно

$$\frac{(1-\lambda)^2}{N\lambda(1-\lambda)} = \frac{1-\lambda}{N\lambda}.$$

Учитывая, что $N\lambda = n_1$ и $n_1 = \min(n_1, n_2)$, получаем, что для рассматриваемого набора констант $c_{N,i}$ условие (3.2.5) выполняется тогда и только тогда, когда

$$\min(n_1, n_2) \to \infty \ \text{ при } \ N \to \infty. \tag{3.3.5}$$

3.3.1. Статистика Уилкоксона. Напомним, что эта статистика имеет вид (см. (1.1.1))

$$W_N = \sum_{i=1}^{n_1} R_{N,i}. \tag{3.3.6}$$

Это ПЛРС вида (3.3.1) с $a_N(i) = i$, $i = 1, \dots, N$. Вводя функцию $\varphi(u) = u$, $0 \leqslant u \leqslant 1$, запишем статистику $W_N/(N+1)$ в виде

$$S_N := \frac{W_N}{N+1} = \sum_{i=1}^{N} c_{N,i}\varphi\left(\frac{R_{N,i}}{N+1}\right), \tag{3.3.7}$$

где $c_{N,i}$ – константы (3.3.2). Предположим, что $\min(n_1, n_2) \to \infty$. Как указано выше, при этом условии константы $c_{N,i}$ удовлетворяют условию (3.2.5).

Заметим, что S_N – ПЛРС с метками $j/(N+1)$, которые в силу первого равенства (2.1.5) представляют собой метки $a_N^{\varphi}(j)$, определяемые равенством (3.1.8) для функции $\varphi(u) = u$, совпадающие для этой функции с приближенными метками (3.2.8). Как и выше, предполагая, что $n_1 \leqslant n_2$, обозначим $\lambda = n_1/N$. Тогда $\bar{c}_N = \lambda$ и $\tilde{c}_{N,i} = c_{N,i} - \bar{c}_N$ даются формулой (3.3.3). Обозначим $\mu_N = ES_N$. Тогда

$$S_N - \mu_N = \sum_{i=1}^{N} \tilde{c}_{N,i}\frac{R_{N,i}}{N+1}$$

по теореме (3.2.2) аппроксимируется в смысле соотношения (3.2.12) суммой

$$T_N = \sum_{i=1}^{N} \tilde{c}_{N,i} U_i,$$

где $U_1, \dots, U_N$ – независимые случайные величины, распределенные равномерно на $[0,1]$. Если $\min(n_1, n_2) \to \infty$, то выполнены условия теоремы 3.2.1 и

$$\frac{S_N - \mu_N}{\sigma_N} \xrightarrow{d} \mathcal{N}(0,1).$$

Выпишем явный вид μ_N и σ_N. Для этого воспользуемся формулами (3.2.6). По первой из них μ_N равно произведению $\bar{c}_N = \lambda$ на $\sum_1^N j/(N+1) = \frac{1}{2}N$, т.е. $\mu_N = n_1/2$. По второй из формул (3.2.6) σ_N^2 есть произведение $N\lambda(1-\lambda)$ (см. (3.3.4)) на

$$\int_0^1 (\varphi(u) - \bar{\varphi})^2 \, du = \int_0^1 \left(u - \frac{1}{2}\right)^2 du = \frac{1}{12}.$$

Таким образом,

$$\sigma_N^2 = \frac{n_1 n_2}{12N}.$$

Возвращаясь к исходной статистике $W_N = S_N(N+1)$, нетрудно подсчитать по формулам (3.1.2) и (3.1.3), что

$$EW_N = \frac{1}{2} n_1 (N+1), \quad \operatorname{var} W_N = \frac{1}{12} n_1 n_2 (N+1).$$

Сравнивая с выражениями для μ_N и σ_N^2, видим, что $EW_N = \mu_N(N+1)$ (т.е. $ES_N = \mu_N$), а $\operatorname{var} W_N = \sigma_N^2 N(N+1)$, т.е. $\sigma_N^2 \neq \operatorname{var} S_N = \operatorname{var} W_N/(N+1)^2$, хотя отличие заключается лишь в множителе $(N+1)/N \to 1$.

3.3.2. Медианный критерий. Это двухвыборочный критерий со статистикой

$$S_N = \sum_{i=1}^{n_1} \operatorname{sign}\left(\frac{R_{N,i}}{N+1} - \frac{1}{2}\right). \tag{3.3.8}$$

Эта статистика имеет вид (3.3.1) с метками, получаемыми по формуле (3.2.8) из функции $\varphi(u) = \operatorname{sign}\big(u - \frac{1}{2}\big)$.

Данная статистика представляет собой число элементов 1-й выборки, превосходящих выборочную медиану объединенной выборки. Ее большие значения свидетельствуют о том, что элементы 1-й выборки группируются в верхней половине вариационного ряда и тем самым имеют тенденцию принимать бóльшие значения по сравнению с элементами 2-й выборки. В книге [3] эта статистика приводится на с. 108–109 вместе с некоторыми ее модификациями (отличающимися учетом слагаемого, соответствующего $R_{N,i} = (N+1)/2$ при нечетном N). Как и статистика Уилкоксона, статистика (3.3.8) асимптотически нормальна, если $\min(n_1, n_2) \to \infty$. Предлагаем читателю по аналогии с предыдущим пунктом самостоятельно выписать аппроксимирующую сумму независимых случайных величин и найти центрирующие и нормирующие константы μ_N, σ_N.

3.3.3. Критерий нормальных меток. Этот критерий основан на статистике

$$S_N = \sum_{i=1}^{n_1} a_N^{\varphi}(R_{N,i}), \tag{3.3.9}$$

имеющей вид (3.3.1) с метками $a_N = a_N^{\varphi}$, определяемыми по формуле (3.1.6) (см. также (3.3.10) ниже), в которой функция φ имеет в данном случае вид $\varphi(u) = \Phi^{-1}(u)$, $0 < u < 1$. Здесь $\Phi^{-1}(u)$ – функция, обратная к функции стандартного нормального распределения. Этот критерий называют также критерием Фишера–Иэйтса–Терри–Гёфдинга по именам статистиков, впервые предложивших и изучавших его. Ссылки на посвященные ему работы и таблицы можно найти, в частности, в книгах [3] и [13].

Известно, что если Z – случайная величина, имеющая непрерывную функцию распределения $F(x)$, то случайная величина $F(Z)$ распределена равномерно на $[0,1]$. И обратно, если U – случайная величина, равномерно распределенная на $[0,1]$, то $Z = F^{-1}(U)$ имеет функцию распределения $F(x)$[2]. Как следствие, если $U_N^{(1)} < \cdots < U_N^{(N)}$ – порядковые статистики выборки объема N из независимых равномерных на $[0,1]$ случайных величин, то $\{Z_N^{(j)} = F^{-1}(U_N^{(j)})\}_{j=1}^{N}$ – порядковые статистики выборки из независимых случайных величин с функцией распределения $F(x)$.

Напомним формулу (3.1.6), определяющую метки $a_N^{\varphi}(j)$ по функции φ:

$$a_N^{\varphi}(j) = E\varphi(U_N^{(j)}), \quad j = 1, \ldots, N. \tag{3.3.10}$$

Применяя приведенное выше рассуждение с $F(x) = \Phi(x)$ и $\varphi(u) = \Phi^{-1}(u)$, получаем[3], что в данном случае $Z = \Phi^{-1}(U) \overset{d}{=} \mathcal{N}(0,1)$ и $\varphi(U_N^{(j)}) = Z_N^{(j)}$, $j = 1, \ldots, N$, следовательно

$$a_N^{\varphi}(j) = E(Z_N^{(j)}), \quad j = 1, \ldots, N,$$

[2] В случае, если $F(x)$ – не строго монотонная функция (имеющая горизонтальные участки), определение обратной функции требует уточнения. Поскольку данное рассуждение применяется к нормальному распределению, будем предполагать F строго монотонной.

[3] На всякий случай отметим, что в настоящем параграфе $\varphi(u)$ всюду обозначает функцию, порождающую метки, и нигде не используется плотность $\varphi(x)$ нормального распределения.

т.е. $a_N^\varphi(j)$ есть математическое ожидание j-й порядковой статистики выборки из нормального распределения.

Это соотношение можно использовать для того, чтобы подсчитать значение $\overline{a_N^\varphi}$, фигурирующее в формуле для ES_N. Имеем (см. (3.1.2))

$$N\overline{a_N^\varphi} = \sum_1^N a_N^\varphi(j) = \sum_1^N E(Z_N^{(j)}) = E\biggl(\sum_1^N Z_N^{(j)}\biggr) = E\biggl(\sum_1^N Z_j\biggr) = 0$$

(очевидно, что сумма порядковых статистик и сумма элементов выборки – это сумма одних и тех же слагаемых, записанных в разном порядке). Таким образом, $\overline{a_N^\varphi} = 0$ и $ES_N = 0$. К сожалению, сумма квадратов меток $a_N^\varphi(j)$, возникающая при вычислении $\operatorname{var} S_N$, не имеет замкнутого явного выражения. Однако нетрудно подсчитать параметры μ_N, σ_N (см. (3.2.6)) асимптотически нормального распределения S_N.

Последняя цепочка равенств показывает, что $\mu_N = 0$. Согласно (3.2.6) для вычисления σ_N^2 нам нужно найти $\sum_1^N (c_{N,i} - \bar{c}_N)^2$ и $\int_0^1 (\varphi(u) - \bar{\varphi})^2\,du$. По формуле (3.3.4)

$$\sum_1^N (c_{N,i} - \bar{c}_N)^2 = N\lambda(1-\lambda) = \frac{n_1 n_2}{N}.$$

Далее, используя замену переменных $u = \Phi(x)$, получаем

$$\bar{\varphi} = \int_0^1 \varphi(u)\,du = \int_0^1 \Phi^{-1}(u)\,du = \int x\,d\Phi(x) = 0$$

и

$$\int_0^1 (\varphi(u) - \bar{\varphi})^2\,du = \int_0^1 [\Phi^{-1}(u)]^2\,du = \int x^2\,d\Phi(x) = 1. \qquad (3.3.11)$$

Таким образом, $\sigma_N^2 = n_1 n_2/N$, и при условии $\min(n_1, n_2) \to \infty$ распределение $S_N/\sqrt{n_1 n_2/N}$ сходится к стандартному нормальному распределению $\mathcal{N}(0,1)$.

Приведем вид аппроксимирующей суммы T_N в данном примере. Напомним, что (см. (3.2.11))

$$T_N = \sum_{i=1}^N \tilde{c}_{N,i}\varphi(U_i). \qquad (3.3.12)$$

При этом $\tilde{c}_{N,i} = c_{N,i} - \bar{c}_N$ выражаются формулой (3.3.3), а в силу проведенных выше рассуждений $Z_i := \varphi(U_i) = \Phi^{-1}(U_i) \overset{d}{=} \mathcal{N}(0,1)$, так что, учитывая, что в (3.3.3) $\lambda = n_1/N$ и $1-\lambda = n_2/N$, получаем

$$T_N = \frac{n_1 n_2}{N}(\overline{Z}_1 - \overline{Z}_2), \quad \overline{Z}_1 = \frac{1}{n_1}\sum_{i=1}^{n_1} Z_i, \quad \overline{Z}_2 = \frac{1}{n_2}\sum_{i=n_1+1}^{N} Z_i,$$

т.е. с точностью до обозначений и неслучайных множителей T_N совпадает со статистикой, полученной в п. 8.1.4 (асимптотически эквивалентной t-статистике Стьюдента), которая, как показано в п. 8.1.4, определяет равномерно наиболее мощный критерий в задаче сравнения двух нормальных выборок, в которых дисперсии наблюдений предполагаются равными и известными.

Отметим еще родственный критерий Ван дер Вардена со статистикой

$$S_N = \sum_{i=1}^{n_1} \Phi^{-1}\left(\frac{R_{N,i}}{N+1}\right),$$

получаемой применением приближенных меток (3.2.8). Он обладает теми же асимптотическими свойствами, что и критерий нормальных меток.

3.3.4. ПЛРС для регрессионных альтернатив. Здесь мы укажем общий вид ранговых статистик, предназначенных для задачи регрессии, рассмотренной в п. 8.1.5 Приложения. А именно, мы предполагаем, что наблюдения $X_1, \ldots, X_N$ имеют вид

$$X_i = \theta + \eta c_{N,i} + e_i, \quad i = 1, \ldots, N, \tag{3.3.13}$$

где $e_1, \ldots, e_N$ – независимые одинаково распределенные «ошибки», которые (в отличие от п. 8.1.5) могут иметь произвольную плотность распределения, а $c_{N,1}, \ldots, c_{N,N}$, как всегда, – заданные константы, играющие в данном случае роль «независимых переменных» («регрессоров»). Проверяется гипотеза $H_0\colon \eta = 0$ против альтернативы $H_1\colon \eta > 0$. В этой задаче применяются ранговые статистики вида (3.1.1), т.е.

$$S_N = \sum_{1}^{N} c_{N,i} a_N(R_{N,i}),$$

где $c_{N,i}$ – «регрессоры», входящие в (3.3.13), а в качестве $a_N(R_{N,i})$ могут быть выбраны, в частности, метки, использованные в двухвыборочных статистиках предыдущих пунктов. Предоставляем читателю, ограничиваясь этими тремя примерами, проследить, аналогично двухвыборочному случаю, что

$$S_N - \mu_N = \sum_1^N \tilde{c}_{N,i} a_N(R_{N,i})$$

аппроксимируются суммами (ср. (3.2.11))

$$T_N = \sum_{i=1}^N \tilde{c}_{N,i} \varphi(U_i), \tag{3.3.14}$$

где соответствие между метками a_N и функцией φ – то же, что и в предыдущих пунктах.

Отметим простой частный случай, когда $c_{N,i}$ образуют арифметическую прогрессию (например, они представляют собой следующие с равными промежутками моменты времени проведения измерений), а метки – те же, как в критерии Уилкоксона. Тогда статистика S_N приводится к виду (1.1.3), упоминавшемуся в конце § 1.1:

$$S_N = \sum_{i=1}^N i R_{N,i}. \tag{3.3.15}$$

Глава 4. Контигуальные последовательности распределений

Эта и следующая главы посвящены одному из важнейших приемов получения предельных распределений статистик критериев при альтернативах. Содержание этих глав носит общий характер и не ограничено применением к ранговым критериям. (Ранговые критерии будут рассматриваться в главе 6.) В основе излагаемых здесь методов лежат понятия, введенные Л. Ле Камом [24], и результаты этой работы.

§ 4.1. Предварительные соображения

В п. 8.1.3 Приложения рассмотрен пример построения НМ критерия в задаче проверки гипотезы о параметре сдвига нормального распределения при известном σ. Приведем некоторые результаты этого раздела.

Рассматривается задача проверки гипотезы $H_0\colon \mu = \mu_0$ против альтернативы $H_1\colon \mu > \mu_0$ по выборке $X_1, \ldots, X_N$ независимых случайных величин, имеющих нормальное распределение $\mathcal{N}(\mu, \sigma^2)$, где σ считается известным. С помощью леммы Неймана–Пирсона (см. п. 8.1.2) строится равномерно наиболее мощный (РНМ) критерий, отвергающий H_0, когда

$$S_N > z_\alpha, \tag{4.1.1}$$

где $\alpha > 0$ – заданный уровень значимости, z_α – квантиль уровня $1-\alpha$ стандартного нормального распределения, т.е. $\Phi(z_\alpha) = 1-\alpha$, а S_N – сумма наблюдений, центрированная и нормированная так, что при гипотезе H_0 она имеет распределение $\mathcal{N}(0,1)$ (обычно в данной задаче применяется среднее арифметическое наблюдений $\overline{X}$; нормировка, стандартизирующая распределение $\overline{X}$, и дает S_N). Поскольку здесь мы рассматриваем последовательность статистических задач, индексированную объемом выборки N, мы будем снабжать этим индексом объекты, фигурирующие в п. 8.1.3, такие как L, Λ, S, и т.д. В п. 8.1.3 введен нормированный параметр t, который будет служить основной мерой отклонения альтернативного значения параметра от μ_0. В отличие от обозначений п. 8.1.3, будем обозначать $\mu_N = \mu_0 + t/\sqrt{N}$. Для мощности НМ критерия против такой альтернативы в п. 8.1.3 получена формула

$$\beta_N(t) = \Phi(t/\sigma - z_\alpha), \tag{4.1.2}$$

где $\Phi(x)$ – ф.р. стандартного нормального закона.

В рассмотренном примере статистика S_N имеет в точности нормальное распределение как при H_0, так и при $\mu \in H_1$. Благодаря этому удается построить критерий, имеющий точный уровень значимости α, и найти точное выражение для его мощности. Однако для ранговых статистик (как и для многих других) нормальность распределения имеет место лишь в пределе при $N \to \infty$. Доказанная в п. 3.2.2 асимптотическая нормальность ранговых статистик при гипотезе H_0 и доказываемая в дальнейшем их асимптотическая нормальность при локальных альтернативах позволят получить формулу для мощности, аналогичную (4.1.2), как предельную при $N \to \infty$. А именно, мощность асимптотически наиболее мощного критерия будет выражаться формулой

$$\beta_N(t) = \Phi(t\sqrt{I} - z_\alpha) + o(1) \ \text{ при } \ N \to \infty \tag{4.1.3}$$

(величина I будет определена ниже).

Вообще сходимость *функций* требует определенных оговорок относительно того, в каком смысле ее следует понимать. Чтобы избежать осложнений такого рода, мы будем задаваться последовательностями $\{t_N\}$ и рассматривать соотношение (4.1.3) для соответствующих последовательностей аргумента, т.е.

$$\beta_N(t_N) = \Phi(t_N\sqrt{I} - z_\alpha) + o(1) \ \text{ при } \ N \to \infty. \tag{4.1.4}$$

Если $t_N \to \infty$, то это соотношение означает, что мощность стремится к 1, и присутствие функции нормального распределения в правой части становится бессодержательным, поскольку правую часть можно записать просто как $1 + o(1)$. Свойство сходимости мощности критерия к 1 при какой-либо последовательности альтернатив называется *состоятельностью* критерия при данных альтернативах[1]. Нашей целью является вычисление асимптотической мощности ранговых критериев и на основе этого – сравнение их между собой и с «параметрическими» критериями по их асимптотической эффективности. Точное определение этого понятия будет дано в свое время, пока же скажем, что относительная эффективность двух критериев измеряется обратным отношением объемов выборок, требуемых для достижения ими одинаковой мощности. Приведем пример, показывающий, что состоятельными могут быть и неэффективные критерии и что наиболее информативным является поведение функции мощности в области ее промежуточных значений, где она не приближается к 1.

В задаче проверки гипотез о среднем значении нормального распределения рассмотрим критерий, использующий только половину имеющихся наблюдений, а именно, критерий, основанный на $\overline{X}' = N_1^{-1} \sum_1^{N_1} X_i$, где $N_1/N \to 1/2$ при $N \to \infty$. Предоставляем читателю убедиться, что этот критерий состоятелен при любой последовательности альтернатив с $t_N \to \infty$ (ср. аргумент в левой части (4.1.4)), а также выписать его мощность $\beta'_N(\mu)$ и показать, что $\beta'_{2N}(\mu) \to \beta_N(\mu)$ при каждом фиксированном t. Последнее означает, что его асимптотическая относительная эффективность по отношению к критерию, основанному на $\overline{X}$, равна $1/2$ (что, впрочем, достаточно очевидно из его определения).

Поэтому мы будем рассматривать последовательности альтернатив, при которых мощность отделена от 1. В данном примере такие альтернативы описываются условием ограниченности последовательности $\{t_N\}$. Однако в связи с ранговыми критериями, применяемыми для проверки гипотезы одинаковой распределенности наблюдений, альтернативы, описывающие разнораспределенность, могут задаваться большим числом параметров (как в случае альтернатив тренда, упомянутых в § 1.1) и требуемые условия на вектор параметров отнюдь не очевидны. Поэтому мы

[1] В литературе имеются подходы, при которых критерий характеризуется скоростью убывания вероятности ошибки 2-го рода (т.е. стремления мощности к 1) при фиксированной альтернативе или скоростью убывания вероятности ошибки 1-го рода, когда при фиксированной альтернативе фиксировано значение мощности. Эти подходы используют методы теории больших уклонений. Мы не затрагиваем здесь эти подходы и изучаем асимптотику функции мощности на основе результатов типа центральной предельной теоремы.

вернемся к задаче о среднем значении нормального распределения и выделим условие непосредственно на последовательности распределений на выборочном пространстве, обеспечивающее требуемое свойство.

Рассмотрим равенство (4.1.2) при некотором фиксированном $t > 0$. До сих пор $\alpha > 0$ было фиксированным числом. Будем теперь писать $\beta_N(t \mid \alpha)$ вместо $\beta_N(t)$, возьмем какую-нибудь последовательность $\alpha_N \downarrow 0$ и посмотрим, как ведет себя мощность $\beta_N(t \mid \alpha_N)$. По формуле (4.1.2) $\beta_N(t \mid \alpha_N) = \Phi(t/\sigma - z_{\alpha_N})$. По определению, $z_{\alpha_N} = \Phi^{-1}(1 - \alpha_N)$ и таким образом, $z_{\alpha_N} \to \infty$ при $\alpha_N \downarrow 0$. Следовательно, $\beta_N(t \mid \alpha_N) \to 0$ при $\alpha_N \downarrow 0$.

С другой стороны, нетрудно убедиться, что для любой последовательности $t_N \to \infty$ можно подобрать последовательность $\alpha_N \to 0$ так, что при этом $\beta_N(t_N \mid \alpha_N) \to 1$. Таким образом, свойство, характеризующее альтернативы, при которых мощность остается отделенной от 1, состоит в том, что при $\alpha \to 0$ мощность при таких альтернативах также стремится к нулю.

Чтобы сформулировать искомое свойство без привлечения критериев и их характеристик, проанализируем еще раз ту же задачу о нормальном распределении. Вспомним, что критерий (4.1.1) является равномерно наиболее мощным (РНМ) в задаче проверки гипотезы $H_0 \colon \mu = \mu_0$ против альтернативы $H_1 \colon \mu > \mu_0$. Зафиксируем $t > 0$ и будем рассматривать последовательность альтернатив $\mu_N = \mu_0 + t/\sqrt{N}$. То, что критерий (4.1.1) является РНМ, означает, что если мы возьмем произвольное борелевское множество $A \subset \mathbb{R}^N$ такое, что $P_{\mu_0}(\mathbf{X} \in A) = \alpha$ (или $\leqslant \alpha$), то $P_{\mu_N}(\mathbf{X} \in A) \leqslant \beta_N(t)$ (см. доказательство леммы Неймана–Пирсона в п. 8.1.2). Учитывая приведенные выше свойства β_N при $\alpha \to 0$, получаем следующую характеризацию «локальных» альтернатив $\{\mu_N\}$, при которых мощность остается отделенной от 1: для любой последовательности борелевских множеств $A_N \subset \mathbb{R}^N$ такой, что $P_{N,\mu_0}(A_N) \to 0$, будет также $P_{N,\mu_N}(A_N) \to 0$.

Это свойство, выведенное нами на частном примере нормального семейства, будет принято в следующем параграфе за определение важного свойства последовательностей пар распределений, которые будут рассматриваться в качестве гипотезы и альтернативы в последующей асимптотической теории.

§ 4.2. Условия контигуальности распределений

Рассмотрим последовательность измеримых пространств $(\mathbb{X}_N, \mathcal{A}_N)$, $N = 1, 2, \dots$, на каждом из которых заданы две вероятностные меры P_N и Q_N.

В дальнейшем, при применении результатов настоящей главы к ранговым критериям, в качестве $(\mathbb{X}_N, \mathcal{A}_N)$ будет рассматриваться $(\mathbb{R}^N, \mathcal{B}(\mathbb{R}^N))$, т.е. пространство значений вектора наблюдений $\mathbf{X}_N = (X_1, \dots, X_N)$ с борелевской σ-алгеброй, а в роли P_N и Q_N будут совместные распределения этого вектора при гипотезе H_0 и выбранной для данного N альтернативе.

Определение 4.2.1. Последовательность $\{Q_N\}$ называется *контигуальной* по отношению к $\{P_N\}$ (обозначается $Q_N \lhd P_N$), если $Q_N(A_N) \to 0$ для любой последовательности множеств $A_N \in \mathcal{A}_N$ таких, что $P_N(A_N) \to 0$ при $N \to \infty$.

Это определение можно эквивалентным образом переформулировать на языке ε–δ.

Определение 4.2.2. $Q_N \lhd P_N$, если для любого $\varepsilon > 0$ найдется $\delta > 0$ такое, что для любой последовательности $A_N \in \mathcal{A}_N$ такой, что $P_N(A_N) < \delta$, имеет место $Q_N(A_N) < \varepsilon$ при всех достаточно больших N.

Определение 4.2.3. Если $Q_N \lhd P_N$ и $P_N \lhd Q_N$, то последовательности $\{P_N\}$ и $\{Q_N\}$ называют взаимно контигуальными (или просто контигуальными) и обозначают $P_N \lhd\rhd Q_N$.

Будем предполагать, что меры P_N и Q_N имеют плотности p_N и q_N относительно некоторой σ-конечной меры ν_N на $\mathbb{X}_N$, т.е. для любого $A \in \mathcal{A}_N$

$$P_N(A) = \int_A p_N(\mathbf{x})\, d\nu_N, \quad Q_N(A) = \int_A q_N(\mathbf{x})\, d\nu_N. \tag{4.2.1}$$

В настоящем параграфе будет доказана теорема, дающая необходимые и достаточные условия контигуальности последовательностей $\{P_N\}$ и $\{Q_N\}$. Предварительно в п. 4.2.1 эта теорема доказывается при упрощающем дополнительном предположении. Общий случай рассматривается в следующем п. 4.2.2.

4.2.1. Случай, когда плотности p_N и q_N имеют общий носитель. Носитель плотности – это множество тех точек $\mathbf{x}$, где она положительна. Дополнительное предположение, о котором сказано выше, состоит в том, что

$$B_N := \{\mathbf{x}\colon p_N(\mathbf{x}) > 0\} = \{\mathbf{x}\colon q_N(\mathbf{x}) > 0\}. \tag{4.2.2}$$

Напомним, что мера Q_N *абсолютно непрерывна* относительно P_N (обозначается $Q_N \ll P_N$), если для всякого множества $A \in \mathcal{A}_N$ такого, что $P_N(A) = 0$, имеет место $Q_N(A) = 0$. При условии (4.2.2) меры P_N и Q_N *взаимно абсолютно непрерывны* (или *эквивалентны*) (обозначается $P_N \sim Q_N$), т.е. $P_N \ll Q_N$ и $Q_N \ll P_N$ (проверьте самостоятельно). Напомним еще, что всякая функция, отличающаяся, скажем, от p_N на множестве ν_N-меры нуль, также является плотностью распределения P_N (это видно из (4.2.1): такое отличие не меняет меры $P_N(A)$). Поэтому, если выполнено (4.2.1), то P_N (Q_N) имеет бесконечно много вариантов плотности p_N (q_N), отличающихся один от другого на множестве ν_N-меры нуль. Если $P_N \sim Q_N$, то плотности p_N и q_N можно выбрать так, чтобы выполнялось (4.2.2) (проверьте самостоятельно). В этом смысле условие (4.2.2) равносильно условию $P_N \sim Q_N$ и мы иногда будем говорить о (4.2.2) как о предположении, что P_N и Q_N эквивалентны (взаимно абсолютно непрерывны).

Таким образом, $P_N(B_N) = Q_N(B_N) = 1$, а множество $B_N^c = \{\mathbf{x}\colon p_N(\mathbf{x}) = 0\} = \mathbb{X}_N \setminus B_N$ имеет меру нуль как относительно P_N, так и относительно Q_N. Поэтому мы можем отбросить множество B_N^c и считать, что $\mathbb{X}_N = B_N$, т.е. что $p_N > 0$ и $q_N > 0$ на всем пространстве $\mathbb{X}_N$. (Например, если распределения P_N и Q_N отвечают выборкам неотрицательных случайных величин, то в качестве выборочного пространства естественно рассматривать не все $\mathbb{R}^N$, а множество точек $\mathbb{R}_+^N$ с неотрицательными координатами.)

Чтобы сформулировать теорему, дающую необходимые и достаточные условия контигуальности $Q_N \lhd P_N$, нам потребуется ввести некоторые обозначения. Для упрощения записи аргумент $\mathbf{x}$ у функций p_N и q_N и связанных с ними функций того же аргумента будем, как правило, опускать.

Введем отношение правдоподобия (ОП) и логарифм ОП

$$L_N = \frac{q_N}{p_N}, \quad \Lambda_N = \log L_N. \tag{4.2.3}$$

(L_N и Λ_N определены на всем пространстве $\mathbb{X}_N$ в силу соглашения, что $\mathbb{X}_N = B_N$.) Отметим равенство, вытекающее непосредственно из этого определения:

$$E(e^{\Lambda_N} \mid P_N) = E(L_N \mid P_N) = E\left(\frac{q_N}{p_N} \mid P_N\right)$$
$$= \int \frac{q_N}{p_N} p_N \, d\nu_N = \int q_N \, d\nu_N = 1. \qquad (4.2.4)$$

Отметим еще простое следствие этого равенства[2]:

$$\text{Последовательность } \{\mathcal{L}[L_N \mid P_N]\} \text{ плотна.} \qquad (4.2.5)$$

Действительно, согласно определению 8.2.4 (см. п. 8.2.3) и в силу положительности L_N, чтобы доказать (4.2.5), нужно проверить, что для любого $\varepsilon > 0$ найдется $a > 0$ такое, что $P_N(L_N > a) < \varepsilon$. Но аналогично неравенству Чебышева из (4.2.4) имеем $P_N(L_N > a) < 1/a$, и достаточно взять любое число $a > 1/\varepsilon$.

Следствие 4.2.1. *Согласно теореме Прохорова (теорема 8.2.4) из (4.2.5) следует, что последовательность функций распределения $F_N(t) = P_N(L_N < t)$ относительно компактна. Это значит, что для любой подпоследовательности $\{m\} \subset \{N\}$ найдется подпоследовательность $\{m'\} \subset \{m\}$ и ф.р. F такие, что $F_{m'} \xrightarrow{w} F$.*

Напомним еще важное свойство математического ожидания, которое будет постоянно использоваться в дальнейшем. Пусть $T = T(\mathbf{x})$ – случайная величина, определенная на вероятностном пространстве $(\mathbb{X}, \mathcal{A}, P)$. Тогда, по определению, ее математическое ожидание – это интеграл $ET = \int_{\mathbb{X}} T(\mathbf{x}) \, dP$. Свойство, о котором идет речь, состоит в том, что это математическое ожидание может быть вычислено интегрированием по ф.р. $F(x) = P(T < x)$, а именно, $ET = \int_{\mathbb{R}} x \, dF(x)$. Точно так же для функции от T, скажем, $g(T)$, имеет место равенство

$$Eg(T) = \int_{\mathbb{X}} g(T(\mathbf{x})) \, dP = \int_{\mathbb{R}} g(x) \, dF(x). \qquad (4.2.6)$$

(Действительно, работая с какой-либо статистикой T, мы в первую очередь пытаемся найти ее ф.р. $F(x)$ и плотность $p(x)$ (если

[2] В этом параграфе используются результаты о сходимости распределений и интегралов, приведенные в пп. 8.2.1, 8.2.3 и 8.2.4 Приложения.

она существует), чтобы вычислять ее характеристики, например, моменты ET^m, как интегралы по числовой прямой $ET^m = \int_{\mathbb{R}} x^m\, dF(x) = \int_{\mathbb{R}} x^m p(x)\, dx$, а не по (многомерному) выборочному пространству.)

В силу этого свойства равенство (4.2.4) может быть записано в терминах ф.р. $F_N(t) = P_N(L_N < t)$ как

$$\int_0^\infty t\, dF_N(t) = 1 \tag{4.2.7}$$

(мы указываем область интегрирования $[0,\infty)$, чтобы подчеркнуть, что, поскольку ОП $L_N > 0$, его распределение сосредоточено на положительной полуоси).

Теперь мы переходим к формулировке и доказательству теоремы, дающей необходимые и достаточные условия контигуальности.

Теорема 4.2.1. *$Q_N \lhd P_N$ тогда и только тогда, когда выполнено любое из следующих четырех условий:*

1. *Для произвольной последовательности случайных величин T_N из сходимости $T_N \to 0$ по вероятности P_N следует сходимость $T_N \to 0$ по вероятности Q_N.*
2. *Для произвольной последовательности случайных величин T_N из того, что последовательность $\{\mathcal{L}[T_N \mid P_N]\}$ плотна, следует, что и последовательность $\{\mathcal{L}[T_N \mid Q_N]\}$ плотна (см. определение 8.2.4).*
3. *Последовательность $\{\mathcal{L}[\Lambda_N \mid Q_N]\}$ плотна.*
4. *Для любой ф.р. F, являющейся пределом некоторой подпоследовательности функций распределения F_N (см. следствие 4.2.1), выполняется равенство $\int t\, dF = 1$.*

Доказательство. Предлагаем читателю самостоятельно доказать эквивалентность условий 1 и 2 условию контигуальности $Q_N \lhd P_N$. Для этого нужно всего лишь привлечь определения соответствующих понятий.

Доказательство эквивалентности остальных условий проведем по следующей схеме: $3 \Rightarrow [Q_N \lhd P_N] \Rightarrow 4 \Rightarrow 3$.

Импликация $3 \Rightarrow [Q_N \lhd P_N]$. Согласно определению 4.2.2 нам надо доказать, что для любого $\varepsilon > 0$ найдется $\delta > 0$ такое, что $P_N(A_N) < \delta \Rightarrow Q_N(A_N) < \varepsilon$. По условию 3 можно выбрать $a > 0$ так, чтобы $Q_N(S_N) < \varepsilon/2$, где $S_N = \{\mathbf{x}\colon \Lambda_N(\mathbf{x}) > a\}$. Для произвольного множества $A_N \in \mathcal{A}$ имеем

$$Q_N(A_N) = Q_N(A_N \cap S_N) + Q_N(A_N \cap S_N^c),$$

где $S_N^c = \mathbb{R} \setminus S_N$. Очевидно, $Q_N(A_N \cap S_N) \leqslant Q_N(S_N) < \varepsilon/2$. Оценим второе слагаемое. Если $\mathbf{x} \in S_N^c$, то $\Lambda_N(\mathbf{x}) \leqslant a$ и $q_N/p_N = \log \Lambda_N \leqslant \log a$. Поэтому

$$\begin{aligned} Q_N(A_N \cap S_N^c) &= \int_{A_N \cap S_N^c} q_N(\mathbf{x})\, d\nu_N = \int_{A_N \cap S_N^c} \frac{q_N(\mathbf{x})}{p_N(\mathbf{x})} p_N(\mathbf{x})\, d\nu_N \\ &\leqslant \log a \int_{A_N} p_N(\mathbf{x})\, d\nu_N = P_N(A_N) \log a. \end{aligned}$$

Положим $\delta = \varepsilon/(2\log a)$. Теперь, если $P_N(A_N) < \delta$, то $Q_N(A_N \cap S_N^c) < \varepsilon/2$. Из полученных неравенств следует $Q_N(A_N) < \varepsilon$, что и требовалось доказать.

Импликация $[Q_N \lhd P_N] \Rightarrow 4$. Согласно (4.2.4) $\int t\, dF_N = EL_N = 1$. Покажем, что функция $g(t) = t$ равномерно интегрируема по $\{F_N\}$. Тогда искомое утверждение будет следовать из п. 2 теоремы 8.2.5. Возьмем произвольное $\varepsilon > 0$. Тогда по условию контигуальности существуют $\delta > 0$ и N_1 такие, что $P_N(A) < \delta \Rightarrow Q_N(A) < \varepsilon$ для произвольного $A \in \mathcal{A}_N$ при всех $N > N_1$. Мы уже отмечали (см. абзац после (4.2.5)), что из $E(L_N \mid P_N) = 1$ следует $P_N(L_N > a) < 1/a$ для любого $a > 0$. Применяя это неравенство с $a = 1/\delta$ вместе с условием контигуальности, видим, что $Q_N(L_N > 1/\delta) < \varepsilon$. Теперь мы можем установить равномерную интегрируемость t по $\{F_N\}$:

$$\begin{aligned} \int_{1/\delta}^{\infty} t\, dF_N &= E(L_N \mathbf{1}_{\{L_N > 1/\delta\}} \mid P_N) \\ &= \int_{\{L_N > 1/\delta\}} L_N\, dP_N = \int_{\{L_N > 1/\delta\}} \frac{q_N}{p_N} p_N\, d\nu_N \\ &= Q_N(L_N > 1/\delta) < \varepsilon \end{aligned} \tag{4.2.8}$$

(здесь $\mathbf{1}_A$ обозначает функцию-индикатор множества A, равную 1 при $\mathbf{x} \in A$ и 0 вне A; первое равенство выполнено в силу (4.2.6)). Из равномерной интегрируемости следует, что

$$\int t\, dF = \lim_N \int t\, dF_N = 1,$$

т.е. свойство 4.

Импликация $4 \Rightarrow 3$. Для доказательства нам нужно доказать малость вероятностей $Q_N(\Lambda_N < -c)$ и $Q_N(\Lambda_N > c)$ при достаточно больших c. Рассмотрим сначала первую из этих вероятностей и покажем, что для нее справедлив аналог неравенства $P_N(L_N > a) < 1/a$, в силу которого ее малость достигается без всяких условий. Применяя (4.2.4) и (4.2.5) с переменой ролей P_N и Q_N, имеем

$$E(e^{-\Lambda_N} \mid Q_N) = E\left(\frac{p_N}{q_N} \mid Q_N\right) = \int \frac{p_N}{q_N} q_N \, d\nu_N = \int p_N \, d\nu_N = 1. \tag{4.2.9}$$

Отсюда, как и выше, получаем

$$Q_N(\Lambda_N < -c) = Q_N(\log(p_N/q_N) > c) = Q_N(p_N/q_N > e^c) < e^{-c}.$$

Следовательно, выбором достаточно большого c эта вероятность может быть сделана сколь угодно малой и содержательная часть доказываемой импликации состоит в малости $Q_N(\Lambda_N > c)$ при достаточно больших c.

Рассмотрим сначала подпоследовательность $\{m\} \subset \{N\}$, по которой $F_m \xrightarrow{w} F$. (Напомним, что $L_N > 0$ и, следовательно, $F_N(t) = 0$ при $t \leqslant 0$.) Согласно (4.2.4) $\int t \, dF_m = E(L_m \mid P_m) = 1$, а по условию 4 $\int t \, dF = 1$. Поэтому $\int t \, dF_m \to \int t \, dF$ (на самом деле имеет место равенство этих интегралов) и согласно п. 3 теоремы 8.2.5 функция t равномерно интегрируема по $F_m(t)$. Это значит, что для любого $\varepsilon > 0$ найдется $b > 0$ такое, что $\int_b^\infty t \, dF_m < \varepsilon$. Как и в (4.2.8), получаем, что этот интеграл равен $Q_m(L_m \geqslant b)$, откуда $Q_m(\Lambda_m \geqslant \log b) < \varepsilon$, что и доказывает утверждение.

Но мы пока доказали, что $\{\mathcal{L}[\Lambda_m \mid Q_m]\}$ плотна для всякой подпоследовательности $\{m\} \subset \{N\}$, по которой $F_m \xrightarrow{w} F$. Доказательство того, что вся последовательность $\{\mathcal{L}[\Lambda_N \mid Q_N]\}$ плотна, проведем от противного. Предположим, что существуют $\varepsilon > 0$ и последовательности $\{m\} \subset \{N\}$ и $c_m \to \infty$ такие, что $Q_m(\Lambda_m > c_m) \geqslant \varepsilon$. Мы отмечали, что последовательность $\{F_N\}$ (а следовательно, и любая ее подпоследовательность) всегда плотна, а тогда по теореме Прохорова она относительно компактна. Это значит, что из последовательности $\{F_m\}$ можно выбрать сходящуюся подпоследовательность $F_{m'} \xrightarrow{w} F$. По доказанному ранее найдется $b > 0$ такое, что $\int_b^\infty t \, dF_{m'} < \varepsilon$ в противоречии с нашим допущением.

Теорема доказана. $\square$

4.2.2. Общий случай, когда носители плотностей \$p_N\$ и \$q_N\$ возможно различны. Общий случай, когда носители плотностей p_N и q_N возможно различны

Обозначим

$$B_N = \{\mathbf{x}\colon p_N(\mathbf{x}) > 0\}, \quad C_N = \{\mathbf{x}\colon q_N(\mathbf{x}) > 0\}. \tag{4.2.10}$$

Положим $D_N = C_N \setminus B_N$. Это множество, на котором $p_N = 0$, а $q_N > 0$. Как следствие, $P_N(D_N) = 0$, а $Q_N(D_N) \geqslant 0$ (равенство $Q_N(D_N) = 0$ возможно, если $\nu_N(D_N) = 0$).

Как и в предыдущем пункте, за пределами множества $B_N \cup C_N$ обе плотности равны нулю, поэтому $P_N((B_N \cup C_N)^c) = Q_N((B_N \cup C_N)^c) = 0$, и мы снова отбросим это дополнение и отождествим $\mathbf{X}_N$ с $B_N \cup C_N$.

Поскольку мы не предполагаем, что $B_N = C_N$, возникают множества, где ОП $L_N = q_N/p_N$ и $\Lambda_N = \log L_N$ (см. (4.2.3)) не определены. Примем следующие определения:

$$L_N(\mathbf{x}) = \begin{cases} \dfrac{q_N(\mathbf{x})}{p_N(\mathbf{x})}, & \mathbf{x} \in B_N, \\ +\infty, & \mathbf{x} \in D_N, \end{cases} \tag{4.2.11}$$

и

$$\Lambda_N(\mathbf{x}) = \begin{cases} \log L_N(\mathbf{x}), & \mathbf{x} \in B_N \cap C_N, \\ +\infty, & \mathbf{x} \in D_N, \\ -\infty, & \mathbf{x} \in B_N \setminus C_N. \end{cases} \tag{4.2.12}$$

Введем соглашения относительно операций над символами $\pm\infty$. Условимся, что $e^{-\infty} = 0$, тогда последняя строка в (4.2.12) позволяет распространить соотношение $L_N = e^{\Lambda_N}$ на множество $B_N \setminus C_N$, на котором $p_N > 0$, $q_N = 0$, поскольку на этом множестве $L_N = q_N/p_N = 0$. Будем, кроме того, считать, что $+\infty > c$ для любого $c \in \mathbb{R}$.

Чтобы пояснить естественность соответствующего определения в (4.2.11) и (4.2.12), рассмотрим задачу проверки гипотезы, в которой P_N и Q_N играют роль гипотезы и альтернативы. В п. 8.1.2 Приложения приводится лемма Неймана–Пирсона для случая, рассмотренного в предыдущем пункте, когда носители этих распределений совпадают. В таком случае критическое множество имеет вид $S_N = \{\mathbf{x}\colon L_N(\mathbf{x}) > c\}$, где c выбирается так, чтобы для заданного $\alpha > 0$ было $P_N(S_N) = \alpha$. Мощность этого критерия равна $Q_N(S_N)$. В случае настоящего пункта наиболее мощный критерий для проверки P_N против Q_N имеет критическую область $S'_N = S_N \cup D_N$, где S_N задается тем же условием $L_N > c$, но только на множестве B_N, где ОП $L_N = q_N/p_N$ определено в собственном смысле, а множество D_N представляет собой «бесплатную добавку» к S_N в том отношении, что его присоединение к S_N не увеличивает уровня значимости α (т.к. оно имеет нулевую P_N-вероятность), но при этом, вообще говоря, увеличивает мощность, равную теперь $Q_N(S_N)+Q_N(D_N)$ (лемма Неймана–Пирсона в рассматриваемом здесь общем случае имеется в книге Лемана [7]). Определение $L_N = +\infty$ на множестве D_N дает возможность единообразного задания критической области S'_N в виде $S'_N = \{\mathbf{x} \in \mathbb{X}\colon L_N(\mathbf{x}) > c\}$.

Отметим некоторые отличия от предыдущих формул, возникающие в данном случае. Для любой измеримой функции f

$$E(f \mid P_N) = \int f(\mathbf{x}) p_N(\mathbf{x})\, d\nu_N = \int_{B_N} f(\mathbf{x}) p_N(\mathbf{x})\, d\nu_N, \qquad (4.2.13)$$

поскольку

$$\int_{B_N^c} f(\mathbf{x}) p_N(\mathbf{x})\, d\nu_N = 0. \qquad (4.2.14)$$

Мы считаем, что равенства (4.2.13) и (4.2.14) выполняются также для функций f, которые могут принимать бесконечные значения. Таким образом, например, равенство (4.2.4) переходит в неравенство (ср. с цепочкой равенств в (4.2.4))

$$E(e^{\Lambda_N} \mid P_N) = E(L_N \mid P_N) = \int L_N(\mathbf{x}) p_N(\mathbf{x})\, d\nu_N$$
$$= \int_{B_N} q_N\, d\nu_N = Q_N(B_N) \leqslant 1. \qquad (4.2.15)$$

Отметим, что из этого неравенства по-прежнему следует свойство (4.2.5), доказательство которого было основано на неравенстве $P_N(L_N > a) < 1/a$, которое также вытекает из (4.2.15).

При сделанных выше определениях и соглашениях формулировка теоремы 4.2.1 остается без изменений. Ниже мы приводим доказательство этой теоремы в условиях настоящего пункта, состоящее в сведении к случаю взаимно абсолютно непрерывных распределений, рассмотренному в предыдущем пункте.

Доказательство теоремы 4.2.1. Рассмотрим сначала случай, когда $C_N \subseteq B_N$. В этом случае $D_N = \varnothing$ и $Q_N \ll P_N$. Анализируя доказательство предыдущего пункта, нетрудно убедиться, что при этом предположении оно проходит почти без изменений. Единственное изменение касается равенства (4.2.9), которое аналогично (4.2.15) переходит в неравенство

$$E(e^{-\Lambda_N} \mid Q_N) = \int_{C_N} p_N \, d\nu = P_N(C_N) \leqslant 1,$$

что не влияет на последующие рассуждения. Отметим еще, что множество, где $\Lambda_N = \pm\infty$, имеет Q_N-меру нуль и тем самым введенные здесь соглашения не сказываются на доказательстве предыдущего пункта.

Поэтому мы можем считать, что теорема 4.2.1 доказана в случае, когда $C_N \subseteq B_N$. Рассмотрим теперь случай, когда $D_N = C_N \setminus B_N \neq \varnothing$.

Покажем прежде всего, что из свойства $Q_N \lhd P_N$ и каждого из условий 1–4 теоремы следует, что

$$Q_N(D_N) \to 0. \tag{4.2.16}$$

Из контигуальности $Q_N \lhd P_N$ следует (4.2.16) по определению, поскольку $P_N(D_N) = 0$ (а тем самым $P_N(D_N) \to 0$).

Доказательства (4.2.16) при условиях 1–4 проводятся от противного. При условиях 1 и 2 достаточно привести примеры последовательностей T_N, не удовлетворяющих этим условиям, если $Q_N(D_N) \nrightarrow 0$. Например, можно положить $T_N = N$ на множестве D_N (а в случае условия 1 можно взять $T_N = \mathrm{const} \neq 0$). Предлагаем читателю проверить это самостоятельно.

При условии 3 также допустим, что $Q_N(D_N) \nrightarrow 0$. Тогда существуют $\varepsilon > 0$ и подпоследовательность $\{m\} \subset \{N\}$ такие, что $Q_m(D_m) > \varepsilon$. По определению (4.2.12) $\Lambda_m = \infty$ на множестве D_m и тем самым с Q_m-вероятностью, большей ε, Λ_m лежит вне любого компакта в противоречии с условием 3 (см. определение 8.2.4 в п. 8.2.3).

Напомним, что в условии 4 рассматриваются функции распределения $F_N(t) = P_N(L_N < t)$. Как и выше, рассуждая от противного, найдем $\varepsilon > 0$ и подпоследовательность $\{m\} \subset \{N\}$ для которой $Q_m(D_m) > \varepsilon$. Согласно следствию 4.2.1 найдется подпоследовательность $\{m'\} \subset \{m\}$ и ф.р. F такие, что $F_{m'}(t) \to F(t)$. По условию 4 $\int t \, dF(t) = 1$, а по теореме 8.2.5, п. 1,

$$\liminf_{m' \to \infty} \int t \, dF_{m'}(t) \geqslant \int t \, dF(t) = 1.$$

В то же время в силу (4.2.15)

$$\int t\,dF_{m'}(t) = E(L_{m'} \mid P_{m'}) = Q_{m'}(B_{m'}) = 1 - Q_{m'}(D_{m'}) \leqslant 1 - \varepsilon$$

в противоречии с предыдущим неравенством.

Введем условное распределение

$$\widetilde{Q}_N(A) = Q_N(A \mid B_N) = \frac{Q_N(A \cap B_N)}{Q_N(B_N)}, \quad A \in \mathcal{A}. \tag{4.2.17}$$

Это распределение имеет плотность

$$\tilde{q}_N(\mathbf{x}) = \frac{q_N(\mathbf{x})}{Q_N(B_N)}, \quad \mathbf{x} \in B_N, \tag{4.2.18}$$

и $\tilde{q}_N(\mathbf{x}) = 0$ вне B_N. Поэтому, если ввести множества $\widetilde{C}_N = \{\mathbf{x}\colon \tilde{q}_N(\mathbf{x}) > 0\}$, то $\widetilde{C}_N \subseteq B_N$, и распределения $(P_N, \widetilde{Q}_N)$ удовлетворяют условию, рассмотренному в начале доказательства, при котором $\widetilde{Q}_N \lhd P_N$.

Обозначим $\delta_N = 1 - Q_N(B_N) = Q_N(D_N)$. Согласно (4.2.16) $\delta_N \to 0$ при $N \to \infty$. Из (4.2.18) получаем

$$\widetilde{\Lambda}_N(\mathbf{x}) = \frac{\tilde{q}_N(\mathbf{x})}{p_N(\mathbf{x})} = \frac{\Lambda_N(\mathbf{x})}{1 - \delta_N} \tag{4.2.19}$$

на множестве B_N.

Далее, по формуле полной вероятности для произвольного $A \in \mathcal{A}$

$$Q_N(A) = Q_N(A \mid B_N)(1 - \delta_N) + Q_N(A \mid B_N^c)\delta_N.$$

Отсюда

$$|Q_N(A) - Q_N(A \mid B_N)| = \delta_N |Q_N(A \mid B_N^c) - Q_N(A \mid B_N)| \leqslant \delta_N.$$

Следовательно,

$$\sup_{A \in \mathcal{A}} |\widetilde{Q}_N(A) - Q_N(A)| \leqslant \delta_N \to 0 \text{ при } N \to \infty. \tag{4.2.20}$$

Нетрудно проверить (предоставляем это читателю), что благодаря (4.2.19) и (4.2.20) условие $Q_N \lhd P_N$ и каждое из условий 1–4 для последовательности пар (P_N, Q_N) эквивалентны (т.е. выполнены тогда и только тогда ...) условию $\widetilde{Q}_N \lhd P_N$ или соответствующему условию 1–4 для последовательности пар $(P_N, \widetilde{Q}_N)$. Если мы обозначим через A и B любые два из пяти указанных выше условий, относящихся к (P_N, Q_N), а через $\widetilde{A}$ и $\widetilde{B}$ соответствующие условия, относящиеся к $(P_N, \widetilde{Q}_N)$, то в силу сказанного $A \Longleftrightarrow \widetilde{A}$ и $B \Longleftrightarrow \widetilde{B}$, а из отмеченной выше справедливости теоремы для $(P_N, \widetilde{Q}_N)$ следует $\widetilde{A} \Longleftrightarrow \widetilde{B}$. Следовательно, $A \Longleftrightarrow B$, что и требовалось доказать. □

Для большей логической ясности мы рассмотрели «одностороннюю» контигуальность $Q_N \lhd P_N$, хотя нас обычно будет интересовать взаимная контигуальность, когда $Q_N \lhd P_N$ и $P_N \lhd Q_N$ (обозначается $P_N \lhd\rhd Q_N$). Контигуальность $P_N \lhd Q_N$ устанавливается с помощью той же теоремы (4.2.1), в которой P_N и Q_N следует поменять ролями. Распределения P_N, отвечающие гипотезе H_0, как правило устроены проще, чем распределения Q_N, отвечающие гипотезе H_1. Так, в связи с ранговыми критериями P_N это совместное распределение N независимых *одинаково распределенных* случайных величин, тогда как Q_N соответствует неодинаково распределенным величинам. Поэтому при проверке контигуальности $Q_N \lhd P_N$ удобно пользоваться условием 4, формулируемым в терминах распределений P_N, а при проверке $P_N \lhd Q_N$ – условием 3, которое при перемене ролей P_N и Q_N также формулируется в терминах распределений P_N.

В § 3.2 была доказана теорема об асимптотической нормальности ранговых статистик (точнее, простых линейных ранговых статистик (ПЛРС)) при гипотезе H_0. В этой теореме было получено представление ПЛРС в виде нормированной суммы независимых случайных величин, к которой применима ц.п.т., с остаточным членом, сходящимся к нулю по вероятности. В последующем мы получим распределения тех же статистик при контигуальных альтернативах. Но уже сейчас отметим полезное свойство этого понятия: согласно п. 1 теоремы 4.2.1 остаточный член, сходящийся к нулю при гипотезе H_0, сходится к нулю и при контигуальных альтернативах. Установить сходимость остаточного члена к нулю при альтернативе каким-либо иным способом было бы совсем непростой задачей.

Терминологическое замечание. Термин *contiguous* был введен в работе Ле Кама [24]. Довольно долго не удавалось подобрать подходящего русского эквивалента этого термина. Чаще всего альтернативы, о которых идет речь, назывались «близкими» (а в настоящее время их зачастую называют «локальными»). Это слово, близкое по смыслу к английскому *contiguous* (смежный, прилегающий), воспринимается в контексте асимптотической теории как «сближающиеся». Это верно в отношении параметров, описывающих распределения (в рассмотренном примере параметры контигуальных распределений отличаются на величину порядка $1/\sqrt{N}$), но термин относится к совместным распределениям, которые вовсе не обязаны сближаться, а должны не «разбегаться», не удаляться друг от друга. Русский термин «контигуальный» был введен при переводе книги Я. Гаека и З. Шидака [3], где это понятие играет важную роль. В отличие от английского слова, взятого из повседневного английского языка, его русский перевод сконструирован искусственно.

Мы будем употреблять выражения «контигуальные альтернативы» и «локальные альтернативы» как синонимы, в зависимости от того, идет ли речь о совместных распределениях, контигуальных относительно нулевой гипотезы, или о соответствующих множествах значений параметра.

Глава 5. Локальная асимптотическая нормальность

§ 5.1. Введение

Напомним представление (8.1.11) для логарифма ОП при локальных альтернативах $\mu_N = \mu_{N,t} = \mu_0 + t/\sqrt{N}$ в случае нормального распределения, полученное в п. 8.1.3 Приложения. Мы несколько изменим обозначения плотностей нормального распределения, определяемых формулами (8.1.6) и (8.1.8), а именно, вместо p_N и q_N будем писать $p_{N,0}$ и $p_{N,t}$, указывая на то, что это плотности нормальных распределений со средними $\mu_{N,t}$, отвечающими $t = 0$ и произвольному $t > 0$. Обозначим

$$Z_N = \frac{1}{\sigma^2\sqrt{N}} \sum_1^N (X_i - \mu_0). \tag{5.1.1}$$

Тогда (8.1.11) можно переписать как

$$\Lambda_{N,t} = tZ_N - \frac{1}{2}\frac{t^2}{\sigma^2} \tag{5.1.2}$$

(здесь мы тоже добавляем индекс при Λ, указывающий на зависимость от t; обратим внимание на то, что Z_N не зависит от t). Выражение (5.1.2) показывает, что рассматриваемое семейство N-мерных нормальных распределений, зависящее от параметра t, образует экспоненциальное семейство с достаточной статистикой Z_N.

Вспоминая определение (4.2.3), формуле (5.1.2) можно придать вид, связанный с *критерием факторизации*:

$$p_{N,t}(\mathbf{X}) = p_{N,0}(\mathbf{X}) \exp(-t^2/2\sigma^2) \exp(tZ_N), \tag{5.1.3}$$

где первый сомножитель в правой части есть функция только от наблюдений (не зависящая от параметра), второй сомножитель зависит только от параметра, и третий зависит от параметра и статистики Z_N, которая и является достаточной статистикой.

В п. 8.1.3 показано также, что по лемме Неймана–Пирсона наиболее мощный критерий проверки гипотезы $H_0\colon \mu = \mu_0$ против любой простой альтернативы $H_{N,t}\colon \mu = \mu_{N,t}$, $t > 0$, отвергает H_0 при

$$Z_N > z_\alpha/\sigma \tag{5.1.4}$$

(см. (8.1.16) в п. 8.1.3; сравнивая формулы (8.1.17) и (5.1.1), имеем $\sigma Z_N = S_N$).

Мы столь подробно разбираем общеизвестный пример, относящийся к семейству нормальных распределений, зависящих от параметра сдвига, по той причине, что в асимптотической теории, которая будет строиться далее, рассмотренное семейство нормальных распределений служит в определенном смысле предельной статистической моделью. А именно, мы будем рассматривать локальные альтернативы к гипотезе H_0 одинаковой распределенности наблюдений, с которой мы имели дело в главе 3. Будет показано, что при таких альтернативах логарифм ОП допускает представление, аналогичное (5.1.2), с тем отличием, что распределение статистики Z_N (имеющей теперь, естественно, другую форму) не нормально, как выше, но сходится к нормальному, второе слагаемое имеет вид $-\frac{1}{2}t^2 I$, где I – некоторый функционал от распределения наблюдений (информация Фишера, равная $1/\sigma^2$ в рассмотренном примере), и правая часть содержит остаточный член, сходящийся к нулю по вероятности.

Семейства распределений, для которых логарифм ОП допускает такое представление, называются *локально асимптотически нормальными* (ЛАН). Мы увидим, что статистика Z_N, входящая в это представление, обладает свойствами, аналогичными (в асимптотическом смысле) свойствам статистики (5.1.1), входящей в представление (5.1.2), в частности, критерий, основанный на статистике Z_N, является локально асимптотически наиболее мощным (точное определение этого будет дано ниже). Кроме того, что особенно важно, указанный аналог представления (5.1.2) лежит в основе приема, позволяющего переходить от распределений статистик при гипотезе к их распределениям при локальных альтернативах.

§ 5.2. ЛАН семейства в задачах регрессионного типа

5.2.1. Постановка задачи. Как и в § 2.1, мы имеем вектор наблюдений $\mathbf{X} = (X_1, \ldots, X_N)$, состоящий из независимых случайных величин, но теперь, в отличие от § 2.1, эти величины распределены, вообще говоря, неодинаково. Будем предполагать, что распределения величин $X_1, \ldots, X_N$ имеют плотности по мере Лебега, принадлежащие некоторому семейству $p_\theta(x)$, $\theta \in \Theta$, где Θ – множество, содержащее интервал вида $\Delta = [-\delta, \delta]$, $\delta > 0$. Таким образом, каждое наблюдение X_i имеет плотность $p_{\theta_i}(x)$ с некоторым $\theta_i \in \Theta$. Предположим, что при каждом N задан набор констант $c_{N,1}, \ldots, c_{N,N}$, нормированных так, чтобы

$$\sum_{i=1}^{N} c_{N,i}^2 = 1, \tag{5.2.1}$$

и удовлетворяющих условию равномерной малости

$$\max_i |c_{N,i}| \to 0 \text{ при } N \to \infty. \tag{5.2.2}$$

Для некоторого $t > 0$ рассмотрим проверку гипотезы

$$H_0\colon \theta_i = 0, \quad i = 1, \ldots, N, \tag{5.2.3}$$

против альтернативы

$$H_1\colon \theta_i = c_{N,i}t, \quad i = 1, \ldots, N. \tag{5.2.4}$$

(Таким образом, при H_1 случайные величины $X_1, \ldots, X_N$ имеют распределения, определяемые набором констант $c_{N,1}, \ldots, c_{N,N}$, изменяющимся с изменением N. Поэтому точнее было бы писать $X_{N,1}, \ldots, X_{N,N}$, но для упрощения обозначений мы не будем вводить двойных индексов при X.) Если $c_{N,1} = \cdots = c_{N,N}$, то согласно (5.2.1) имеем $c_{N,i} = 1/\sqrt{N}$. В этом случае наблюдения $X_1, \ldots, X_N$ одинаково распределены как при гипотезе H_0 (5.2.3), так и при альтернативе H_1 (5.2.4), и такая альтернатива неотличима от H_0 при помощи ранговых критериев. Тем не менее, поскольку наши ближайшие рассмотрения, посвященные свойству локальной асимптотической нормальности (ЛАН), носят в этом отношении общий характер, мы не исключаем этот случай, а при переходе к ранговым критериям константы $c_{N,1}, \ldots, c_{N,N}$ будут подчинены дополнительным требованиям.

5.2.2. Предварительный вывод свойства ЛАН. В этом пункте мы проведем упрощенный вывод свойства ЛАН в описанной выше модели. А именно, не формулируя точных условий, будем предполагать, что участвующие в наших выводах функции дифференцируемы необходимое число раз, а возникающие при этом интегралы сходятся и допускают дифференцирование по параметру под знаком интеграла. Этот вывод имеет целью показать естественность свойства ЛАН. В следующем пункте мы приводим формулировки двух теорем, дающих достаточные условия для выполнения этого свойства. Эти условия гораздо менее ограничительны, чем те, которые потребовались бы для оправдания проводимого здесь вывода. Доказательству этих теорем будет посвящена глава 7.

Для простоты зафиксируем некоторое t (предполагая для определенности $t > 0$) и рассмотрим последовательность задач проверки простой гипотезы H_0 (5.2.3) против простых альтернатив H_1 (5.2.4), определяемых данным значением t. Как сказано выше, каждое наблюдение X_i имеет плотность $p_{\theta_i}(x)$ с $\theta_i = c_{N,i}t$, $i = 1, \ldots, N$, где гипотезе H_0 соответствует $t = 0$, а альтернативе H_1 – выбранное значение $t > 0$. Тогда совместная плотность распределения вектора $\mathbf{X} = (X_1, \ldots, X_N)$ равна

$$p_{N,t}(\mathbf{x}) = \prod_{i=1}^{N} p_{c_{N,i}t}(x_i), \tag{5.2.5}$$

где $p_{N,0}(\mathbf{x})$ соответствует гипотезе H_0. Примем дополнительное условие: множество

$$\{x \in \mathbb{R} \colon p_\theta(x) > 0\} \ \text{ не зависит от } \ \theta \in \Theta \tag{5.2.6}$$

(аналог условия, использованного в п. 4.2.1). Тогда определено отношение правдоподобия (ОП)

$$L_{N,t} = \frac{p_{N,t}(\mathbf{X})}{p_{N,0}(\mathbf{X})} \tag{5.2.7}$$

и логарифм ОП $\Lambda_{N,t} = \log L_{N,t}$. Введем обозначение $l_\theta(x) = \log p_\theta(x)$. Тогда в силу (5.2.5) и (5.2.7)

$$\Lambda_{N,t} = \log\left(\frac{p_{N,t}(\mathbf{X})}{p_{N,0}(\mathbf{X})}\right) = \sum_{i=1}^{N} \bigl(l_{c_{N,i}t}(X_i) - l_0(X_i)\bigr). \tag{5.2.8}$$

Примем следующие обозначения: частные производные функции $l_\theta(x)$ по θ будем обозначать верхним индексом в скобках, указывающим порядок производной, т.е.

$$l_\theta^{(j)}(x) = \frac{\partial^j l_\theta(x)}{\partial \theta^j}, \quad j = 1, 2, \dots . \tag{5.2.9}$$

Когда $\theta = 0$, индекс 0 будем опускать, т.е., например, писать $l(X_i)$ вместо $l_0(X_i)$; точно так же $l^{(j)}(x) = l_\theta^{(j)}(x)|_{\theta=0}$ и т.п. Такое же обозначение частных производных по θ будет использоваться в применении и к другим функциям, скажем, $p_\theta^{(j)}(x)$.

Разложим каждое слагаемое в правой части (5.2.8) по формуле Тейлора:

$$l_{c_{N,i}t}(X_i) - l(X_i) = c_{N,i}t\, l^{(1)}(X_i) + \frac{1}{2}\, c_{N,i}^2 t^2\, l^{(2)}(X_i) + \cdots . \tag{5.2.10}$$

Нам потребуются некоторые хорошо известные тождества, встречающиеся, например, при доказательстве неравенства Рао–Крамера. По определению, $p_\theta(x) = e^{l_\theta(x)}$ и следовательно,

$$\int e^{l_\theta(x)}\, dx \equiv 1, \quad \theta \in \Theta.$$

Дифференцируя это тождество два раза под знаком интеграла, получаем

$$\int l_\theta^{(1)} e^{l_\theta}\, dx = 0, \quad \int [l_\theta^{(2)} + (l_\theta^{(1)})^2]\, e^{l_\theta}\, dx = 0$$

(опуская аргумент x у подынтегральной функции). В частности, при $\theta = 0$ имеем

$$E_0 l^{(1)} = \int l_0^{(1)}(x) p_0(x)\, dx = 0, \tag{5.2.11}$$

и аналогично

$$E_0 l^{(2)} = -E_0 (l^{(1)})^2.$$

Величина

$$I_\theta = E_\theta (l_\theta^{(1)}(\mathbf{X}))^2 = \int (l_\theta^{(1)}(x))^2 p_\theta(x)\, dx, \quad \theta \in \Theta,$$

называется *информацией Фишера* для семейства плотностей $p_\theta(x)$. Следуя принятому соглашению, пишем $I = I_0$, опуская индекс 0. Таким образом,

$$I = E_0 (l^{(1)})^2 = -E_0 l^{(2)}. \tag{5.2.12}$$

Мы предполагаем, что $I < \infty$.

Подставляя разложение (5.2.10) в (5.2.8), получаем

$$\Lambda_{N,t} = tZ_N + \frac{1}{2}t^2 \sum_1^N c_{N,i}^2\, l^{(2)}(X_i) + \cdots\,, \tag{5.2.13}$$

где

$$Z_N = \sum_1^N c_{N,i} l^{(1)}(X_i). \tag{5.2.14}$$

Рассмотрим асимптотическое поведение каждого слагаемого в (5.2.13). К Z_N применима ц.п.т. для сумм специального вида, уже использовавшаяся нами в п. 3.2.2 и приводимая как теорема 8.3.1 в п. 8.3.1 Приложения. Используя условие (5.2.1) и соотношения (5.2.11) и (5.2.12), получаем

$$\mathcal{L}[Z_N \mid H_0] \to \mathcal{N}(0, I). \tag{5.2.15}$$

Применяя закон больших чисел для сумм специального вида, см. п. 8.3.2 Приложения, теорема 8.3.5, получаем

$$\sum_1^N c_{N,i}^2\, l^{(2)}(X_i) \xrightarrow{p} E_0 l^{(2)} = -I, \tag{5.2.16}$$

где $\xrightarrow{p}$ означает сходимость по вероятности.

Наконец, не выписанное явно слагаемое, включающее третью производную функции l с множителями $c_{N,i}^3$, сходится к нулю по вероятности при условиях, налагаемых обычно при доказательстве, например, асимптотической нормальности оценок максимального правдоподобия (заметим, что $|c_{N,i}|^3 \leqslant c_{N,i}^2 \max_j |c_{Nj}|$, а по условию (5.2.2) $\max_j |c_{Nj}| \to 0$ при $N \to \infty$). Предлагаем читателю самостоятельно сформулировать условия и провести доказательство сходимости к нулю этого остаточного члена.

В результате из (5.2.13), заменяя второе слагаемое в правой части его пределом согласно (5.2.16) и включая возникающую при этом погрешность в остаточный член ζ_{Nt}, получаем

$$\Lambda_{N,t} = tZ_N - \frac{1}{2}t^2 I + \zeta_{N,t}, \tag{5.2.17}$$

где $\zeta_{Nt} \xrightarrow{p} 0$ при $N \to \infty$. При выполнении представления (5.2.17) для логарифма ОП говорят, что семейство распределений с плотностями $p_{N,t}(\mathbf{x})$ (см. (5.2.5)) *локально асимптотически нормально*.

5.2.3. Теоремы о достаточных условиях для ЛАН. В этом пункте мы формулируем две теоремы, дающие достаточные условия весьма общего характера для выполнения свойства ЛАН. Как уже говорилось, их доказательства будут даны в гл. 7.

Мы рассматриваем постановку задачи, описанную в п. 5.2.1, не предполагая, в отличие от п. 5.2.2, условия (5.2.6). В первой теореме доказывается свойство ЛАН при условии, что функция $\sqrt{p_\theta(x)}$ дифференцируема по θ в среднем квадратичном в точке $\theta = 0$. Вторая теорема показывает, что достаточным условием этого является существование в окрестности точки $\theta = 0$ информации Фишера, непрерывной при $\theta = 0$. Условия этой теоремы являются по существу минимальными, поскольку информация Фишера входит в само соотношение (5.2.17). Подчеркнем, что в обеих этих теоремах не требуется, чтобы носитель плотности $\{x\colon p_\theta(x) > 0\}$ не зависел от θ.

Первая из этих теорем (в отличие от второй) допускает распространение на случай векторного параметра θ, поэтому ее условия представляют самостоятельный интерес, чем и объясняется раздельная формулировка двух теорем.

Прежде, чем формулировать теорему, поясним ее формулировку. Упомянутое условие дифференцируемости в среднем квадратичном связано с тем, что существование производной функции двух переменных, скажем, $f_\theta(x)$ по θ в точке $\theta = 0$, означает существование предела

$$g(x) = \lim_{\theta\to 0} \frac{f_\theta(x) - f_0(x)}{\theta}. \tag{5.2.18}$$

Рассматриваемая обычно в анализе операция дифференцирования по одной из переменных предполагает существование предела (5.2.18) в фиксированной точке x. В этом случае $g(x) = f_\theta^{(1)}(x)|_{\theta=0}$ (следуя обозначениям, введенным в (5.2.9)). Но сходимость *функций* к предельной функции может пониматься и в смысле сходимости в каком-либо функциональном пространстве с той или иной нормой (метрикой, топологией, и т.п.). В данном случае рассматривается дифференцируемость в среднем квадратичном, которая применительно к (5.2.18) означает сходимость

$$\int \left(\frac{f_\theta(x) - f_0(x)}{\theta} - g(x) \right)^2 dx \to 0. \tag{5.2.19}$$

Функция $g(x)$ в (5.2.19) может не быть производной в обычном смысле (т.е. пределом (5.2.18) в данной точке x). Поэтому мы будем различать производную в среднем квадратичном и обычную производную $f^{(1)}$, используя для них, в частности, разные обозначения.

Наша цель – показать, что выполнение (5.2.19) с $f_\theta(x) = \sqrt{p_\theta(x)}$ и некоторой функцией $g(x)$ влечет выполнение (5.2.17) с

$$I = 4\int g^2(x)\,dx, \quad Z_N = \sum_1^N c_{N,i}h(X_i), \quad h(x) = \frac{2g(x)}{\sqrt{p(x)}}. \tag{5.2.20}$$

Чтобы пояснить эти соотношения, предположим, что плотность $p_\theta(x)$ дифференцируема в обычном смысле. Тогда

$$g = \frac{\partial\sqrt{p_\theta}}{\partial\theta}\Big|_{\theta=0} = \frac{p^{(1)}}{2\sqrt{p}}, \quad h = l^{(1)} = \frac{p^{(1)}}{p} = \frac{2g}{\sqrt{p}},$$
$$I = E\big(l^{(1)}\big)^2 = 4\int\frac{g^2}{p}p\,dx.$$

В теореме доказывается несколько усиленное соотношение (5.2.17), в котором t может меняться вместе с N. Приведем точную формулировку этого утверждения.

Теорема 5.2.1. *Предположим, что в схеме, описанной в п. 5.2.1, константы $c_{N,1},\ldots,c_{N,N}$ удовлетворяют условиям (5.2.1) и (5.2.2), а плотности $p_\theta(x)$, $\theta\in\Theta$, – условию*

$$\int\left(\frac{\sqrt{p_\theta(x)}-\sqrt{p_0(x)}}{\theta} - g(x)\right)^2 dx \to 0 \tag{5.2.21}$$

для некоторой функции $g(x)$ такой, что $\int g^2(x)\,dx < \infty$ и $g(x) = 0$ на множестве $\{x\colon p_0(x) = 0\}$. Тогда для любой ограниченной последовательности $t_N\in\mathbb{R}$

$$\Lambda_{N,t_N} = t_N Z_N - \frac{1}{2}t_N^2 I + \zeta_{N,t_N}, \tag{5.2.22}$$

где Z_N и I определены в (5.2.20), $Eh = 0$, и $\zeta_{N,t_N}\xrightarrow{P} 0$ относительно $P_{N,0}$. Как следствие,

$$\mathcal{L}(Z_N \mid P_{N,0}) \to \mathcal{N}(0, I). \tag{5.2.23}$$

Доказательство этого утверждения (в несколько иной форме) имеется в книге Гаека и Шидака [3]. Предлагаемое в гл. 7 доказательство – более громоздкое, но по нашему мнению оно раскрывает некоторые интересные связи и объясняет, почему именно *корень из плотности* играет такую особую роль в формулировке этого условия.

Теперь мы сформулируем упомянутую выше вторую теорему.

Теорема 5.2.2. *Предположим, что плотность $p_\theta(x)$ удовлетворяет следующим условиям:*

(A) *При каждом $x \in \mathbb{R}$ плотность $p_\theta(x)$ абсолютно непрерывна по θ в некоторой окрестности точки $\theta = 0$;*

(B) *При каждом θ из этой окрестности производная $p_\theta^{(1)}(x) = \frac{\partial}{\partial\theta} p_\theta(x)$ существует при почти всех (по мере Лебега) $x \in \mathbb{R}$;*

(C) *Функция*

$$I(\theta) = E_\theta \left(\frac{p_\theta^{(1)}}{p_\theta} \right)^2 < \infty,$$

положительна и непрерывна в этой окрестности.

Тогда выполнены условия теоремы 5.2.1 с $g(x) = p_0^{(1)}(x)/\big(2 \times \sqrt{p_0(x)}\big)$ на множестве $\{x\colon p_0(x) > 0\}$ и $g(x) = 0$, если $p_0(x) = 0$, а следовательно, утверждения (5.2.22) и (5.2.23) этой теоремы.

Таким образом, при условиях этой теоремы функция $l^{(1)}(x)$ определена при почти всех $x \in \mathbb{R}$ и условие ЛАН состоит в выполнении (5.2.22) с Z_N и I, определенными равенствами (5.2.14) и (5.2.12).

Пример 5.2.1. Приведем пример, показывающий, как «работают» условия дифференцируемости корня из плотности в среднем квадратичном и существования конечной информации Фишера в случае, когда носитель плотности зависит от параметра.

Рассмотрим плотность гамма-распределения $p(x) = (1/\Gamma(\alpha)) \times x^{\alpha-1}e^{-x}$, $\alpha > 0$, при $x > 0$, $p(x) = 0$ при $x \leqslant 0$. Мы будем рассматривать семейство сдвигов этого распределения, т.е. семейство $p_\theta(x) = p(x - \theta)$, $\theta \in \mathbb{R}$. Функция $p(x)$ бесконечно дифференцируема всюду за исключением точки $x = 0$, где она может быть разрывной (при $\alpha \leqslant 1$), недифференцируемой (при $1 < \alpha \leqslant 2$) или иметь ограниченное число производных (зависящее от α). Мы рассматриваем проверку гипотезы $\theta = 0$ при локальных альтернативах $\theta_i = c_{N,i}t$ с $c_{N,i}$, удовлетворяющими условиям (5.2.1) и (5.2.2). Из-за того, что носитель плотности (полупрямая $x > \theta$) зависит от θ, логарифм отношения правдоподобия (ОП), фигурирующий в условии ЛАН, не определен на интервале $(0, \theta)$ (или $(\theta, 0)$ в зависимости от знака θ), когда в числителе плотность равна нулю, а в знаменателе положительна, либо наоборот. Такая неопределенность допустима, если вероятность попадания наблюдения в интервал $(0, \theta)$ (или $(\theta, 0)$) имеет порядок $o(\theta^2)$. В этом случае при $\theta_i = c_{N,i}t$ вероятность попадания хотя бы одного наблюдения в интервал неопределенности равна $o(\sum c_{N,i}^2) = o(1)$, т.е. возможность того, что логарифм ОП не определен, имеет пренебрежимо малую вероятность при $N \to \infty$ (это выражение свойства (4.2.16) (см. п. 4.2.2) в данной схеме наблюдений). Пусть $\theta > 0$. Тогда вероятность попадания наблюдения в интервал $(0, \theta)$ при малых θ имеет порядок

$$\int_0^\theta x^{\alpha-1}\,dx = \frac{1}{\alpha}\theta^\alpha$$

и условие, что эта вероятность имеет порядок $o(\theta^2)$, выполняется при $\alpha > 2$. Это по-существу минимальное требование, при котором асимптотический анализ логарифма ОП имеет смысл.

Посмотрим теперь, при каких α информация Фишера I конечна. Напомним, что $I = E(l^{(1)})^2$, где $l^{(1)} = \partial \log p/\partial\theta$. Но если $p_\theta(x)$ зависит от разности $x - \theta$, то $l^{(1)} = -\partial \log p/\partial x$. Кроме того, в случае параметра сдвига информация Фишера не зависит от θ, поэтому будем вычислять ее при $\theta = 0$. Находим $\log p = \mathrm{const} + (\alpha - 1)\log x - x$, $d\log p/dx = \frac{\alpha-1}{x} - 1$ и

$$I = \int_0^\infty \left(\frac{\alpha-1}{x} - 1\right)^2 p(x)\,dx.$$

Учитывая, что $p(x)$ при малых x имеет порядок $Cx^{\alpha-1}$, подынтегральная функция в окрестности нуля имеет особенность порядка $x^{\alpha-3}$, которая интегрируема, если $\alpha - 3 > -1$, т.е. $\alpha > 2$. Таким образом, конечность информации Фишера эквивалентна полученному выше условию, при котором неопределенность логарифма ОП асимптотически пренебрегаема.

Рассмотрим теперь условие дифференцируемости корня из плотности в среднем квадратичном, т.е. условие (5.2.21) с $g(x) = -d\sqrt{p(x)}/dx$. Здесь мы проверим часть этого условия, касающуюся поведения подынтегральной функции вблизи нуля. Мы видели выше, что при $\alpha > 2$ так определенная функция $g(x)$ интегрируема в квадрате,

$$\int_0^\infty g^2(x)\,dx < \infty.$$

Будем считать для определенности, что $\theta > 0$. Тогда для выполнения (5.2.21) необходимо, чтобы

$$\int \left(\frac{\sqrt{p(x-\theta)} - \sqrt{p(x)}}{\theta} \right)^2 dx < \infty,$$

а интеграл по интервалу $[0, \theta)$ должен сходиться к нулю при $\theta \to 0$. При $x < \theta$ имеем $\sqrt{p(x-\theta)} = 0$, следовательно этот интеграл равен

$$\frac{1}{\theta^2} \int_0^\theta p(x)\,dx \sim \mathrm{const}\,\frac{\theta^\alpha}{\theta^2}$$

и видим, что он стремится к нулю при том же условии $\alpha > 2$. Детальную проверку условия (5.2.21) предоставляем читателю.

§ 5.3. Следствия ЛАН: контигуальность и распределения при альтернативах

В этом параграфе мы рассмотрим связь между контигуальностью и свойством ЛАН, а также следствия этих свойств, касающиеся связи между распределениями статистик при гипотезе и при альтернативе. Эти свойства будут изучаться в общей постановке, введенной в § 4.2, с тем отличием, что при каждом N будет рассматриваться некоторое *семейство* распределений (а не два распределения, как в § 4.2).

Итак, мы предполагаем, что при каждом $N = 1, 2, \ldots$ имеется выборочное пространство $(\mathbb{X}_N, \mathcal{A}_N)$ с σ-конечной мерой ν_N на нем, и на каждом таком пространстве задано семейство распределений $\{P_{N,t}\}$, $t \in \mathcal{T}$, где $\mathcal{T} \subset \mathbb{R}$ – некоторый интервал на прямой $\mathbb{R}$ (возможно, совпадающий со всей прямой $\mathbb{R}$), содержащий точку $t = 0$. Предположим, далее, что распределения $P_{N,t}$ имеют плотности $p_{N,t}$, $t \in \mathcal{T}$, по мере ν_N. (В рассмотренном выше случае, когда $\mathbf{X}$ – вектор, состоящий из N независимых наблюдений, пространство $\mathbb{X}_N = \mathbb{R}^N$ и плотности $p_{N,t}$ являются произведениями одномерных плотностей, см. (5.2.5). Эта схема будет далее рассматриваться в главе 6.)

В случае, когда носитель плотности, $\{\mathbf{x} \in \mathbb{X}_N : p_{N,t}(\mathbf{x}) > 0\}$, не зависит от t, определим отношение правдоподобия (ОП) $L_{N,t}$ и логарифм ОП $\Lambda_{N,t}$ формулами[1]

$$L_{N,t} = L_{N,t}(\mathbf{X}) = \frac{p_{N,t}(\mathbf{X})}{p_{N,0}(\mathbf{X})}, \quad \Lambda_{N,t} = \Lambda_{N,t}(\mathbf{X}) = \log L_{N,t}(\mathbf{X}). \tag{5.3.1}$$

В общем же случае доопределим $L_{N,t}$ и $\Lambda_{N,t}$ аналогично (4.2.11) и (4.2.12). Мы предполагаем, что семейство $\{P_{N,t}\}$ обладает следующим свойством:

Свойство локальной асимптотической нормальности (*ЛАН*). Существуют статистики $Z_N = Z_N(\mathbf{X})$ и константа $I > 0$ такие, что

$$\mathcal{L}(Z_N \mid P_{N,0}) \to \mathcal{N}(0, I) \ \text{ при } \ N \to \infty, \tag{5.3.2}$$

и Λ_{N,t_N} для любой ограниченной последовательности[2] $\{t_N\}$ допускает представление

$$\Lambda_{N,t_N} = t_N Z_N - \frac{1}{2} t_N^2 I + \zeta_{N,t_N}, \tag{5.3.3}$$

где $\zeta_{N,t_N} \xrightarrow{p} 0$ относительно $P_{N,0}$, т.е. для любого $\varepsilon > 0$

$$P_{N,0}(|\zeta_{N,t_N})| > \varepsilon) \to 0 \ \text{ при } \ N \to \infty. \tag{5.3.4}$$

Замечание 5.3.1. Напомним еще раз, что мы не предполагаем, что носитель плотности, $\{\mathbf{x} \in \mathbb{X}_N : p_{N,t}(\mathbf{x}) > 0\}$, не зависит от t. Это значит, что согласно определению, аналогичному (4.2.12), $\Lambda_{N,t}$ может принимать значения $\pm\infty$ с положительной вероятностью. Статистика Z_N предполагается определенной и принимающей конечные значения $P_{N,0}$-почти всюду, т.е. на множестве, имеющем $P_{N,0}$-вероятность 1[3]. Поэтому первые два слагаемых в правой части (5.3.3) конечны $P_{N,0}$-почти всюду, а тогда событие $\{\Lambda_{N,t_N} = \pm\infty\}$ включается в событие $\{|\zeta_{N,t_N})| > \varepsilon\}$

[1] Очевидно, в рассматриваемом сейчас случае множество $\{\mathbf{x} \in \mathbb{X}_N : p_{N,t}(\mathbf{x}) = 0\}$ также не зависит от t. Это множество имеет нулевую вероятность относительно как $P_{N,0}$, так и $P_{N,t}$, поэтому неопределенность в следующей ниже формуле, возникающая на этом множестве, не имеет значения.

[2] Здесь и далее, говоря об ограниченной последовательности $\{t_N\}$, мы предполагаем, что $t_N \in [-C, C] \subset \mathcal{T}$ для всех N.

[3] Если допустить, что Z_N определена с вероятностью $p_N < 1$, то функция распределения $P_{N,0}(Z_N < x) \to p_N$ при $x \to \infty$, где в силу условия (5.3.2) $p_N \to 1$. В таком случае положим, скажем, $Z_N = 0$ на дополнительном множестве, имеющем $P_{N,0}$-вероятность $1 - p_N \to 0$. При таком видоизменении Z_N определена и конечна $P_{N,0}$-почти всюду и, как нетрудно проверить, попрежнему удовлетворяет условиям (5.3.2) – (5.3.4).

и согласно (5.3.4) его $P_{N,0}$-вероятность стремится к нулю при $N \to \infty$. В следующем пункте доказывается контигуальность $P_{N,t_N} \lhd\rhd P_{N,0}$, вследствие которой $P_{N,t_N}(\Lambda_{N,t_N} = \pm\infty)$ также стремится к нулю.

Непосредственно из свойства ЛАН вытекает

Следствие 5.3.1. *Если $t_N \to t \in \mathcal{T}$, то*

$$\mathcal{L}(\Lambda_{N,t_N} \mid P_{N,0}) \to \mathcal{N}\Bigl(-\frac{1}{2}t^2 I, t^2 I\Bigr), \tag{5.3.5}$$

а $L_{N,t_N} \xrightarrow{d} e^{Y_t}$, где Y_t – случайная величина, имеющая нормальное распределение $\mathcal{N}\bigl(-\frac{1}{2}t^2 I, t^2 I\bigr)$.

5.3.1. ЛАН и контигуальность. Следующим важным следствием свойства ЛАН является то, что при его выполнении распределения $P_{N,0}$ и $P_{N,t}$ взаимно контигуальны.

Следствие 5.3.2. *Если выполнено свойство ЛАН, то $P_{N,t_N} \lhd\rhd P_{N,0}$ для любой ограниченной последовательности $t_N \in \mathcal{T}$.*

Доказательство. 1. Докажем, что $P_{N,t_N} \lhd P_{N,0}$. Для этого применим п. 4 теоремы 4.2.1, полагая в ней $P_N := P_{N,0}$, $Q_N := P_{N,t_N}$, $L_N := L_{N,t_N} = p_{N,t}(\mathbf{X})/p_{N,0}(\mathbf{X})$, $\Lambda_N := \Lambda_{N,t_N}$, если носитель плотности не зависит от t, и доопределяя L_N и Λ_N формулами (4.2.11) и (4.2.12) в противном случае. Нетрудно понять, что для подпоследовательности $\{m\} \subset \{N\}$ распределение L_{m,t_m} сходится к пределу, и соответственно последовательность ф.р. $F_m(x) = P_{m,0}(L_{m,t_m} < x)$ сходится к некоторой предельной ф.р. $F(x)$, только тогда, когда последовательность t_m сходится к пределу, скажем, $t_m \to t$. (Если $\{t_m\}$ не сходится, то из нее можно выбрать две подпоследовательности, сходящиеся к разным пределам, скажем, $t' \neq t''$, тогда последовательность $\mathcal{L}(L_{m,t_m} \mid P_{m,0})$ сходится по этим подпоследовательностям к разным пределам и, следовательно, не является сходящейся.) Тогда согласно следствию 5.3.1 $L_{m,t_m} \xrightarrow{d} e^{Y_t}$, где $Y_t \sim \mathcal{N}\bigl(-\frac{1}{2}t^2 I, t^2 I\bigr)$, а следовательно, $F(x) = P(e^{Y_t} < x)$. Таким образом, $\int t\,dF(t) = Ee^{Y_t}$ и нам нужно доказать, что $Ee^{Y_t} = 1$. Но нетрудно проверить прямыми преобразованиями нормальной плотности, что если $Y \overset{d}{=} \mathcal{N}(\mu, \sigma^2)$, то $Ee^Y = \exp(\mu + \sigma^2/2)$. В данном случае $\mu = -\frac{1}{2}t^2 I$, $\sigma^2 = t^2 I$, откуда следует требуемое равенство.

2. Докажем, что $P_{N,0} \lhd P_{N,t_N}$. Теперь мы применяем теорему 4.2.1, полагая $P_N := P_{N,t_N}$, $Q_N := P_{N,0}$. При этом по сравнению с предыдущим пунктом доказательства ОП «переворачивается» (числитель и знаменатель меняются местами), а $\Lambda_N = -\Lambda_{N,t}$. Согласно п. 3 этой теоремы нам надо проверить, что последовательность распределений $\mathcal{L}(\Lambda_{N,t} \mid P_{N,0})$ плотна (очевидно, что перемена знака при Λ_{N,t_N} не влияет на справедливость этого свойства). Предлагаем читателю самостоятельно вывести искомое утверждение из (5.3.2)–(5.3.4). □

Выведенная нами из свойства ЛАН контигуальность влечет выполнение ряда важных свойств. Первое из них вытекает из п. 1 теоремы 4.2.1: *остаточные члены в асимптотических соотношениях, сходящиеся к нулю по вероятности относительно* $P_{N,0}$, *сходятся к нулю по вероятности и относительно* P_{N,t_N} *для любой ограниченной последовательности* t_N.

Замечание 5.3.2. Обозначим $B_N = \{\mathbf{x}\colon p_{N,0}(\mathbf{x})p_{N,t_N}(\mathbf{x}) > 0\}$ (множество, где обе плотности положительны) и $D_N = \mathbf{X}_N \setminus B_N$. Из $P_{N,t_N} \lhd\rhd P_{N,0}$ следует, что $P_{N,0}(D_N) \to 0$ и $P_{N,t_N}(D_N) \to 0$. Введем условные распределения $\widetilde{P}_{N,0}$ и $\widetilde{P}_{N,t_N}$ аналогично (4.2.17). Тогда, повторяя доказательство формулы (4.2.20), получаем

$$\sup_{A\in\mathcal{A}} |\widetilde{P}_{N,0}(A) - P_{N,0}(A)| \to 0, \quad \sup_{A\in\mathcal{A}} |\widetilde{P}_{N,t_N}(A) - P_{N,t_N}(A)| \to 0 \tag{5.3.6}$$

при $N \to \infty$. Плотности $\tilde{p}_{N,0}$ и $\tilde{p}_{N,t_N}$ распределений $\widetilde{P}_{N,0}$ и $\widetilde{P}_{N,t_N}$ имеют общий носитель B_N. В последующих пунктах будут доказываться теоремы о сходимости распределений тех или иных статистик относительно P_{N,t_N}. В силу (5.3.6) для статистики, скажем, T_N, ф.р. $P_{N,t_N}(T_N < x)$ сходится к тому же пределу, что и ф.р. $\widetilde{P}_{N,t_N}(T_N < x)$. Это соображение позволит нам доказывать упомянутые предельные теоремы в предположении, что плотности распределений $P_{N,0}$ и P_{N,t_N} имеют общий носитель.

5.3.2. Распределение ОП при альтернативе. Рассмотрим еще раз пример с проверкой гипотезы о параметре сдвига нормального распределения. При проверке гипотезы $H_0\colon \mu = \mu_0$ против альтернативы $H_1\colon \mu = \mu_1 = \mu_0 + t/\sqrt{N}$ логарифм ОП выражается формулой (5.1.2), где Z_N дается формулой (5.1.1).

Непосредственно из этих формул видно, что

$$\mathcal{L}(\Lambda_N \mid P_{\mu_0}) = \mathcal{N}\left(-\frac{1}{2\sigma^2}t^2, \frac{1}{\sigma^2}t^2\right). \tag{5.3.7}$$

(Эта формула выведена также в п. 8.1.3, см. (8.1.12).) Из (5.1.2), (5.1.1) нетрудно вывести также распределение Λ_N при H_1. В отличие от предыдущего случая математическое ожидание каждого слагаемого $X_i - \mu_0$ в (5.1.1) равно $t/\sqrt{N}$, дисперсия же этого слагаемого не меняется и остается равной σ^2. В результате математическое ожидание первого слагаемого в правой части (5.1.2) становится равным t^2/σ^2 и мы получаем

$$\mathcal{L}(\Lambda_N \mid P_{\mu_1}) = \mathcal{N}\left(\frac{1}{2\sigma^2}t^2, \frac{1}{\sigma^2}t^2\right). \tag{5.3.8}$$

Посмотрим, как связаны нормальные распределения в правых частях (5.3.7) и (5.3.8). Нам будет несущественна конкретная структура математичеких ожиданий и дисперсий этих нормальных распределений, а будет важно только то, что они имеют вид $\mathcal{N}\big(-\frac{1}{2}b^2, b^2\big)$ и $\mathcal{N}\big(\frac{1}{2}b^2, b^2\big)$, $b > 0$. Простые выкладки показывают, что

$$\frac{\varphi\big(y; \frac{1}{2}b^2, b^2\big)}{\varphi\big(y; -\frac{1}{2}b^2, b^2\big)} = e^y, \tag{5.3.9}$$

где $\varphi(y; \mu, \sigma^2)$ обозначает плотность нормального распределения с указанными параметрами (см. (8.1.7)).

Покажем, что соотношения, аналогичные (5.3.7) и (5.3.8), выполняются в пределе, при $N \to \infty$, в общей ситуации настоящего параграфа. Аналог (5.3.7) был получен нами в следствии 5.3.1. Теперь мы получим аналог (5.3.8).

Следствие 5.3.3. *При условии ЛАН, если $t_N \to t \in \mathcal{T}$, то*

$$\mathcal{L}(\Lambda_{N,t_N} \mid P_{N,t_N}) \to \mathcal{N}\left(\frac{1}{2}t^2 I, t^2 I\right). \tag{5.3.10}$$

Доказательство. Используя описанную в замечании 5.3.2 возможность перехода от распределений $P_{N,0}$ и P_{N,t_N} к распределениям $\widetilde{P}_{N,0}$ и $\widetilde{P}_{N,t_N}$, плотности которых имеют общий носитель, будем предполагать, что сами распределения $P_{N,0}$ и P_{N,t_N} обладают этим свойством, т.е.

$$\{\mathbf{x} \in \mathbb{X}_N : p_{N,0}(\mathbf{x}) > 0\} = \{\mathbf{x} \in \mathbb{X}_N : p_{N,t_N}(\mathbf{x}) > 0\}.$$

В этом случае ОП L_{N,t_N} и логарифм ОП Λ_{N,t_N} определены формулой (5.3.1).

Обозначим $G_{N,t_N}(y)$ функцию распределения Λ_{N,t_N} относительно P_{N,t_N},

$$G_{N,t_N}(y) = P_{N,t_N}(\Lambda_{N,t_N} < y).$$

Функцию распределения и плотность $\mathcal{N}(\mu,\sigma^2)$ мы обозначаем $\Phi(y;\mu,\sigma^2)$ и $\varphi(y;\mu,\sigma^2)$ соответственно. Тогда искомое соотношение (5.3.10) можно переписать в виде

$$G_{N,t_N}(y) \to \Phi\left(y; \frac{1}{2}t^2 I, t^2 I\right). \tag{5.3.11}$$

Для простоты обозначений опустим индекс N при t. Легко видеть, что все рассуждения, проводимые как бы для фиксированного t, справедливы для последовательности $t_N \to t$.

Учитывая, что $p_{N,t}(\mathbf{x}) = \exp(\Lambda_{N,t}(\mathbf{x}))p_{N,0}(\mathbf{x})$ по определению (5.3.1), запишем $G_{N,t}(y)$ в виде

$$\begin{aligned} G_{N,t}(y) = P_{N,t}(\Lambda_{N,t} < y) &= \int_{\{\Lambda_{N,t}<y\}} p_{N,t}(\mathbf{x})\,d\mathbf{x} \\ &= \int_{\mathbb{R}^N} \mathbf{1}_{\{\Lambda_{N,t}<y\}} e^{\Lambda_{N,t}(\mathbf{x})} p_{N,0}(\mathbf{x})\,d\mathbf{x} \\ &= E_{N,0}\mathbf{1}_{\{\Lambda_{N,t}<y\}} e^{\Lambda_{N,t}}\,d\mathbf{x}. \end{aligned} \tag{5.3.12}$$

Обозначая $G_{N,0}(y) = P_{N,0}(\Lambda_{N,t} < y)$ (ф.р. $\Lambda_{N,t}$ относительно $P_{N,0}$), перепишем (5.3.5), аналогично (5.3.11), в виде

$$G_{N,0}(y) \to \Phi\left(y; -\frac{1}{2}t^2 I, t^2 I\right). \tag{5.3.13}$$

Математическое ожидание в правой части (5.3.12) равно

$$\int \mathbf{1}_{\{x<y\}} e^x\,dG_{N,0}(x). \tag{5.3.14}$$

Пользуясь сходимостью (5.3.13), перейдем формально к пределу под знаком интеграла:

$$\int \mathbf{1}_{\{x<y\}} e^x\,dG_{N,0}(x) \to \int \mathbf{1}_{\{x<y\}} e^x\,d\Phi\left(x; -\frac{1}{2}t^2 I, t^2 I\right). \tag{5.3.15}$$

Мы покажем, что из (5.3.15) следует (5.3.11) (или, эквивалентно, (5.3.10)), а затем проведем обоснование предельного перехода.

Помня, что левая часть (5.3.15) есть $G_{N,t}(y)$, нам нужно показать, что правая часть (5.3.15) равна $\Phi\bigl(x; \frac{1}{2}t^2 I, t^2 I\bigr)$. Пользуясь (5.3.9), имеем

$$\int \mathbf{1}_{\{x<y\}} e^x \, d\Phi\left(x; -\frac{1}{2}t^2 I, t^2 I\right) = \int_{-\infty}^{y} e^x \varphi\left(x; -\frac{1}{2}t^2 I, t^2 I\right) dx$$
$$= \int_{-\infty}^{y} \varphi\left(x; \frac{1}{2}t^2 I, t^2 I\right) dx = \Phi\left(x; \frac{1}{2}t^2 I, t^2 I\right). \tag{5.3.16}$$

Тем самым (5.3.15) влечет искомое соотношение (5.3.10) и нам остается обосновать законность предельного перехода в (5.3.15).

Для этого мы докажем, что функция e^x равномерно интегрируема относительно $G_{N,0}(x)$; функция $\mathbf{1}_{\{x<y\}} e^x \leqslant e^x$ тем более равномерно интегрируема, и сходимость интегралов тогда следует из п. 2 теоремы 8.2.5 Приложения.

По определению, нам надо доказать, что для любого $\varepsilon > 0$ найдется $A > 0$ такое, что

$$\int \mathbf{1}_{\{|x|>A\}} e^x \, dG_{N,0}(x) < \varepsilon.$$

Этот интеграл отличается от (5.3.14) только областью интегрирования. Переходя к представлению (5.3.12) и проделывая в обратном порядке те же преобразования, как в (5.3.12), получаем, что он равен $P_{N,t}(|\Lambda_{N,t}| > A)$, и мы видим, что требуемое утверждение равносильно условию плотности последовательности распределений $\{\mathcal{L}\,[\Lambda_{N,t} \mid P_{N,t}]\}$. Но в силу доказанной ранее контигуальности $P_{N,0} \lhd\rhd P_{N,t_N}$ (следствие 5.3.2) это свойство следует из п. 3 теоремы 4.2.1. Тем самым следствие 5.3.3 доказано. $\square$

5.3.3. Асимптотически наиболее мощные критерии. Рассмотрим теперь задачу проверки простой гипотезы $H_0\colon t = 0$ против простой альтернативы H_1, задаваемой каким-либо фиксированным значением $t \in \mathcal{T}$. Иначе говоря, мы проверяем гипотезу о том, что распределение наблюдаемого случайного элемента $\mathbf{X}$ пространства $\mathbb{X}_N$ есть $P_{N,0}$ против альтернативы, что это распределение есть $P_{N,t}$. По лемме Неймана–Пирсона (см. п. 8.1.2) наиболее мощный критерий в этой задаче состоит в том, чтобы отвергнуть H_0, если $\mathbf{X} \in S_{N,\alpha}$ (и принять H_0, если $\mathbf{X} \in \mathbb{X}_N \setminus S_{N,\alpha}$), где множество $S_{N,\alpha} \subset \mathbb{X}_N$ имеет вид:

$$S_{N,\alpha} = \{\mathbf{x}\colon \Lambda_{N,t} > c_{N,\alpha}\}, \tag{5.3.17}$$

причем константа $c_{N,\alpha}$ выбрана так, чтобы[4]

$$P_{N,0}(S_{N,\alpha}) = \alpha. \tag{5.3.18}$$

(Буквальное применение правила построения критической области $S_{N,\alpha}$, даваемого формулами (8.1.2), (8.1.3) п. 8.1.2, приводило бы к области вида

$$S_{N,\alpha} = \{\mathbf{x}\colon L_{N,t} > c_{N,\alpha}\} \tag{5.3.19}$$

с тем же условием (5.3.18). Тогда значение $c_{N,\alpha}$ отличалось бы от значения $c_{N,\alpha}$ в (5.3.17), но область $S_{N,\alpha}$ была бы той же, как в (5.3.17). Ср. замечание в п. 8.1.2.)

Замечание 5.3.3. В п. 8.1.2 предполагается, что распределения $P_{N,t}$ имеют носитель, не зависящий от t. В общем случае критическая область НМ критерия имеет вид

$$S_{N,\alpha} \cup \{p_{N,t}(\mathbf{x}) > 0, p_{N,0}(\mathbf{x}) = 0\},$$

где $S_{N,\alpha} \subset \{p_{N,t}(\mathbf{x})p_{N,0}(\mathbf{x}) > 0\}$ определяется формулой (5.3.19) с условием (5.3.18). Но $P_{N,t}\{p_{N,t}(\mathbf{x}) > 0, p_{N,0}(\mathbf{x}) = 0\} \to 0$ в силу контигуальности $P_{N,0}$ и $P_{N,t}$, поэтому это добавочное множество не влияет на предельную мощность критерия. Учитывая замечание 5.3.2, будем предполагать, что распределения $P_{N,t}$ при различных t взаимно абсолютно непрерывны (т.е. имеют носитель, не зависящий от t). Тогда НМ критерий имеет вид (5.3.17)–(5.3.18).

За редкими исключениями точное распределение ОП $L_{N,t}$ или его логарифма $\Lambda_{N,t}$ неизвестно и мы можем только пользоваться предельными теоремами, позволяющими находить приближенные критические константы $\tilde{c}_{N,\alpha}$ так, чтобы критическая область

$$\widetilde{S}_{N,\alpha} = \{\mathbf{x}\colon \Lambda_{N,t} > \tilde{c}_{N,\alpha}\} \tag{5.3.20}$$

удовлетворяла условию

$$P_{N,0}(\widetilde{S}_{N,\alpha}) \to \alpha \text{ при } N \to \infty. \tag{5.3.21}$$

[4]Следует иметь в виду, что $S_{N,\alpha}$ и $c_{N,\alpha}$, как и вводимая ниже константа $\tilde{c}_{N,\alpha}$, также зависят от t, чего мы не указываем явно, чтобы не усложнять обозначений.

А именно, воспользуемся следствием 5.3.1 и определим $\tilde{c}_{N,\alpha}$ так, как если бы формула (5.3.5) была точным равенством. Тогда мы имели бы

$$P_{N,0}(\Lambda_{N,t} < x) = \Phi\left(x; -\frac{1}{2}t^2 I, t^2 I\right),$$

нормированная величина $\left(\Lambda_{N,t}+\frac{1}{2}t^2 I\right)/t\sqrt{I}$ имела бы стандартное нормальное распределение $\mathcal{N}(0,1)$, и, определяя, как и раньше, константу z_α равенством $\Phi(z_\alpha) = 1-\alpha$, мы имели бы

$$P_{N,0}\left(\frac{\Lambda_{N,t}+\frac{1}{2}t^2 I}{t\sqrt{I}} > z_\alpha\right) = \alpha.$$

Полагая

$$\tilde{c}_{N,\alpha} = z_\alpha t\sqrt{I} - \frac{1}{2}t^2 I, \tag{5.3.22}$$

мы получили бы

$$P_{N,0}(\Lambda_{N,t} > \tilde{c}_{N,\alpha}) = \alpha.$$

Но в «реальной жизни» вместо точного равенства мы имеем сходимость

$$P_{N,0}(\Lambda_{N,t} < x) \to \Phi\left(x; -\frac{1}{2}t^2 I, t^2 I\right), \tag{5.3.23}$$

из которой следует, что, определяя $\tilde{c}_{N,\alpha}$ формулой (5.3.22), мы получим

$$P_{N,0}(\Lambda_{N,t} > \tilde{c}_{N,\alpha}) \to \alpha,$$

что эквивалентно соотношению (5.3.21). Кроме того, из (5.3.23) следует, что

$$c_{N,\alpha} \to \tilde{c}_{N,\alpha}, \tag{5.3.24}$$

где $c_{N,\alpha}$ – (неизвестное нам) точное решение уравнения

$$P_{N,0}(\Lambda_{N,t} > c_{N,\alpha}) = \alpha$$

(см. (5.3.17), (5.3.18)).

Теперь, обозначая через $\beta_N^*(t)$ мощность НМ критерия при альтернативе H_1,

$$\beta_N^*(t) = P_{N,t}(\Lambda_{N,t} > c_{N,\alpha}), \tag{5.3.25}$$

найдем ее предел при $N \to \infty$.

Заметим, что $\beta_N^*(t)$ как функция t не есть функция мощности какого-то одного критерия: при каждом t это мощность критерия, основанного на $\Lambda_{N,t}$ и имеющего критическую область $S_{N,\alpha}$ (см. (5.3.17)), зависящую, вообще говоря, от t. Эта критическая область не зависит от t в исключительных случаях, как, например, в задаче п. 8.1.3 проверки гипотезы о параметре сдвига нормального распределения. В таких случаях существует *равномерно наиболее мощный* критерий. В общем же случае мы построим его асимптотический аналог, так называемый *локально асимптотически наиболее мощный* критерий.

Согласно следствию 5.3.3

$$P_{N,t}(\Lambda_{N,t} < x) \to \Phi\left(x; \frac{1}{2}t^2 I, t^2 I\right). \tag{5.3.26}$$

Из этого соотношения и (5.3.24) следует, что

$$\beta_N^*(t) \to 1 - \Phi\left(\tilde{c}_{N,\alpha}; \frac{1}{2}t^2 I, t^2 I\right).$$

Подставляя сюда выражение для $\tilde{c}_{N,\alpha}$ (см. (5.3.22)) и пользуясь равенством

$$\Phi(x; \mu, \sigma^2) = \Phi\left(\frac{x-\mu}{\sigma^2}\right),$$

получим

$$\beta_N^*(t) \to \beta^*(t) := 1 - \Phi(z_\alpha - t\sqrt{I}) = \Phi(t\sqrt{I} - z_\alpha) \tag{5.3.27}$$

в силу симметрии нормального распределения.

Замечание 5.3.4. В некоторых случаях, когда альтернативы задаются иначе, чем в (5.2.4), например, задающий их набор малых параметров не разделяется на набор констант типа $c_{N,i}$ и общий множитель t, либо константы нормированы иначе, чем в (5.2.1), удается тем не менее построить аналог (5.3.3)–(5.3.4) вида

$$\Lambda_N = \mathcal{Z}_N - \frac{1}{2}A^2 + \zeta_N, \tag{5.3.28}$$

где $\mathcal{L}(\mathcal{Z}_N \mid P_{N,0}) \to \mathcal{N}(0, A^2)$ и $\zeta_N \xrightarrow{p} 0$ относительно $P_{N,0}$. В таком случае (ср. замечание 8.1.2) мощность НМ критерия сходится к пределу

$$\beta_N^* \to \beta^* := \Phi(A - z_\alpha), \tag{5.3.29}$$

т.е. величина A, определяющая предельную мощность, есть стандартное отклонение первого (случайного) члена разложения Λ_N (5.3.28).

Вспомним теперь наше исходное предположение о выполнении соотношений (5.3.2), (5.3.3), (5.3.4) (свойство ЛАН). Из (5.3.3) видно, что $\Lambda_{N,t}$ является линейной функцией статистики Z_N с точностью до остаточного члена, сходящегося к нулю по вероятности. Поэтому следует ожидать, что критерий, основанный на Z_N, будет иметь такую же предельную мощность, как $\beta_N^*(t)$. Действительно, из (5.3.2) следует, что

$$P_{N,0}(Z_N > z_\alpha \sqrt{I}) \to \alpha \ \text{ при } \ N \to \infty.$$

Таким образом, критерий, отвергающий H_0, когда $Z_N > z_\alpha \sqrt{I}$, имеет предельный уровень значимости α. Чтобы вычислить его предельную мощность, воспользуемся соотношениями (5.3.3), (5.3.4) и (5.3.26), из которых находим, что

$$P_{N,t}(Z_N < x) \to \Phi(x; tI, I).$$

Отсюда для мощности $\beta_{N,Z}(t)$ рассматриваемого критерия получаем

$$\begin{aligned} \beta_{N,Z}(t) = P_{N,t}(Z_N > z_\alpha \sqrt{I}) \to \\ \to 1 - \Phi(z_\alpha - t\sqrt{I}) = \Phi(t\sqrt{I} - z_\alpha) = \beta^*(t). \end{aligned} \tag{5.3.30}$$

Таким образом, при каждом $t > 0$ предел мощности нашего критерия равен пределу $\beta^*(t)$ мощности наиболее мощного критерия для данного t, т.е. пределу наибольшей достижимой мощности при данной альтернативе. Критерий, обладающий такими свойствами, называется *локально асимптотически наиболее мощным* (ЛАМН). (Слово «локально» в этом названии связано с тем, что переменная t описывает малые отклонения исходного параметра от значения, отвечающего H_0, при этом во всех предельных переходах t должно оставаться ограниченным. Как уже обсуждалось ранее, при таких t распределения $P_{N,0}$ и $P_{N,t}$ взаимно контигуальны.)

Критерий, основанный на статистике Z_N, является не единственным ЛАМН критерием. Любая статистика, которая при подходящей нормировке отличается от Z_N на величину, сходящуюся к нулю по вероятности $P_{N,0}$, порождает критерий с теми же свойствами. В дальнейшем мы построим ЛАМН ранговые критерии.

5.3.4. Третья лемма Ле Кама. В этом пункте мы установим результат, позволяющий из асимптотической нормальности какой-либо статистики относительно $P_{N,0}$ вывести ее асимптотическую нормальность относительно $P_{N,t}$. Этот результат (в более общей формулировке) был установлен в основополагающей статье Ле Кама [24]. Важнейшие результаты Ле Кама были подробно изложены в книге Гаека и Шидака [3] в виде «трех лемм Ле Кама»[5], из которых третья получила наибольшую известность ввиду ее широкой применимости.

Итак, в той же постановке, как в предыдущих пунктах, рассмотрим последовательность статистик $\{T_N\}$, о которой известно, что их распределения сходятся при гипотезе H_0 к нормальному $\mathcal{N}(0, \sigma^2)$ со средним 0 (для простоты) и дисперсией σ^2, т.е.

$$\mathcal{L}(T_N \mid P_{N,0}) \to \mathcal{N}(0, \sigma^2). \tag{5.3.31}$$

Наша цель – получить предельное распределение T_N при альтернативе, т.е. относительно распределений $P_{N,t}$. Заметим, что, вообще говоря, это отнюдь не простая задача, даже когда T_N представляет собой сумму независимых случайных величин. Дело в том, что распределения этих слагаемых меняются при изменении N и тем самым слагаемые составляют то, что в теории суммирования независимых случайных величин называется «схемой серий», когда при каждом N имеется свой набор случайных величин со своими распределениями. Общую теорию такого рода предельных теорем можно найти, например, в книге Лоэва [8]. Использование условия ЛАН позволяет получить желаемый результат очень просто (при дополнительном условии, формулируемом ниже). Технические трудности, связанные с использованием этого метода, сосредоточены в разделе, посвященном обоснованию условий ЛАН (гл. 7).

Упомянутое дополнительное условие состоит в том, что не только каждая из последовательностей T_N и $\Lambda_{N,t}$ асимптотически нормальна (см. (5.3.31) и (5.3.5)), но и их совместное распределение сходится к двумерному нормальному распределению:

$$\mathcal{L}[(\Lambda_{N,t}, T_N) \mid P_{N,0}] \to \mathcal{N}\left(-\frac{1}{2}t^2 I, 0, t^2 I, \sigma^2, \sigma_{1,2}\right), \tag{5.3.32}$$

[5] Название «леммы Ле Кама» дано по-видимому авторами книги, у самого Ле Кама соответствующие утверждения не назывались леммами.

где в качестве параметров предельного нормального распределения перечислены известные из (5.3.5) и (5.3.31) математические ожидания и дисперсии и указана некоторая ковариация $\sigma_{1,2}$. (Относительно сходимости распределений на плоскости см. п. 8.2.2.)

Условие (5.3.32) выполнено в том важном для нас случае, когда $\Lambda_{N,t}$ имеет вид (5.2.17) с Z_N, определяемым формулой (5.2.14), а T_N также представляет собой сумму некоторых функций от X_i. Этот более конкретный случай будет рассматриваться в § 6.1.

Лемма 5.3.1 (Третья лемма Ле Кама). *Если выполнено свойство ЛАН (5.3.2)–(5.3.4), то для произвольной последовательности $t_N \to t$, $t \in \mathcal{T}$, $t \neq 0$, из условия (5.3.32) с $t = t_N$ в левой части следует*

$$\mathcal{L}(T_N \mid P_{N,t_N}) \to \mathcal{N}(\sigma_{1,2}, \sigma^2). \tag{5.3.33}$$

Замечание 5.3.5. На самом деле имеет место более сильное утверждение, а именно, сходимость совместного распределения (T_N, Λ_{N,t_N}) к двумерному нормальному закону,

$$\mathcal{L}[(\Lambda_{N,t_N}, T_N) \mid P_{N,t_N}] \to \mathcal{N}\left(\frac{1}{2}t^2 I, \sigma_{1,2}, t^2 I, \sigma^2, \sigma_{1,2}\right). \tag{5.3.34}$$

Доказательство (5.3.33) несколько проще и это утверждение достаточно для последующих применений.

Доказательство. Будем предполагать, что коэффициент корреляции $\rho = \sigma_{1,2}/(t\sqrt{I}\sigma)$ отличен от ± 1, т.е. $|\rho| < 1$. В противном случае двумерное нормальное распределение в (5.3.32) (и (5.3.34)) вырождено (сосредоточено на прямой) и не имеет плотности распределения, которая будет использоваться при доказательстве. Случай $\rho = \pm 1$ возникает, когда T_N и $\Lambda_{N,t}$ линейно связаны с точностью до слагаемых, стремящихся к нулю по вероятности. В этом случае утверждение леммы следует из результатов предыдущего пункта.

Наша цель – доказать, что для любого $y \in \mathbb{R}$

$$P_{N,t_N}(T_N < y) \to \Phi(y; \sigma_{1,2}, \sigma^2). \tag{5.3.35}$$

Напомним, что $T_N = T_N(\mathbf{X})$, где $\mathbf{X}$ – случайная точка пространства $\mathbb{X}_N$, имеющая распределение $P_{N,0}$ с плотностью $p_{N,0}$, либо P_{N,t_N} с плотностью p_{N,t_N}. Пользуясь замечанием 5.3.2, будем предполагать, что плотности $p_{N,0}$ и p_{N,t_N} имеют общий носитель, т.е.

$$\{\mathbf{x}\colon p_{N,0} > 0\} = \{\mathbf{x}\colon p_{N,t_N} > 0\}.$$

Это множество имеет вероятность 1 относительно $P_{N,0}$ и P_{N,t_N}, поэтому мы можем отождествить с ним пространство $\mathbb{X}_N$. На этом множестве определены ОП $L_{N,t_N}(\mathbf{x}) = p_{N,t_N}(\mathbf{x})/p_{N,0}(\mathbf{x})$ и его логарифм $\Lambda_{N,t_N}(\mathbf{x}) = \log L_{N,t_N}$, и таким образом, $p_{N,t_N} = e^{\Lambda_{N,t_N}} p_{N,0}$.

Для простоты обозначений снова будем опускать индекс N при t (отметим, что вводимая ниже ф.р. $G_{N,0}(x,y)$ зависит от t_N и сходимость (5.3.41) обеспечивается условием леммы о выполнении (5.3.32) с $t = t_N$).

Запишем рассматриваемую функцию распределения в виде

$$\begin{aligned} P_{N,t}(T_N < y) = &\int_{\mathbb{R}^N} \mathbf{1}_{\{T_N<y\}}(\mathbf{x}) p_{N,t}(\mathbf{x})\,d\mathbf{x} \\ = &\int_{\mathbb{R}^N} \mathbf{1}_{\{T_N<y\}}(\mathbf{x}) e^{\Lambda_{N,t}(\mathbf{x})} p_{N,0}(\mathbf{x})\,d\mathbf{x} \\ = &E_{N,0}\mathbf{1}_{\{T_N<y\}} e^{\Lambda_{N,t}}. \end{aligned} \tag{5.3.36}$$

Обозначим $G_{N,0}$ совместное распределение $(\Lambda_{N,t}, T_N)$, фигурирующее в левой части (5.3.32) (т.е. $G_{N,0}(A) = P_{N,0}((\Lambda_{N,t}, T_N) \in A)$, $A \in \mathcal{B}(\mathbb{R}^2)$). Как и в п. 8.2.2 Приложения, обозначим $A(x,y) = \{(x_1,x_2)\colon x_1 < x, x_2 < y\}$ и введем функцию распределения $G_{N,0}(x,y) = G_{N,0}(A(x,y))$, т.е.

$$G_{N,0}(x,y) = P_{N,0}(\Lambda_{N,t} < x, T_N < y).$$

Покажем, что функцию распределения (5.3.36) можно представить в виде

$$E_{N,0}\mathbf{1}_{\{T_N<y\}} e^{\Lambda_{N,t}} = \int_{-\infty}^{\infty} e^x\, G_{N,0}(dx,y), \tag{5.3.37}$$

где запись $G_{N,0}(dx,y)$ обозначает, что $G_{N,0}(x,y)$ рассматривается при фиксированном y как функция аргумента x, по которому и производится интегрирование. Тем самым $G_{N,0}(x,y)$ как функция x образует несобственную функцию распределения, изменяющуюся от $G(-\infty,y) = 0$ до $G(\infty,y) = P_{N,0}(T_N < y)$.

Введем условную функцию распределения

$$G_{N,0}(x \mid y) = G_{N,0}(\Lambda_{N,t} < x \mid T_N < y) = \frac{G_{N,0}(x,y)}{P_{N,0}(T_N < y)}. \qquad (5.3.38)$$

По формуле полного математического ожидания

$$E_{N,0}\mathbf{1}_{\{T_N<y\}}e^{\Lambda_{N,t}} = P_{N,0}(T_N < y)E_{N,0}(\mathbf{1}_{\{T_N<y\}}e^{\Lambda_{N,t}} \mid T_N < y)$$

(аналогичное слагаемое с условием $T_N \geqslant y$ обращается в нуль, поскольку при этом условии $\mathbf{1}_{\{T_N<y\}} = 0$). В последнем выражении множитель $\mathbf{1}_{\{T_N<y\}}$ под знаком математического ожидания можно опустить, так как при условии $T_N < y$ он тождественно равен 1. По формуле (5.3.38) тогда получаем (5.3.37), откуда с учетом (5.3.36) следует

$$P_{N,t}(T_N < y) = \int_{-\infty}^{\infty} e^x\, G_{N,0}(dx,y). \qquad (5.3.39)$$

Обозначим Φ_0 и $\varphi_0(x_1,x_2)$ нормальное распределение в правой части (5.3.32) и его плотность. Введем также двумерную функцию распределения $\Phi_0(x,y) = \Phi_0(A(x,y))$. Имеем

$$\Phi_0(x,y) = \int_{-\infty}^{x}\int_{-\infty}^{y} \varphi_0(s,t)\, dsdt. \qquad (5.3.40)$$

По определению 8.2.2 (см. также лемму 8.2.2) условие (5.3.32) означает, что

$$G_{N,0}(x,y) \to \Phi_0(x,y) \qquad (5.3.41)$$

для каждой точки $(x,y) \in \mathbb{R}^2$. Из п. 2 теоремы 8.2.5 тогда следует, что

$$\int_{-\infty}^{\infty} e^x\, G_{N,0}(dx,y) \to \int_{-\infty}^{\infty} e^x\, \Phi_0(dx,y). \qquad (5.3.42)$$

Равномерная интегрируемость обосновывается так же, как при доказательстве леммы 5.3.3. В силу (5.3.39) остается показать, что

$$\int_{-\infty}^{\infty} e^x\, \Phi_0(dx,y) = \Phi(y;\sigma_{1,2},\sigma^2). \qquad (5.3.43)$$

Напомним, что мы обозначили $\varphi_0(x,y) = \varphi\big(x,y;-\frac{1}{2}t^2I,0,t^2I,$ $\sigma^2,\sigma_{1,2}\big)$, а $\Phi_0(x,y)$ выражается через плотность $\varphi_0(x,y)$ в виде интеграла (5.3.40). Далее, Φ_0 как функция x при фиксированном y имеет плотность, получаемую дифференцированием этого выражения по x и равную

$$\int_{-\infty}^{y} \varphi_0(x,t)\,dt.$$

Таким образом, интеграл в левой части (5.3.43) можно переписать в виде

$$\int_{-\infty}^{y} \left(\int_{-\infty}^{\infty} e^{x} \varphi_0(x,t)\,dx \right) dt. \tag{5.3.44}$$

Покажем, что

$$e^{x}\varphi_0(x,y) = \varphi_1(x,y) := \varphi\Big(x, y; \frac{1}{2}t^2 I, \sigma_{1,2}, t^2 I, \sigma^2, \sigma_{1,2}\Big). \tag{5.3.45}$$

Тогда интеграл (5.3.44) равен

$$\int_{-\infty}^{y} \left(\int_{-\infty}^{\infty} \varphi\Big(x, t; \frac{1}{2}t^2 I, \sigma_{1,2}, t^2 I, \sigma^2, \sigma_{1,2}\Big)\,dx \right) dt,$$

внутренний интеграл равен частной плотности $\varphi(t; \sigma_{1,2}, \sigma^2)$, интеграл от которой от $-\infty$ до y есть функция распределения $\Phi(y; \sigma_{1,2}, \sigma^2)$. Этим (5.3.43) будет доказано, что и завершит доказательство леммы.

Конечно, доказать (5.3.45) можно путем непосредственных преобразований двумерной нормальной плотности, умноженной на e^x. Мы, однако, проведем доказательство (5.3.45), использующее некоторые известные формулы и не требующее практически никаких выкладок.

Введем случайный вектор (Λ_t, T), имеющий плотность распределения $\varphi_0(x,y)$. Обозначим через $\varphi_{T|\Lambda_t}(y \mid x)$ условную плотность распределения T при фиксированном $\Lambda_t = x$ и через $\varphi_0(x) = \varphi\big(x; -\frac{1}{2}t^2 I, t^2 I\big)$ плотность Λ_t. Тогда

$$\varphi_0(x,y) = \varphi_{T|\Lambda_t}(y \mid x)\varphi_0(x). \tag{5.3.46}$$

Мы видели раньше, что $e^x\varphi_0(x) = \varphi_1(x) := \varphi\big(x; \frac{1}{2}t^2 I, t^2 I\big)$. Следовательно,

$$e^x\varphi_0(x,y) = \varphi_{T|\Lambda_t}(y \mid x)\varphi_1(x). \tag{5.3.47}$$

Заметим, что $\varphi_1(x)$ есть частная плотность распределения Λ_t, если вектор (Λ_t, T) имеет плотность распределения $\varphi_1(x,y)$. Покажем, что в этом случае условная плотность T при фиксированном $\Lambda_t = x$, равная

$$\varphi_{1,T|\Lambda_t}(y \mid x) = \frac{\varphi_1(x,y)}{\varphi_1(x)},$$

совпадает с условной плотностью $\varphi_{T|\Lambda_t}(y \mid x)$, отвечающей случаю, когда (Λ_t, T) имеет плотность $\varphi_0(x,y)$, т.е.

$$\varphi_{1,T|\Lambda_t}(y \mid x) = \varphi_{T|\Lambda_t}(y \mid x).$$

Тогда, заменяя в (5.3.47) $\varphi_{T|\Lambda_t}$ на $\varphi_{1,T|\Lambda_t}$, получаем, что произведение в правой части (5.3.47) равно $\varphi_1(x,y)$, что и доказывает (5.3.45).

Для доказательства утверждения о равенстве условных плотностей проверим, что параметры условных распределений $(T \mid \Lambda_t = x)$ относительно $\varphi_0(x,y)$ и $\varphi_1(x,y)$ совпадают.

Приведем известные формулы для параметров условного нормального распределения. Если случайный вектор (X,Y) имеет двумерное нормальное распределение со средними μ_1, μ_2, дисперсиями σ_1^2, σ_2^2 и коэффициентом корреляции ρ, то условное распределение Y при фиксированном $X = x$ нормально со средним

$$\mu = \mu_2 + \frac{\sigma_{1,2}}{\sigma_1^2}(x - \mu_1) \quad \text{и дисперсией} \quad \sigma_2\sqrt{1-\rho^2}$$

(где ковариация $\sigma_{1,2} = \rho\sigma_1\sigma_2$). Эти выражения можно найти, например, в учебнике [1], пример 6.5. Плотность φ_0 имеет параметры $\mu_1 = -\frac{1}{2}t^2 I$, $\mu_2 = 0$, $\sigma_1^2 = t^2 I$, $\sigma_2^2 = \sigma^2$ и $\sigma_{1,2}$, принятое в наших обозначениях, а плотность φ_1 – параметры $\mu_1' = \frac{1}{2}t^2 I$, $\mu_2' = \sigma_{1,2}$ и те же σ_1^2, σ_2^2 и $\sigma_{1,2}$. Последнее означает, что дисперсии условных распределений одинаковы. Выпишем математические ожидания условных распределений, обозначая их μ^0 и μ^1 соответственно. Имеем

$$\mu^0 = \frac{\sigma_{1,2}}{t^2 I}\left(x + \frac{1}{2}t^2 I\right), \quad \mu^1 = \sigma_{1,2} + \frac{\sigma_{1,2}}{t^2 I}\left(x - \frac{1}{2}t^2 I\right).$$

Видно, что $\mu^0 = \mu^1$, откуда следует заявленное равенство условных распределений, что завершает доказательство леммы. $\square$

Глава 6. Асимптотическая мощность и эффективность «параметрических» и ранговых критериев

§ 6.1. Свойства «параметрических» критериев при локальных альтернативах

Имея в виду использование аппроксимации ранговых статистик суммами независимых случайных величин, введенной в § 3.2, мы рассмотрим в данном параграфе асимптотические свойства критериев, основанных на суммах вида (3.2.11), при локальных альтернативах. Эти результаты будут применены к ранговым критериям в следующем параграфе.

В настоящей главе мы предполагаем, что плотности $p_\theta(x)$, $\theta \in \Theta$, удовлетворяют условиям теоремы 5.2.2, обеспечивающей выполнение свойства ЛАН (5.3.2)–(5.3.4) применительно к рассматриваемой здесь постановке задачи (см. п. 5.2.1). Как следствие, в этой постановке справедливы результаты, полученные в § 5.3.

6.1.1. Проверка простой гипотезы $\boldsymbol{H_0}$. Мы рассматриваем постановку задачи, описанную в п. 5.2.1. А именно, задано семейство распределений на прямой $\mathbb{R}$, имеющих плотности $p_\theta(x)$, $\theta \in \Theta$, по мере Лебега, где Θ – интервал, содержащий точку $\theta = 0$. Наблюдается вектор $\mathbf{X} = (X_i)_{i=1}^N$ независимых случайных величин, имеющих соответственно плотности $p_{\theta_i}(x)$, $i = 1, \ldots, N$. При каждом N задан набор констант $c_{N,1}, \ldots, c_{N,N}$, удовлетворяющих условиям (5.2.1) и (5.2.2), и проверяется гипотеза H_0 (5.2.3) против альтернативы H_1 (5.2.4), которые для удобства напомним здесь:

$$H_0 \colon \theta_i = 0, \qquad i = 1, \ldots, N, \tag{6.1.1}$$

$$H_1 \colon \theta_i = c_{N,i} t, \quad i = 1, \ldots, N, \quad t > 0. \tag{6.1.2}$$

Как и раньше, пишем $p(x) = p_0(x)$ и обозначаем $P_{N,0}$ и $P_{N,t}$ распределения вектора $\mathbf{X}$ при гипотезах H_0 и H_1 соответственно[1].

Замечания 5.3.2 и 5.3.3 позволяют без потери общности предполагать, что распределения $P_{N,t}$ имеют носитель, не зависящий от t, что эквивалентно предположению, что множество $\{p_\theta(x) > 0\}$ не зависит от $\theta \in \Theta$ (условие (5.2.6)). Тогда формулами (5.2.5)–(5.2.8) определены плотность $p_{N,t}$ распределения $P_{N,t}$, ОП $L_{N,t}$ и его логарифм $\Lambda_{N,t}$. Как уже говорилось, мы предполагаем выполненными условия теоремы 5.2.2, согласно которой имеет место свойство ЛАН. Для удобства сформулируем здесь это свойство. По теореме 5.2.2 (совместно с теоремой 5.2.1) логарифм ОП Λ_{N,t_N} для любой ограниченной последовательности $t_N \in \mathbb{R}$ допускает представление

$$\Lambda_{N,t_N} = t_N Z_N - \frac{1}{2} t_N^2 I + \zeta_{N,t_N}, \quad \zeta_{N,t_N} \xrightarrow{p} 0 \text{ при } N \to \infty, \tag{6.1.3}$$

где (см. (5.2.14) и (5.2.12))

$$Z_N = \sum_1^N c_{N,i} l^{(1)}(X_i), \quad I = E_0 (l^{(1)})^2 < \infty, \tag{6.1.4}$$

$l^{(1)}(x) = l_0^{(1)}(x)$, $l_\theta^{(1)}(x) = \partial l_\theta(x)/\partial\theta$, $l_\theta(x) = \log p_\theta(x)$. При этом $El^{(1)} = 0$ и

$$\mathcal{L}(Z_N \mid P_{N,0}) \to \mathcal{N}(0, I), \quad \mathcal{L}(\Lambda_{N,t_N} \mid P_{N,0}) \to \mathcal{N}\Big(-\frac{1}{2} t^2 I, t^2 I\Big). \tag{6.1.5}$$

Тем самым выполнено свойство ЛАН (5.3.2)–(5.3.4), сформулированное в более общей постановке в начале § 5.3, и как следствие применительно к рассматриваемой нами схеме справедливы результаты, полученные в § 5.3.

В частности, при данном $t > 0$ НМ критерий для проверки H_0 (6.1.1) против альтернативы H_1 (6.1.2) отвергает H_0, когда $\mathbf{X} = (X_1, \ldots, X_N) \in S_{N,\alpha}(t)$, где

$$S_{N,\alpha}(t) = \{\mathbf{x}\colon \Lambda_{N,t} > c_{N,\alpha}\}, \text{ и } c_{N,\alpha} \text{ таково, что } P_{N,0}(S_{N,\alpha}(t)) = \alpha \tag{6.1.6}$$

[1]Точка $\theta = 0$ в качестве гипотезы H_0 выбрана для простоты записи. Точно так же может проверяться гипотеза H_0 относительно произвольной точки $\theta_0 \in \Theta$. В этом случае гипотеза (6.1.1) и альтернатива (6.1.2) переходят в $H_0\colon \theta_i = \theta_0$, и $H_1\colon \theta_i = \theta_0 + c_{N,i}t$, $i = 1, \ldots, N$. При этом, разумеется, в формулировках результатов происходят соответствующие изменения, например, в формуле (6.1.4) $l^{(1)}(x) = l_0^{(1)}(x)$ заменяется на $l_{\theta_0}^{(1)}(x)$, а I – на $I(\theta_0) = E_{\theta_0}(l_{\theta_0}^{(1)}(X))^2$, и т.п.

(см. (5.3.17), (5.3.18)). Мощность этого НМ критерия для данного $t > 0$ мы обозначали $\beta_N^*(t)$,

$$\beta_N^*(t) = P_{N,t}(S_{N,\alpha}(t)) \tag{6.1.7}$$

(см. (5.3.25)). В п. 5.3.3 было показано, что

$$\beta_N^*(t) \to \beta^*(t) = \Phi(t\sqrt{I} - z_\alpha). \tag{6.1.8}$$

Кроме того, было показано, что к тому же пределу сходится мощность критерия, основанного на статистике Z_N (см. (5.3.30)), который тем самым является локально асимптотически наиболее мощным (ЛАМН) критерием (см. обсуждение этого термина после формулы (5.3.30)). Было отмечено также, что любая статистика, которая при соответствущей нормировке отличается от Z_N на величину, стремящуюся к нулю по вероятности, тоже определяет ЛАМН критерий.

Как уже отмечалось, $\beta_N^*(t)$ не есть функция мощности какого-либо критерия – при каждом $t > 0$ это мощность «своего» НМ критерия для соответствующей альтернативы H_1 (6.1.2), и эти критерии различны для различных t (за исключением случая, когда существует РНМ критерий, как, например, в п. 8.1.3). Функцию $\beta_N^*(t)$ называют *огибающей функцией мощности.* Чтобы пояснить это название, построим для фиксированного $t_1 > 0$ критерий вида (6.1.6) с критической областью $S_{N,\alpha}(t_1)$. Обозначим $\beta_{N,t_1}(t)$ его функцию мощности, т.е. мощность при альтернативах (6.1.2) как функцию $t > 0$. Рассмотрим теперь семейство критериев указанного вида, индексированное параметром $t_1 > 0$, и соответствующее семейство функций мощности $\beta_{N,t_1}(t)$. При фиксированном $t > 0$ максимальное значение этих функций мощности по всевозможным $t_1 > 0$ достигается при $t_1 = t$, поскольку $S_{N,\alpha}(t)$ есть критическая область НМ критерия против альтернативы t. Таким образом, это максимальное значение равно $\beta_{N,t}(t) = \beta_N^*(t)$ (сравните (6.1.7) и (6.1.9)). Иначе говоря, при каждом $t_1 > 0$ график функции $\beta_{N,t_1}(t)$ лежит под кривой $\beta_N^*(t)$, соприкасаясь с ней в точке $t = t_1$. Этим объясняется приведенное выше название (более педантичным было бы *огибающая функций мощности*).

Укажем попутно, что при каждых $t_1 > 0$ и $t > 0$

$$\beta_{N,t_1}(t) = P_{N,t}(S_{N,\alpha}(t_1)) \to \beta^*(t) \tag{6.1.9}$$

(докажите самостоятельно), т.е. для каждого $t_1 > 0$ критерий с критической областью $S_{N,\alpha}(t_1)$ является ЛАНМ критерием. Таким образом, мы получили пример семейства ЛАНМ критериев.

Перейдем к рассмотрению сумм независимых случайных величин, аппроксимирующих ранговые статистики, и изучению мощности основанных на них критериев. Напомним, что формула (3.2.11) определяет сумму вида

$$T_N = \sum_{i=1}^{N} c_{N,i}\varphi(U_i). \tag{6.1.10}$$

Здесь мы будем, однако, рассматривать несколько видоизмененные суммы. А именно, вместо функций $\varphi(U_i)$ от равномерно распределенных случайных величин U_i будем рассматривать функции $g(X_i)$ от наших наблюдений X_i, $i = 1, \ldots, N$, где g – некоторая функция, удовлетворяющая условиям

$$E_0 g(X_1) = \int g(x)p(x)\,dx = 0, \tag{6.1.11}$$

$$\sigma_g^2 := E_0 g^2(X_1) = \int g^2(x)p(x)\,dx < \infty \tag{6.1.12}$$

(аналог условия (3.1.5)). Переход к величинам U_i будет рассмотрен в следующем параграфе, когда результаты настоящего параграфа будут применяться к ранговым критериям. Еще одно отличие будет введено в следующем пункте. Таким образом, мы рассматриваем статистики

$$T_N = \sum_{i=1}^{N} c_{N,i} g(X_i) \tag{6.1.13}$$

в качестве статистик критериев проверки простой гипотезы H_0 против альтернативы H_1.

Мы подчеркиваем, что в данном пункте рассматривается задача проверки *простой* гипотезы H_0, потому что к этой задаче, вообще говоря, неприменимы ранговые статистики рассматриваемого вида (ПЛРС). Например, как уже отмечалось, частным случаем задачи проверки гипотезы (6.1.1) против альтернативы (6.1.2), получаемым при $c_{N,1} = \cdots = c_{N,N}$, является задача проверки гипотезы о значении параметра θ по выборке из независимых одинаково распределенных наблюдений (как при гипотезе, так и при альтернативе). Статистики же вида (1.1.2) как раз и предназначены для проверки *сложной* гипотезы о том, что наблюдения X_i одинаково распределены, против тех или иных альтернатив разнораспределенности. Различие между двумя задачами аналогично различию между задачами, обсуждаемыми в пп. 8.1.3 и 8.1.4. Рассмотрения настоящего пункта, представляющие и самостоятельный интерес, будут использованы ниже при переходе к задачам проверки сложной гипотезы, в которых будут применяться ранговые критерии.

Нашей ближайшей целью будет получить предельные распределения статистики T_N (6.1.13) при гипотезе H_0 (6.1.1) и альтернативе H_1 (6.1.2), которые позволят нам выписать асимптотическую мощность критерия, основанного на T_N.

Учитывая условия (5.2.1) и (5.2.2), в силу (6.1.12) имеем $E_0 T_N = 0$, $\operatorname{var}_0 T_N = \sigma_g^2$, и по теореме 8.3.1 получаем

$$\mathcal{L}(T_N \mid P_{N,0}) \to \mathcal{N}\big(0, \sigma_g^2\big) \text{ при } N \to \infty. \tag{6.1.14}$$

Для получения предельного распределения T_N при альтернативе воспользуемся третьей леммой Ле Кама (лемма 5.3.1). Чтобы применить эту лемму, нам требуется предельное совместное распределение вектора $(\Lambda_{N,t}, T_N)$.

Соотношения (6.1.5) показывают, что предельные распределения $\Lambda_{N,t}$ и Z_N связаны между собой так, как если бы имела место линейная зависимость $\Lambda_{N,t} = tZ_N - \frac{1}{2}t^2 I$. Нам будет удобнее найти предельное распределение (Z_N, T_N), из которого, пользуясь указанным соображением, легко получается предельное распределение $(\Lambda_{N,t}, T_N)$.

Обозначим

$$\sigma(l^{(1)}, g) = E_0(l^{(1)} g) = \int l^{(1)}(x) g(x) p(x)\, dx. \tag{6.1.15}$$

Лемма 6.1.1. *Пусть Z_N и T_N определены равенствами (6.1.4) и (6.1.13) с константами $c_{N,1}, \ldots, c_{N,N}$, удовлетворяющими условиям (5.2.1) и (5.2.2), и функцией g, удовлетворяющей условиям (6.1.11) и (6.1.12). Тогда*

$$\mathcal{L}[(Z_N, T_N) \mid P_{N,0}] \to \mathcal{N}\big(0, 0, I, \sigma_g^2, \sigma(l^{(1)}, g)\big). \tag{6.1.16}$$

Доказательство. Воспользуемся теоремой 8.2.3, согласно которой для доказательства леммы достаточно показать, что для любого $t = (t_1, t_2) \in \mathbb{R}^2$

$$\mathcal{L}(t_1 Z_N + t_2 T_N \mid P_{N,0}) \to \mathcal{L}(t_1 Z + t_2 T),$$

где (Z, T) – случайный вектор, имеющий двумерное нормальное распределение, стоящее в правой части (6.1.16). При $(t_1, t_2) = (0, 0)$ в обеих частях этого соотношения стоят вырожденные распределения, сосредоточенные в точке 0, так что оно представляет собой тривиальное равенство. Поэтому в дальнейшем будем предполагать, что $(t_1, t_2) \neq (0, 0)$.

Как известно, всякая линейная комбинация компонент двумерного нормального вектора распределена нормально. Найдем параметры распределения $t_1 Z + t_2 T$. Согласно (6.1.16)

$$E(t_1 Z + t_2 T) = 0, \tag{6.1.17}$$

$$\operatorname{var}(t_1 Z + t_2 T) = E(t_1 Z + t_2 T)^2 = t_1^2 I + t_2^2 \sigma_g^2 + 2 t_1 t_2 \sigma(l^{(1)}, g). \tag{6.1.18}$$

Таким образом, нам нужно доказать, что распределение $t_1 Z_N + t_2 T_N$ сходится к нормальному со средним 0 и дисперсией (6.1.18).

Согласно (6.1.13) и (6.1.4)

$$t_1 Z_N + t_2 T_N = \sum_1^N c_{N,i} (t_1 l^{(1)}(X_i) + t_2 g(X_i)). \tag{6.1.19}$$

Применим теорему 8.3.1, полагая $V_i = t_1 l^{(1)}(X_i) + t_2 g(X_i)$. Условия этой теоремы на коэффициенты $c_{N,i}$ выполнены в силу предположений (5.2.1) и (5.2.2). Легко подсчитать, что дисперсия V_i равна правой части (6.1.18). Следовательно, требуемое утверждение вытекает из теоремы 8.3.1. $\square$

Пользуясь соотношением (6.1.3), из леммы 6.1.1 выводим

Следствие 6.1.1. *Предположим, что выполнено условие ЛАН* (6.1.3), (6.1.4), *выполнены условия* (5.2.1) *и* (5.2.2), *а* T_N *определено равенством* (6.1.13) *с условиями* (6.1.11) *и* (6.1.12). *Тогда*

$$\mathcal{L}[(\Lambda_{N,t}, T_N) \mid P_{N,0}] \to \mathcal{N}\left(-\frac{1}{2} t^2 I, 0, t^2 I, \sigma_g^2, t\sigma(l^{(1)}, g)\right). \tag{6.1.20}$$

Таким образом, выполнено соотношение (5.3.32) с $\sigma^2 = \sigma_g^2$ и $\sigma_{1,2} = t\sigma(l^{(1)}, g)$.

Следствие 6.1.2. *При условиях следствия* 6.1.1

$$\mathcal{L}(T_N \mid P_{N,t}) \to \mathcal{N}\big(t\sigma(l^{(1)}, g), \sigma_g^2\big). \tag{6.1.21}$$

В силу (6.1.20) это следствие непосредственно вытекает из леммы 5.3.1.

Пользуясь соотношениями (6.1.14) и (6.1.21), выпишем предельную мощность критерия, основанного на T_N. Определим z_α уравнением $\Phi(z_\alpha) = 1-\alpha$, где $\Phi(x)$ – функция распределения стандартного нормального закона $\mathcal{N}(0,1)$. Тогда из (6.1.14) следует, что

$$P_{N,0}(T_N > \sigma_g z_\alpha) \to \alpha, \quad N \to \infty.$$

Таким образом, критерий отвергающий H_0, когда $T_N > \sigma_g z_\alpha$, имеет в пределе уровень значимости α. Обозначим $\beta_{N,T}(t) = P_{N,t}(T_N > \sigma_g z_\alpha)$ мощность этого критерия. Согласно (6.1.21) имеем

$$\beta_{N,T}(t) \to \beta_T(t) := 1-\Phi\left(\frac{\sigma_g z_\alpha - t\sigma(l^{(1)},g)}{\sigma_g}\right) = \Phi\left(t\frac{\sigma(l^{(1)},g)}{\sigma_g} - z_\alpha\right). \tag{6.1.22}$$

Сравним эту предельную мощность с предельной мощностью $\beta^*(t)$ наиболее мощного (НМ) критерия, основанного на $\Lambda_{N,t}$, см. (5.3.27) (или, что то же, с предельной мощностью локально асимптотически наиболее мощного (ЛАМН) критерия, основанного на Z_N, см. (5.3.30)). Согласно (5.3.27)

$$\beta^*(t) = \Phi(t\sqrt{I} - z_\alpha). \tag{6.1.23}$$

Поскольку I, σ_g^2 и $\sigma(l^{(1)},g)$ – дисперсии и ковариация случайных величин $l^{(1)}(X_1)$ и $g(X_1)$, то, как известно[2],

$$\rho(l^{(1)},g) := \frac{\sigma(l^{(1)},g)}{\sqrt{I}\sigma_g} \leqslant 1, \tag{6.1.24}$$

причем равенство достигается только если $g(x) = al^{(1)}(x)$, $a>0$. (Вообще из равенства 1 коэффициента корреляции величин X и Y следует, что они связаны линейной зависимостью вида $Y = aX+b$, но в нашем случае свободный член $b=0$, поскольку $E_0 g = E_0 l^{(1)} = 0$.)

Из (6.1.13) и (6.1.4) видно, что статистика Z_N ЛАМН критерия является частным случаем статистики T_N при $g(x) = l^{(1)}(x)$.

[2]Следующее далее неравенство, вообще говоря, имеет вид $|\rho(l^{(1)},g)| \leqslant 1$, но случай $\rho(l^{(1)},g) < 0$, т.е. $\sigma(l^{(1)},g) < 0$, не представляет интереса с точки зрения статистики: он происходит от неудачного выбора знака функции g, и перемена знака увеличивает мощность критерия.

Отношение коэффициентов при t в (6.1.22) и (6.1.23), равное $\rho(l^{(1)}, g)$, характеризует уменьшение мощности, вызванное применением статистики критерия с «неоптимальной» (отличной от $l^{(1)}$) функцией g.

В заключение этого раздела отметим, что многие статистики критериев не являются в точности суммами вида T_N, а допускают представление, в котором такого рода сумма служит главным членом, а остаток стремится к нулю по вероятности. Благодаря свойству контигуальности это стремление к нулю имеет место как при гипотезе, так и при альтернативе, и для таких критериев справедливы полученные выше выражения для мощности и их следствия.

6.1.2. Асимптотическая эффективность по Питмену. Как уже отмечалось, частным случаем рассматриваемой задачи проверки гипотез является задача проверки гипотезы о значении параметра распределения по простой выборке, т.е. по выборке, состоящей из независимых одинаково распределенных наблюдений (как при гипотезе, так и при альтернативе). Эта задача получается при $c_{N,i} = 1/\sqrt{N}$, $i = 1, \ldots, N$. Эти константы очевидным образом удовлетворяют условиям (5.2.1) и (5.2.2), и критерии, основанные на статистиках

$$T_N = \frac{1}{\sqrt{N}} \sum_1^N g(X_i) \quad \text{и} \quad Z_N = \frac{1}{\sqrt{N}} \sum_1^N l^{(1)}(X_i), \tag{6.1.25}$$

имеют асимптотические мощности $\beta_T(t)$ и $\beta^*(t)$, даваемые формулами (6.1.22) и (6.1.23). Напомним (см. (6.1.2)), что в этом случае все X_i, $i = 1, \ldots, N$, имеют при гипотезе H_0 плотность $p(x, 0)$, а при альтернативе H_1 плотность $p(x, t/\sqrt{N})$.

Зададим теперь некоторое ω, $0 < \omega < 1 - \alpha$, и посмотрим, при каких t (а следовательно, при каких $\theta = t/\sqrt{N}$) та и другая мощность достигают уровня $1 - \omega$ (т.е. вероятность ошибки 2-го рода равна ω). Для этого решим уравнения

$$\beta_T(t) = 1 - \omega \quad \text{и} \quad \beta^*(t) = 1 - \omega,$$

и эти решения будут представлять асимптотику искомых значений t при $N \to \infty$ (для краткости изложения мы не проводим детального обоснования наших рассуждений). Пользуясь выражениями (6.1.22) и (6.1.23), получаем соответственно

$$t_g \to (z_\alpha + z_\omega) \frac{\sigma_g}{\sigma(l^{(1)}, g)}, \quad t^* \to \frac{z_\alpha + z_\omega}{\sqrt{I}}. \tag{6.1.26}$$

В терминах параметра $\theta = t/\sqrt{N}$ соотношения (6.1.26) можно переписать как

$$\theta_g \sim \frac{(z_\alpha + z_\omega)}{\sqrt{N}} \frac{\sigma_g}{\sigma(l^{(1)}, g)}, \quad \theta^* \sim \frac{z_\alpha + z_\omega}{\sqrt{NI}}. \tag{6.1.27}$$

Теперь поставим вопрос: какой объем выборки требуется, чтобы отличить «близкую» альтернативу H_1 от гипотезы H_0 с заданными вероятностями ошибок 1-го и 2-го рода? «Близкая» альтернатива означает, что рассматривается асимптотика при $\theta \to 0$ и вопрос состоит в том, при каких объемах выборок, зависящих от θ, мощность того и другого критерия достигает уровня $1 - \omega$. Обозначим $N_g(\theta)$ и $N^*(\theta)$ искомые объемы выборок для критериев, основанных на статистиках T_N и Z_N. Тогда из (6.1.27) получаем асимптотические соотношения

$$\theta \sim \frac{(z_\alpha + z_\omega)}{\sqrt{N_g(\theta)}} \frac{\sigma_g}{\sigma(l^{(1)}, g)}, \quad \theta \sim \frac{z_\alpha + z_\omega}{\sqrt{N^*(\theta) I}},$$

откуда следует, что при $\theta \to 0$

$$N_g(\theta) \sim \frac{(z_\alpha + z_\omega)^2}{\theta^2} \frac{\sigma_g^2}{\sigma^2(l^{(1)}, g)}, \quad N^*(\theta) \sim \frac{(z_\alpha + z_\omega)^2}{\theta^2 I}. \tag{6.1.28}$$

Предел отношения необходимых объемов выборок $N^*(\theta)/N_g(\theta)$ при $\theta \to 0$ называется *асимптотической эффективностью по Питмену* (или *питменовской асимптотической эффективностью*) критерия, основанного на статистике T_N. Из (6.1.28) следует, что эта асимптотическая эффективность равна

$$e_g := \lim_{\theta \to 0} \frac{N^*(\theta)}{N_g(\theta)} = \frac{\sigma^2(l^{(1)}, g)}{I \sigma_g^2} = \rho^2(l^{(1)}, g) \tag{6.1.29}$$

(см. (6.1.24)). Полученный ответ имеет простой наглядный смысл: понятно, что ЛАНМ критерий, основанный на Z_N, требует (асимптотически) меньше (во всяком случае не больше) наблюдений для достижения той же мощности, чем критерий, основанный на T_N, и качество последнего, измеряемое отношением необходимых объемов выборок, тем выше, чем ближе функция g к $l^{(1)}$ в смысле близости их коэффициента корреляции к 1.

Формула (6.1.28) позволяет также находить питменовскую *асимптотическую относительную эффективность* (АОЭ) двух критериев, основанных на статистиках вида T_N, скажем,

$$T_N(g) = \frac{1}{\sqrt{N}} \sum_1^N g(X_i) \quad \text{и} \quad T_N(f) = \frac{1}{\sqrt{N}} \sum_1^N f(X_i).$$

Из первой формулы в (6.1.28) сразу получаем, что АОЭ критерия, основанного на $T_N(f)$ по отношению к критерию, основанному на $T_N(g)$, равна

$$e_{f,g} := \lim_{\theta \to 0} \frac{N_g(\theta)}{N_f(\theta)} = \frac{\sigma_g^2 \sigma^2(l^{(1)}, f)}{\sigma^2(l^{(1)}, g)\sigma_f^2} = \frac{\rho^2(l^{(1)}, f)}{\rho^2(l^{(1)}, g)}.$$

Приведем пример построения ЛАНМ критерия и вычисления асимптотической эффективности конкурирующего критерия.

Пример 6.1.1. Рассмотрим распределение Лапласа (или, иначе, симметрично-экспоненциальное распределение) с плотностью $p(x) = \frac{1}{2}e^{-|x|}$ и семейство сдвигов этого распределения с плотностями

$$p_\theta(x) = p(x - \theta) = \frac{1}{2}e^{-|x-\theta|}, \quad \theta \in \mathbb{R}. \tag{6.1.30}$$

Для этого семейства рассмотрим задачу проверки гипотезы H_0: $\theta = 0$ против альтернативы $H_1\colon \theta > 0$ по простой выборке $X_1, \ldots, X_N$. Нетрудно подсчитать, что

$$l^{(1)}(x) = \frac{\partial}{\partial \theta} \log p_\theta(x)|_{\theta=0} = -\frac{p'(x)}{p(x)} = \operatorname{sign} x \tag{6.1.31}$$

(производная в точке $x = 0$ не существует и мы доопределяем $l^{(1)}(x)$ в этой точке произвольным образом, полагая ее равной нулю). Статистика Z_N ЛАМН критерия в этом случае имеет вид (см. (6.1.25))

$$Z_N = \frac{1}{\sqrt{N}} \sum_1^N \operatorname{sign}(X_i) = \frac{\#\{X_i > 0\} - \#\{X_i < 0\}}{\sqrt{N}}, \tag{6.1.32}$$

где $\#\{\ldots\}$ обозначает число элементов данного множества. Фишеровская информация равна

$$I = E_0(l^{(1)}(x))^2 = \int (\operatorname{sign} x)^2 p(x)\, dx = 1. \tag{6.1.33}$$

Следовательно, $Z_N \xrightarrow{d} \mathcal{N}(0,1)$ при гипотезе H_0 и критерий, отвергающий H_0 при $Z_N > z_\alpha$, имеет при альтернативе $\theta = t/\sqrt{N}$ асимптотическую мощность (см. (6.1.23))

$$\beta^*(t) = \Phi(t - z_\alpha). \tag{6.1.34}$$

Посмотрим, как в этой задаче «работает» критерий, являющийся НМ критерием, когда $p(x)$ – плотность нормального распределения, а именно критерий, основанный на статистике $\overline{X} = \sum_1^N X_i/N$, или, эквивалентно, на

$$T_N = \frac{1}{\sqrt{N}} \sum_1^N X_i. \tag{6.1.35}$$

Эта статистика имеет вид T_N в (6.1.25) с $g(x) = x$. Мощность этого критерия дается формулой (6.1.22), в которой

$$\sigma(l^{(1)}, g) = \int x \operatorname{sign}(x) p(x)\, dx = \int_0^\infty x e^{-x}\, dx = 1,$$
$$\sigma_g^2 = \int_0^\infty x^2 e^{-x}\, dx = 2.$$

Следовательно, асимптотическая мощность согласно (6.1.22) равна

$$\beta_T(t) = \Phi\left(t\frac{\sigma(l^{(1)}, g)}{\sigma_g} - z_\alpha\right) = \Phi\left(\frac{t}{\sqrt{2}} - z_\alpha\right),$$

а асимптотическая эффективность критерия, основанного на $\overline{X}$, по отношению к асимптотически НМ критерию знаков, основанному на Z_N (6.1.32), равна (см. (6.1.29))

$$\rho^2(l^{(1)}, g) = \frac{\sigma^2(l^{(1)}, g)}{I\sigma_g^2} = \frac{1}{2}. \tag{6.1.36}$$

Если бы (как обычно бывает в практических задачах) распределение зависело еще и от параметра масштаба, т.е. имело плотность

$$\frac{1}{\sigma} p\left(\frac{x-\theta}{\sigma}\right),$$

то статистика Z_N (6.1.32) была бы применима без всяких изменений, а вместо критерия, основанного на $\overline{X}$ или T_N (6.1.35), использовался критерий Стьюдента, который имел бы ту же асимптотическую эффективность $\frac{1}{2}$. Формула (6.1.36) показывает, что «на чужом поле» критерий Стьюдента может «работать» с весьма низкой эффективностью.

В п. 6.2.1 мы сравним этот результат с асимптотической эффективностью критерия нормальных меток (см. п. 3.3.3), который, как и критерий Стьюдента, асимптотически эффективен в нормальном случае.

6.1.3. Проверка простой гипотезы $\boldsymbol{H_0}$ (продолжение). Коэффициенты $c_{N,i}$ при $g(X_i)$ в статистике T_N (6.1.13) совпадают с константами, определяющими альтернативу H_1 (см. (6.1.2)). В этом пункте мы рассмотрим суммы T_N, в которые $g(X_i)$ входят с коэффициентами, отличными от $c_{N,i}$. А именно, рассмотрим последовательность сумм

$$T_N^d = \sum_{i=1}^{N} d_{N,i} g(X_i) \tag{6.1.37}$$

с коэффициентами $d_{N,i}$, удовлетворяющими условиям

$$\sum_{i=1}^{N} d_{N,i}^2 = 1 \quad \text{и} \quad \max_i |d_{N,i}| \to 0 \text{ при } N \to \infty \tag{6.1.38}$$

(аналог условий (5.2.1) и (5.2.2)). Статистику T_N предыдущего пункта будем теперь обозначать T_N^c.

Целью такого видоизменения является выяснить, как влияет выбор коэффициентов на свойства критерия. Для этого мы, как и в предыдущем пункте, найдем предельные распределения T_N^d (6.1.37) при H_0 и H_1 и асимптотическую мощность критерия, основанного на этой статистике.

Для простоты рассуждений будем дополнительно предполагать, что существует предел

$$\tau = \lim_{N \to \infty} \tau_N, \quad \text{где} \quad \tau_N := \sum_{i=1}^{N} c_{N,i} d_{N,i}. \tag{6.1.39}$$

(Здесь $c_{N,i}$ – константы, входящие в (6.1.2).) Рассматривая наборы коэффициентов $\mathbf{c}_N = (c_{N,1}, \ldots, c_{N,N})$ и $\mathbf{d}_N = (d_{N,1}, \ldots, d_{N,N})$ как векторы в евклидовом пространстве $\mathbb{R}^N$, можно представить условия (5.2.1) и (6.1.38) как $\|\mathbf{c}_N\| = \|\mathbf{d}_N\| = 1$, откуда следует, что скалярное произведение $\tau_N = (\mathbf{c}_N, \mathbf{d}_N)$ не превосходит 1 по абсолютной величине, $|\tau_N| \leqslant 1$.

Заметим сразу, что в случае $|\tau| = 1$ статистики T_N^c и T_N^d асимптотически эквивалентны. А именно, если $\tau = 1$, то при гипотезе H_0 аналогично выводу (6.1.14) получаем

$$\mathrm{var}_0(T_N^c - T_N^d) = E_0(T_N^c - T_N^d)^2 = \sigma_g^2(2 - 2\tau_N) \to 0,$$

следовательно, $T_N^c - T_N^d \xrightarrow{p} 0$ и T_N^d имеет то же предельное распределение, что и T_N^c (см. (6.1.14)). В силу контигуальности $T_N^c - T_N^d \xrightarrow{p} 0$ и при альтернативе H_1, поэтому T_N^c и T_N^d имеют при H_1 также одно и то же предельное распределение (см. (6.1.14)) и как следствие, соответствующие критерии имеют одну и ту же мощность (6.1.22). В случае $\tau = -1$ сказанное выше справедливо в отношении статистики, получаемой при замене $d_{N,i}$ на $-d_{N,i}$ в T_N^d. Таким образом, случай $|\tau| = 1$ сводится к рассмотренному в предыдущем пункте, и в дальнейшем будем предполагать, что $|\tau| < 1$.

В этом случае статистики T_N^c и T_N^d не сближаются между собой, тем не менее соображения, приводящие к сходимости (6.1.14), справедливы и для T_N^d (с использованием условий (6.1.38) вместо (5.2.1) и (5.2.2)), поэтому

$$\mathcal{L}(T_N^d \mid P_{N,0}) \to \mathcal{N}(0, \sigma_g^2) \text{ при } N \to \infty. \tag{6.1.40}$$

Следуя схеме предыдущего пункта, найдем сначала предел совместного распределения (Z_N, T_N^d) при гипотезе H_0. Следующая лемма является аналогом леммы 6.1.1. Напомним, что $\sigma(l^{(1)}, g)$ определено формулой (6.1.15).

Лемма 6.1.2. *Пусть Z_N и T_N^d определены равенствами (6.1.4) и (6.1.37) с функцией g, удовлетворяющей условиям (6.1.11) и (6.1.12), и коэффициентами $d_{N,1}, \dots, d_{N,N}$, удовлетворяющими условиям (6.1.38) и (6.1.39). Тогда*

$$\mathcal{L}[(Z_N, T_N^d) \mid P_{N,0}] \to \mathcal{N}\big(0, 0, I, \sigma_g^2, \tau\sigma(l^{(1)}, g)\big). \tag{6.1.41}$$

Доказательство использует тот же прием Крамера–Уолда (теорема 8.2.3), что и доказательство леммы 6.1.1. Повторяя почти дословно начало этого доказательства и обозначая (Z, T^d) аналог вектора (Z, T), получаем, что для любого $t = (t_1, t_2) \in \mathbb{R}^2$ нужно доказать сходимость

$$\mathcal{L}(t_1 Z_N + t_2 T_N^d \mid P_{N,0}) \to \mathcal{L}(t_1 Z + t_2 T^d). \tag{6.1.42}$$

Как и в доказательстве леммы 6.1.1, предполагаем, что $(t_1, t_2) \neq (0, 0)$.

Распределение величины $t_1 Z + t_2 T^d$ – нормальное с параметрами

$$E(t_1 Z + t_2 T^d) = 0, \tag{6.1.43}$$

$$\operatorname{var}(t_1 Z + t_2 T^d) = \sigma^2(t_1, t_2) := t_1^2 I + t_2^2 \sigma_g^2 + 2t_1 t_2 \tau \sigma(l^{(1)}, g). \tag{6.1.44}$$

Таким образом, требуется доказать, что распределение $t_1 Z_N + t_2 T_N^d$ сходится к $\mathcal{N}(0, \sigma^2(t_1, t_2))$.

Дальнейшее доказательство отличается от доказательства леммы 6.1.1. Аналогом (6.1.19) теперь служит

$$t_1 Z_N + t_2 T_N^d = \sum_1^N (c_{N,i} t_1 l^{(1)}(X_i) + d_{Ni} t_2 g(X_i)), \tag{6.1.45}$$

и к сумме в (6.1.45) неприменима теорема 8.3.1. Поэтому для доказательства сходимости (6.1.42) мы используем непосредственно теорему Линдеберга–Феллера (теорема 8.3.2). Нетрудно подсчитать, что

$$E(t_1 Z_N + t_2 T_N^d) = 0, \tag{6.1.46}$$

$$\operatorname{var}(t_1 Z_N + t_2 T_N^d) = \sigma_N^2(t_1, t_2) := t_1^2 I + t_2^2 \sigma_g^2 + 2t_1 t_2 \tau_N \sigma(l^{(1)}, g). \tag{6.1.47}$$

По предположению (6.1.39) $\tau_N \to \tau$, следовательно, $\sigma_N^2(t_1, t_2) \to \sigma^2(t_1, t_2)$, и нам достаточно доказать, что

$$\frac{t_1 Z_N + t_2 T_N^d}{\sigma_N(t_1, t_2)} \xrightarrow{d} \mathcal{N}(0, 1) \tag{6.1.48}$$

при гипотезе H_0.

Если уж мы делим на $\sigma_N(t_1, t_2)$, нужно проверить, что $\sigma_N(t_1, t_2) \neq 0$. Покажем, что $\sigma(t_1, t_2) > 0$ при любых $(t_1, t_2) \neq (0, 0)$ (тогда и $\sigma_N(t_1, t_2) > 0$ при всех достаточно больших N). Возьмем произвольную точку $(t_1, t_2) \neq (0, 0)$. Хотя бы одна из ее компонент t_1, t_2 отлична от нуля, пусть для определенности $t_1 \neq 0$. Покажем, что $\min_t \sigma^2(t_1, t) > 0$, откуда будет следовать требуемое утверждение. Учитывая выражение (6.1.44), имеем

$$\frac{d}{dt} \sigma^2(t_1, t) = 2t\sigma_g^2 + 2t_1 \tau \sigma(l^{(1)}, g).$$

Приравнивая производную к нулю, находим

$$t_{\min} = -t_1 \frac{\tau\sigma(l^{(1)}, g)}{\sigma_g^2}$$

и, подставляя $t_2 = t_{\min}$ в выражение для $\sigma(t_1, t_2)$, получаем

$$\begin{aligned}\min_t \sigma^2(t_1, t) &= t_1^2 I + t_1^2 \frac{\tau^2\sigma^2(l^{(1)}, g)}{\sigma_g^2} - 2t_1^2 \frac{\tau^2\sigma^2(l^{(1)}, g)}{\sigma_g^2} \\ &= t_1^2\biggl(I - \frac{\tau^2\sigma^2(l^{(1)}, g)}{\sigma_g^2}\biggr).\end{aligned}$$

Учитывая, что $\sigma^2(l^{(1)}, g) \leqslant I\sigma_g^2$ и, по предположению, $\tau^2 < 1$, заключаем отсюда, что $\min_t \sigma^2(t_1, t) > 0$ и, следовательно, $\sigma(t_1, t_2) > 0$ при любых $(t_1, t_2) \neq (0, 0)$.

Положим (см. (6.1.45))

$$Y_{Ni} = \frac{c_{N,i} t_1 l^{(1)}(X_i) + d_{Ni} t_2 g(X_i)}{\sigma_N(t_1, t_2)}.$$

Тогда левая часть (6.1.48) равна $\sum_i Y_{Ni}$, причем $EY_{Ni} = 0$, $i = 1, \ldots, N$, и $\sum_i EY_{Ni}^2 = 1$. Тем самым Y_{Ni} удовлетворяют условиям теоремы 8.3.2, и для доказательства (6.1.48) нам нужно проверить условие

$$\sum_{i=1}^{N} EY_{N,i}^2 \mathbf{1}_{\{|Y_{N,i}| > \varepsilon\}} \to 0 \ \text{ при } \ N \to \infty. \tag{6.1.49}$$

Прежде всего, очевидно, что знаменатель $\sigma_N(t_1, t_2)$, сходящийся к постоянной $\sigma(t_1, t_2) > 0$, не влияет на сходимость (6.1.49) и вместо $Y_{N,i}$ в этом соотношении можно рассматривать числители $c_{N,i} t_1 l^{(1)}(X_i) + d_{Ni} t_2 g(X_i)$. Воспользуемся элементарным неравенством $(a + b)^2 \leqslant 2(a^2 + b^2)$. Тогда задача сводится к проверке условия

$$\begin{aligned}\sum_{i=1}^{N} E\bigl[(c_{N,i} t_1 l^{(1)}(X_i))^2 + (d_{Ni} t_2 g(X_i))^2\bigr] \times \\ \times \mathbf{1}_{\{(c_{N,i} t_1 l^{(1)}(X_i))^2 + (d_{Ni} t_2 g(X_i))^2 > \varepsilon^2/2\}} \to 0.\end{aligned}$$

Обозначим $C_N = \max[\max_i c_{N,i}^2, \max_i d_{Ni}^2]$. Из условий (5.2.2) и (6.1.38) следует, что $C_N \to 0$ при $N \to \infty$. Далее, положим

$V_i = ((t_1 l^{(1)}(X_i))^2 + (t_2 g(X_i))^2)^{1/2}$. Тогда индикатор в последнем соотношении не превосходит $\mathbf{1}_{\{V_i^2 > \varepsilon^2/2C_N^2\}}$ и аналогично переходу от (8.3.5) к (8.3.6) в доказательстве теоремы 8.3.1 левая часть этого соотношения не превосходит

$$EV_1^2 \mathbf{1}_{\{|V_1| > \varepsilon/\sqrt{2}C_N\}}.$$

Сходимость к нулю этого выражения доказывается теми же рассуждениями, как в доказательстве теоремы 8.3.1. □

Как и в п. 6.1.1, из соотношения (6.1.3) и леммы 6.1.2 вытекает

Следствие 6.1.3. *Предположим, что выполнено условие ЛАН* (6.1.3), (6.1.4), *выполнены условия* (5.2.1) *и* (5.2.2), *а* T_N^d *определено равенством* (6.1.37) *с условиями* (6.1.11), (6.1.12) *и* (6.1.38). *Тогда*

$$\mathcal{L}[(\Lambda_{N,t}, T_N^d) \mid P_{N,0}] \to \mathcal{N}\Big(-\frac{1}{2}t^2 I, 0, t^2 I, \sigma_g^2, t\tau\sigma(l^{(1)}, g)\Big). \quad (6.1.50)$$

Теперь из леммы 5.3.1 получаем

Следствие 6.1.4. *При условиях следствия* 6.1.3

$$\mathcal{L}(T_N^d \mid P_{N,t}) \to \mathcal{N}\big(t\tau\sigma(l^{(1)}, g), \sigma_g^2\big). \quad (6.1.51)$$

Аналогично выводу (6.1.22) обозначим $\beta_{N,T}^d(t) = P_{N,t}(T_N^d > \sigma_g z_\alpha)$ мощность критерия, основанного на T_N^d. Согласно (6.1.51) имеем

$$\beta_{N,T}^d(t) \to \beta_T^d(t) := \Phi\Big(t\frac{\tau\sigma(l^{(1)}, g)}{\sigma_g} - z_\alpha\Big). \quad (6.1.52)$$

§ 6.2. Асимптотическая мощность и эффективность ранговых критериев

Ранговые статистики[3] применяются в тех случаях, когда распределение наблюдений точно не известно, и применение «параметрических» критериев, рассчитанных на какую-либо конкретную форму этого распределения, может приводить к ошибочным

[3] Как уже говорилось, в данном курсе рассматриваются только простые линейные ранговые статистики (ПЛРС) вида (1.1.2), которые мы зачастую называем ранговыми статистиками без дополнительных оговорок.

заключениям. Так, свойство свободы от распределения (см. § 2.2) позволяет строить ранговые критерии, имеющие заданный уровень значимости, не зная ничего о распределении наблюдений при гипотезе H_0 (кроме его непрерывности). Тем не менее представляет интерес выяснение мощности ранговых критериев при альтернативах, определяемых теми или иными конкретными распределениями. Такого рода результаты позволяют, в частности, сравнить свойства «параметрических» и ранговых критериев.

В п. 6.2.1 для заданного семейства распределений и заданного рангового критерия находится асимптотическая мощность этого критерия при локальных альтернативах. Этот результат позволяет, в частности, вычислять асимптотическую относительную эффективность ранговых критериев между собой и по отношению к «параметрическим» критериям.

В пп. 6.2.2 и 6.2.3 рассматриваются примеры применения этих результатов к конкретным ранговым критериям.

В п. 6.2.4 для заданного семейства распределений строится асимптотически наиболее мощный (АНМ) ранговый критерий, который имеет асимптотическую эффективность 1 по отношению к АНМ «параметрическому» критерию. В частности, будет показано, что если наши наблюдения распределены нормально, то критерий нормальных меток (см. п. 3.3.3) имеет ту же асимптотическую мощность, что и критерий Стьюдента. С другой стороны, при любом отклонении распределения от нормальности критерий нормальных меток более эффективен, чем критерий Стьюдента[4]. Поэтому, если предположение о нормальности распределений не имеет под собой достаточно веских оснований, то применение рангового критерия предпочтительнее.

Для простоты изложения мы ограничимся семействами распределений, зависящими от параметра сдвига.

6.2.1. Асимптотическая мощность ранговых критериев. Для заданной плотности $p(x)$, $x \in \mathbb{R}$, будем рассматривать семейство распределений на прямой, порожденное сдвигами этой плотности, т.е. семейство распределений с плотностями $p_\theta(x) = p(x-\theta)$, $\theta \in \mathbb{R}$. Как и раньше, имеется выборка $X_1, \ldots, X_N$

[4]Известный в литературе результат, который будет приведен без доказательства в п. 6.2.5.

с распределениями из нашего семейства, $X_i \sim p_{\theta_i}(x)$. Рассматривается задача проверки гипотезы об одинаковой распределенности элементов выборки, т.е. гипотезы

$$H_0\colon \theta_1 = \cdots = \theta_N = \theta, \quad \theta \in \mathbb{R}. \tag{6.2.1}$$

Это сложная гипотеза, поскольку она оставляет произвольным общее значение параметра θ. Эта гипотеза проверяется против (сложной) альтернативы

$$H_1\colon \theta_i = \theta + tc_{N,i}, \quad \theta \in \mathbb{R}, \quad t > 0, \tag{6.2.2}$$

где θ – неизвестный параметр, а $(c_{N,i})_{i=1}^N$, как и раньше, – заданный набор регрессионных констант, удовлетворяющих условиям (5.2.1) и (5.2.2).

Наша цель – исследовать асимптотическое поведение ранговой статистики вида

$$S_N = \sum_{i=1}^N c_{N,i} a_N(R_{N,i}) \tag{6.2.3}$$

(см. (3.1.1)) при альтернативе H_1.

Обозначим $F(x)$ функцию распределения, соответствующую плотности $p(x)$, и $F_\theta(x) = F(x-\theta)$ – функцию распределения, соответствующую $p_\theta(x) = p(x-\theta)$. Заметим, что если $X \sim F(x)$, то $X + \theta \sim F_\theta(x)$ (где знак $\sim$ указывает на соответствие между случайной величиной и ее функцией распределения). Поэтому мы можем представлять, что при гипотезе H_0 наши наблюдения имеют вид $X_1+\theta, \ldots, X_N+\theta$, $\theta \in \mathbb{R}$, где $X_1, \ldots, X_N$ имеют «стандартную» ф.р. $F(x)$. Добавление одного и того же слагаемого θ к каждому элементу выборки не меняет взаимного расположения элементов выборки и следовательно, их рангов, а следовательно и значения самой статистики S_N. Тем самым распределение S_N при гипотезе H_0 не зависит от θ (частный случай свойства свободы от распределения). Точно так же при гипотезе H_1 распределение S_N не зависит от θ (считая фиксированными t и набор констант $(c_{N,i})_{i=1}^N$). Поэтому мы можем изучать распределение S_N при удобных нам простых подгипотезах гипотез H_0 и H_1. В качестве таких простых подгипотез мы возьмем гипотезы (6.1.1) и (6.1.2), которые, чтобы отличать их от сложных гипотез H_0 и H_1, обозначим H_0' и H_1':

$$H_0'\colon \theta_i = 0, \qquad i = 1, \ldots, N, \tag{6.2.4}$$

$$H_1'\colon \theta_i = \tilde{c}_{N,i} t, \quad i = 1, \ldots, N, \quad t > 0, \tag{6.2.5}$$

где обозначение $\tilde{c}_{N,i} = c_{N,i} - \bar{c}_N$, $\bar{c}_N = \sum c_{N,i}/N$, было введено в (3.2.10) и неоднократно употреблялось в дальнейшем.

Кроме того, мы можем центрировать и нормировать коэффициенты $c_{N,i}$ в ранговых статистиках (6.2.3). Из структуры ПЛРС (6.2.3) видно, что умножение коэффициентов $c_{N,i}$ на одно и то же число приводит к умножению всей статистики на это же число, а переход от $c_{N,i}$ к, скажем, $c_{N,i} + b$ добавляет к S_N неслучайное слагаемое $b\sum_{j=1}^{N} a_N(j)$. Таким образом, нормирование и центрирование коэффициентов ранговой статистики приводит к статистике, определяющей тот же самый критерий, который как следствие имеет ту же самую мощность.

Поэтому мы сразу возьмем и в качестве коэффициентов ранговой статистики и в качестве констант, задающих альтернативу, центрированные и нормированные коэффициенты, которые снова будем обозначать $c_{N,i}$, получаемые из $\tilde{c}_{N,i}$ делением на корень из суммы их квадратов, т.е. положим

$$c_{N,i} := \frac{c_{N,i} - \bar{c}}{\sqrt{\sum_i (c_{N,i} - \bar{c})^2}}. \tag{6.2.6}$$

Так переопределенные коэффициенты $c_{N,i}$ удовлетворяют условиям

$$\sum_{i=1}^{N} c_{N,i} = 0, \quad \sum_{i=1}^{N} c_{N,i}^2 = 1 \tag{6.2.7}$$

и при таких коэффициентах $c_{N,i}$ альтернатива (6.2.5) принимает вид

$$H_1'\colon \theta_i = c_{N,i} t, \quad i = 1, \ldots, N, \quad t > 0. \tag{6.2.8}$$

Будем также считать, что константы $c_{N,i}$ (6.2.6) входят в определение сложной альтернативы (6.2.2). Как уже отмечалось, мощность критерия, основанного на S_N, в задаче (6.2.1)–(6.2.2) равна его мощности в задаче (6.2.4)–(6.2.8).

Асимптотическое распределение S_N при гипотезе H_0' дается теоремой 3.2.1, согласно которой

$$\frac{S_N - \mu_N}{\sigma_N} \xrightarrow{d} \mathcal{N}(0,1), \tag{6.2.9}$$

где μ_N и σ_N^2 даются формулой (3.2.6). Эта теорема является следствием теоремы 3.2.2 об аппроксимации статистики S_N суммой независимых случайных величин. Эта аппроксимация в силу контигуальности остается в силе при альтернативе H_1, и мы используем ее для получения соответствующего предельного распределения статистики S_N и асимптотической мощности основанного на ней критерия.

Приведем формулировку теоремы 3.2.2 и связанные с ней условия и обозначения.

Предполагается, что метки $a_N(j)$ связаны с некоторой функцией $\varphi(u)$, $u \in (0,1)$, либо равенством (3.1.6), т.е. $a_N(j) = a_N^{\varphi}(j)$, либо, при выполнении соответствующих условий на функцию $\varphi(u)$, одним из равенств (3.1.7) или (3.2.8) (см. замечание 3.2.2).

Относительно функции $\varphi(u)$ будем предполагать, что

$$\bar{\varphi} = \int_0^1 \varphi(u)\,du = 0 \quad \text{и} \quad 0 < \sigma_{\varphi}^2 := \int_0^1 \varphi^2(u)\,du < \infty \tag{6.2.10}$$

(в теореме 3.2.2 не предполагалось, что $\bar{\varphi} = 0$, но переход от φ к $\varphi - \bar{\varphi}$ не изменяет статистик S_N (6.2.3) и T_N (см. (6.2.13) ниже) и упрощает обозначения; второе условие соответствует условиям (3.2.3) и (3.2.4)).

На константы $\{c_{N,i}\}$ накладывается условие (3.2.5), которое в силу (6.2.7) формулируется как

$$\max_i |c_{N,i}| \to 0 \ \text{ при } \ N \to \infty. \tag{6.2.11}$$

Учитывая (6.2.7), имеем $\bar{c}_N = 0$, и (3.2.6) принимает вид

$$ES_N = \mu_N = 0, \quad \sigma_N^2 = \sigma_{\varphi}^2. \tag{6.2.12}$$

По теореме 3.2.2 статистика S_N аппроксимируется суммой независимых случайных величин (см. (3.2.11))

$$T_N = \sum_{i=1}^{N} c_{N,i}\varphi(U_i), \tag{6.2.13}$$

где $U_1, \ldots, U_N$ – независимые случайные величины, распределенные равномерно на отрезке $[0,1]$, а $R_{N,1}, \ldots, R_{N,N}$, входящие в S_N, – соответствующие им ранги. Теорема 3.2.2 утверждает сходимость нормированной разности $(S_N - T_N)/\sigma_N$ к нулю в среднеквадратическом смысле (см. (3.2.12)), но нам потребуется только следствие из этого утверждения, состоящее в том, что эта нормированная разность сходится к нулю по вероятности:

$$S_N - T_N \xrightarrow{p} 0 \text{ при } N \to \infty \tag{6.2.14}$$

(в силу (6.2.12) $\sigma_N = \mathrm{const}$ не влияет на сходимость).

Как, однако, связать этот результат, выраженный в терминах вспомогательных случайных величин $U_1, \ldots, U_N$, с задачей проверки гипотез по выборке $X_1, \ldots, X_N$? Прием, устанавливающий такую связь, уже возникал у нас в п. 3.3.3, где в качестве одного из примеров применения теоремы 3.2.1 рассматривался критерий нормальных меток. Там использовался тот факт, что если случайная величина X имеет непрерывную функцию распределения $F(x)$, то случайная величина $F(X)$ распределена равномерно на $[0,1]$. Докажем несколько более общий факт: пусть $F(x)$ и $G(x)$ – две непрерывные строго возрастающие функции распределения, принимающие значения строго между 0 и 1 (т.е. соответствующие случайные величины могут принимать любые значения от $-\infty$ до $+\infty$). Пусть X имеет функцию распределения $F(x)$. Как распределена случайная величина $Y = G(X)$? В силу сделанных предположений Y принимает значения в интервале $(0,1)$ и ее функция распределения равна

$$F_Y(u) = P(G(X) < u) = P(X < G^{-1}(u)) = F(G^{-1}(u)), \ \ 0 \leqslant u \leqslant 1.$$

В частном случае, когда $G(x) = F(x)$, имеем

$$F_Y(u) = F(F^{-1}(u)) = u, \quad 0 \leqslant u \leqslant 1,$$

т.е. $F(X)$ имеет равномерное распределение на $[0,1]$[5].

Возьмем теперь в качестве $F(x)$ функцию распределения наших наблюдений $X_1, \ldots, X_N$, которую они имеют при гипотезе

[5]Предоставляем читателю проверить, что то же рассуждение проходит, если оба распределения сосредоточены на одном и том же интервале (a,b), $-\infty \leqslant a < b \leqslant \infty$, т.е. $F(x) = G(x) = 0$ при $x \leqslant a$, $F(x) = G(x) = 1$ при $x \geqslant b$, и обе функции строго монотонны при $a < x < b$. Предположение, что, например, $b = \infty$, означает, что, скажем, $F(x) < 1$ при всех $x > a$, т.е. соответствующее распределение сосредоточено на полупрямой $\{x > a\}$. Указанным предположениям удовлетворяют, например, экспоненциальное или гамма-распределения, сосредоточенные на положительной полуоси. Утверждение о равномерной распределенности $F(X)$ справедливо на самом деле для любой непрерывной ф.р. F, не обязательно строго монотонной. Предоставляем проверить это читателю.

H_0' (6.2.4). В силу сказанного величины $U_i = F(X_i)$, $i = 1, \ldots, N$, распределены при гипотезе H_0' равномерно на $[0, 1]$, и мы можем считать, что T_N (6.2.13) имеет вид

$$T_N = \sum_{i=1}^{N} c_{N,i}\varphi(F(X_i)). \tag{6.2.15}$$

При этом в S_N (см. (6.2.3)) входят ранги $R_{N,1}, \ldots, R_{N,N}$ величин $U_1, \ldots, U_N$, где $U_i = F(X_i)$, которые в силу монотонности функции F являются в то же время рангами величин $X_1, \ldots, X_N$.

Заметим, что статистика T_N (6.2.15) имеет вид (6.1.13), рассматриваемый в п. 6.1.1, с $g(X_i) = \varphi(F(X_i))$. При этом задача (6.2.4) – (6.2.8) это в точности задача (6.1.1) – (6.1.2). Поэтому к статистике T_N и основанному на ней критерию применимы результаты п. 6.1.1. В частности, следствие 6.1.2 дает распределение T_N при альтернативе H_1', из которого выводится формула (6.1.22) для мощности соответствующего критерия. Таким образом, мы получаем, что мощность $\beta_{N,T}(t)$ критерия, основанного на T_N (6.2.15), при альтернативе (6.2.8) сходится к пределу

$$\beta_{N,T}(t) \to \beta_T(t) = \Phi\left(t\,\frac{\sigma(l^{(1)}, g)}{\sigma_g} - z_\alpha\right), \tag{6.2.16}$$

где σ_g^2 и $\sigma(l^{(1)}, g)$ определены формулами (6.1.12) и (6.1.15) с $g(x) = \varphi(F(x))$:

$$\sigma_g^2 = E_0(\varphi(F(X_1)))^2, \quad \sigma(l^{(1)}, g) = E_0(l^{(1)}(X_1), \varphi(F(X_1))). \tag{6.2.17}$$

Здесь E_0 обозначает математическое ожидание, отвечающее гипотезе H_0' (6.2.4) и вычисляемое интегрированием по ф.р. F. Имеем

$$\sigma_g^2 = \int \varphi^2(F(x))\,dF(x) = \int_0^1 \varphi^2(u)\,du = \sigma_\varphi^2 \tag{6.2.18}$$

(см. (6.2.10)), а $\sigma(l^{(1)}, g)$ с $g = \varphi(F)$ переобозначим как $\sigma(l^{(1)}, \varphi(F))$,

$$\sigma(l^{(1)}, \varphi(F)) = \int l^{(1)}(x)\varphi(F(x))\,dF(x) = \int_0^1 l^{(1)}(F^{-1}(x))\varphi(u)\,du. \tag{6.2.19}$$

Как уже отмечалось, аппроксимация (6.2.14), полученная в теореме 3.2.2 для случая, когда выполнена гипотеза H_0 одинаковой распределенности наблюдений $X_1, \ldots, X_N$, сохраняется при контигуальных альтернативах (6.2.2) или (6.2.8). Поэтому статистика S_N имеет при этих альтернативах то же предельное распределение, что и T_N, а следовательно, критерий, основанный на этой статистике, имеет ту же предельную мощность $\beta_T(t)$. Обозначая $\beta_{N,S}(t)$ мощность нашего рангового критерия, имеем

$$\beta_{N,S}(t) \to \Phi\left(t\,\frac{\sigma(l^{(1)},\varphi(F))}{\sigma_\varphi} - z_\alpha\right), \tag{6.2.20}$$

где σ_φ и $\sigma(l^{(1)},\varphi(F))$ даются формулами (6.2.10) и (6.2.19).

В следующих двух пунктах мы приведем несколько примеров применения формулы (6.2.20) к ранговым статистикам в задаче сравнения двух выборок, рассмотренным в § 3.3.

В этой задаче центрированные и нормированные коэффициенты $c_{N,i}$ (6.2.6) получаются из коэффициентов (3.3.2) делением центрированных коэффициентов (3.3.3) на корень из суммы их квадратов (3.3.4), что дает

$$c_{N,i} = \begin{cases} \sqrt{n_2/Nn_1} & \text{при } \ i = 1,\ldots,n_1, \\ -\sqrt{n_1/Nn_2} & \text{при } \ i = n_1+1,\ldots,N. \end{cases} \tag{6.2.21}$$

6.2.2. Примеры: Семейство нормальных распределений. В этом пункте мы рассматриваем задачу сравнения двух выборок в предположении, что наблюдения имеют нормальное распределение $\mathcal{N}(\theta,\sigma^2)$, т.е. задачу проверки гипотезы (6.2.1) против альтернативы (6.2.2) по наблюдениям $X_1,\ldots,X_N$, имеющим распределения с плотностями $p_{\theta_i}(x) = \varphi(x;\theta_i,\sigma^2)$. Если σ^2 известно, то, как показано в п. 8.1.4, критерий, основанный на статистике $\overline{X}_1 - \overline{X}_2$, где $\overline{X}_1$ и $\overline{X}_2$ – средние арифметические наблюдений 1-й и 2-й выборок, является равномерно наиболее мощным (РНМ) критерием в этой задаче. В общем случае, когда σ^2 неизвестно, применяется критерий Стьюдента, являющийся РНМ несмещенным критерием в этой задаче (см. ссылку в конце п. 8.1.4). Если $N = n_1 + n_2 \to \infty$ (где n_1 и n_2 – объемы 1-й и 2-й выборок) так, что $\min(n_1,n_2) \to \infty$ (ср. § 3.3), то критерий Стьюдента имеет ту же предельную мощность, что и критерий, основанный на $\overline{X}_1 - \overline{X}_2$. Ниже мы выпишем эту предельную мощность с тем, чтобы сравнить с ней мощности ранговых критериев.

В п. 8.1.4 получено выражение для мощности критерия, основанного на статистике $\overline{X}_1 - \overline{X}_2$ в задаче (6.2.4)–(6.2.8): по формуле (8.1.28) она равна $\Phi(t/\sigma - z_\alpha)$, где $\Phi(z_\alpha) = 1-\alpha$. (То, что рассматриваемая в п. 8.1.4 задача совпадает с (6.2.4)–(6.2.8) и параметр t имеет тот же смысл, что и в настоящем пункте, объясняется в следующем за (8.1.28)– (8.1.29) абзаце.) Обозначим $\beta_{N,T}(t)$ мощность критерия Стьюдента. При $N \to \infty$ она сходится к мощности указанного выше критерия, т.е.

$$\beta_{N,T}(t) \to \Phi\left(\frac{t}{\sigma} - z_\alpha\right). \tag{6.2.22}$$

Заметим, что то же выражение для мощности можно было бы получить как предельное по формуле (5.3.27). Для применения этой формулы требуется вычислить информацию Фишера для семейства нормальных распределений с плотностями $p_\theta(x) = \varphi(x; \theta, \sigma^2)$. Пользуясь формулой (8.1.7), получаем

$$l^{(1)}(x) = \frac{x}{\sigma^2}, \quad I = E(l^{(1)})^2 = \frac{1}{\sigma^2}. \tag{6.2.23}$$

Тем самым коэффициент $1/\sigma$ при t в (6.2.22) есть величина $\sqrt{I}$, входящая в общую формулу (5.3.27).

Приведем теперь примеры применения формулы (6.2.20) к ранговым статистикам, рассмотренным в § 3.3.

Пример 6.2.1. *Критерий Уилкоксона* (см. п. 3.3.1). Статистика Уилкоксона дается формулой (3.3.6), но следующая за ней формула (3.3.7) требует небольшой модификации, поскольку в настоящем пункте мы условились сразу центрировать функцию $\varphi(u)$, вычитая из нее среднее значение $\bar{\varphi}$ (см. (6.2.10)). Поэтому от $\varphi(u) = u$ (см. п. 3.3.1) мы перейдем к $\varphi(u) = u - \frac{1}{2}$ и вместо статистики (3.3.7) будем рассматривать статистику

$$S_N = \sum_{i=1}^{N} c_{N,i}\varphi\left(\frac{R_{N,i}}{N+1}\right) = \sum_{i=1}^{N} c_{N,i}\left(\frac{R_{N,i}}{N+1} - \frac{1}{2}\right). \tag{6.2.24}$$

Для применения к ней формулы (6.2.20) нам нужно вычислить σ_φ и $\sigma(l^{(1)}, \varphi(F))$ по формулам (6.2.10) и (6.2.19) с $F(x) = \Phi(x; 0, \sigma^2) = \Phi\left(\frac{x}{\sigma}\right)$. Имеем

$$\sigma_\varphi^2 = \int_0^1 \left(u - \frac{1}{2}\right)^2 du = \frac{1}{12}. \tag{6.2.25}$$

Далее, по формуле (6.2.19), используя выражение (6.2.23) для $l^{(1)}$ и делая замену $y = x/\sigma$, получаем

$$\sigma(l^{(1)}, \varphi(F)) = \frac{1}{\sigma} \int y \left(\Phi(y) - \frac{1}{2} \right) d\Phi(y). \tag{6.2.26}$$

Во избежание недоразумений будем обозначать плотность стандартного нормального распределения буквой ψ (вместо обычного φ), т.е. $\psi(x) = \Phi'(x) = (1/\sqrt{2\pi}) \exp\left(-\frac{1}{2}x^2\right)$. Тогда в (6.2.26) $d\Phi(x) = \psi(x)\,dx$ (возвращаясь к переменной интегрирования x). Пользуясь хорошо известными (или легко проверяемыми) равенствами

$$\int x\psi(x)\,dx = 0, \quad \psi'(x) = -x\psi(x),$$

(6.2.26) можно переписать как

$$\sigma(l^{(1)}, \varphi(F)) = -\frac{1}{\sigma} \int \Phi(x)\psi'(x)\,dx = \frac{1}{\sigma} \int \psi^2(x)\,dx$$

(последнее равенство получено интегрированием по частям). Пользуясь явным выражением для $\psi(x)$, находим

$$\psi^2(x) = \frac{1}{2\pi} \exp(-x^2) = \frac{1}{\sqrt{2\pi}} \cdot \frac{1}{\sqrt{2}} \psi\left(x; 0, \frac{1}{2}\right),$$

где $\psi(x; \mu, \sigma^2)$ – плотность $\mathcal{N}(\mu, \sigma^2)$. В результате получаем

$$\sigma(l^{(1)}, \varphi(F)) = \frac{1}{2\sigma\sqrt{\pi}}. \tag{6.2.27}$$

Подставляя (6.2.25) и (6.2.27) в (6.2.20) и обозначая мощность критерия Уилкоксона через $\beta_{N,W}(t)$, получаем

$$\beta_{N,W}(t) \to \Phi\left(t\,\frac{\sqrt{12}}{2\sigma\sqrt{\pi}} - z_\alpha\right) = \Phi\left(\frac{t}{\sigma}\sqrt{\frac{3}{\pi}} - z_\alpha\right). \tag{6.2.28}$$

Сравнивая с (6.2.22), видим, что коэффициенты при t в выражениях для предельной мощности критериев Уилкоксона и Стьюдента отличаются множителем $\sqrt{3/\pi}$. Это позволяет нам найти питменовскую асимптотическую эффективность (АЭ) критерия Уилкоксона (см. п. 6.1.2). Напомним, что она определяется как предел отношения объемов выборок, необходимых данному критерию (здесь – Уилкоксона) и асимптотически НМ критерию (здесь – Стьюдента) для достижения одной и той же мощности при «близких» альтернативах (точные формулировки см. в п. 6.1.2). Эта же АЭ может рассматриваться и как асимптотическая относительная эффективность (АОЭ) критерия Уилкоксона по отношению к критерию Стьюдента (см. тот же п. 6.1.2). Будем обозначать ее $e_{W,T}$. Рассуждения, проведенные в п. 6.1.2, применимы без всяких изменений к рассматриваемой здесь задаче сравнения двух выборок, и мы можем воспользоваться полученными в нем результатами. Из формул (6.1.22), (6.1.23) и (6.1.29) видно, что АОЭ $e_{W,T}$ равна квадрату отношения коэффициентов при t в выражениях для предельной мощности соответствующих критериев. Из полученных выше выражений находим

$$e^N_{W,T} = \frac{3}{\pi} \approx 0.955.$$

Верхний индекс N поставлен в знак того, что эта эффективность вычислена для случая нормально распределенных наблюдений.

Пример 6.2.2. *Медианный критерий* (3.3.8) (см. п. 3.3.2) – это критерий с метками, определяемыми функцией $\varphi(u) = \operatorname{sign}\bigl(u - \frac{1}{2}\bigr)$. Предлагаем читателю аналогично предыдущему примеру найти его асимптотическую мощность и АОЭ $e_{M,T}$ (M – от «медианный») этого критерия относительно критерия Стьюдента и убедиться, что она равна

$$e^N_{M,T} = \frac{2}{\pi} \approx 0.637.$$

Пример 6.2.3. *Критерий нормальных меток* – это ранговый критерий с $\varphi(u) = \Phi^{-1}(u)$ (см. п. 3.3.3). Как и выше, вычисляем σ_φ и $\sigma(l^{(1)}, \varphi(F))$. По формуле (3.3.11) $\sigma_\varphi = 1$. Для вычисления $\sigma(l^{(1)}, \varphi(F))$ подставляем в первое выражение в (6.2.19) $l^{(1)}(x) = x/\sigma^2$ (см. (6.2.23)), $F(x) = \Phi(x/\sigma)$, $\varphi(F(x)) = \Phi^{-1}(\Phi(x/\sigma)) = x/\sigma$. Получаем:

$$\sigma(l^{(1)}, \varphi(F)) = \frac{1}{\sigma^3} \int x^2\, d\Phi\left(\frac{x}{\sigma}\right) = \frac{1}{\sigma}.$$

По формуле (6.2.20) мощность $\beta_{N,\mathcal{N}}(t)$ критерия нормальных меток сходится к пределу

$$\beta_{N,\mathcal{N}}(t) \to \Phi\left(\frac{t}{\sigma} - z_\alpha\right). \tag{6.2.29}$$

Сравнивая с (6.2.22), видим, что критерий нормальных меток имеет ту же предельную мощность, что и критерий Стьюдента. Как следствие, его АОЭ по отношению к критерию Стьюдента равна 1,

$$e^{N}_{\mathcal{N},T} = 1. \tag{6.2.30}$$

6.2.3. Примеры: Семейство распределений Лапласа. В этом пункте мы рассмотрим случай, когда наблюдения имеют распределение Лапласа (6.1.30) (см. пример 6.1.1 в п. 6.1.2), и найдем АОЭ $e_{W,M}$ критерия Уилкоксона и $e_{\mathcal{N},M}$ критерия нормальных меток (см. п. 3.3.3) по отношению к медианному критерию. Для этого мы выпишем асимптотические мощности трех критериев, пользуясь формулой (6.2.20). В этом пункте $F(x)$ обозначает функцию распределения, отвечающую плотности $p(x) = \frac{1}{2}e^{-|x|}$.

Медианный критерий. Как сказано выше, этот критерий задается функцией $\varphi(u) = \operatorname{sign}\bigl(u - \frac{1}{2}\bigr)$. Отсюда $\varphi^2(u) \equiv 1$ и $\sigma^2_\varphi = 1$ (см. (6.2.10)). По формуле (6.1.31) для рассматриваемого семейства $l^{(1)}(x) = \operatorname{sign} x$. Поэтому по формуле (6.2.19)

$$\sigma(l^{(1)}, \varphi(F)) = \int (\operatorname{sign} x)^2\, dF(x) = 1.$$

Следовательно, для мощности $\beta_{N,M}(t)$ медианного критерия по формуле (6.2.20) получаем

$$\beta_{N,M}(t) \to \Phi(t - z_\alpha)$$

(что совпадает с выражением (6.1.34) для предельной мощности аналогичного одновыборочного критерия).

Критерий Уилкоксона. Здесь $\varphi(u) = u - \frac{1}{2}$ и $\sigma^2_\varphi = 1/12$ (см. (6.2.25)). Пользуясь вторым выражением в формуле (6.2.19), находим

$$\sigma(l^{(1)}, \varphi(F)) = \int_0^1 \operatorname{sign}(F^{-1}(u))\Bigl(u - \frac{1}{2}\Bigr)\, du = 2\int_{1/2}^1 \Bigl(u - \frac{1}{2}\Bigr)\, du = \frac{1}{4}.$$

Отсюда по формуле (6.2.20) получаем

$$\beta_{N,W}(t) \to \Phi\biggl(\frac{\sqrt{3}}{2}t - z_\alpha\biggr).$$

Критерий нормальных меток. Это ранговый критерий с $\varphi(u) = \Phi^{-1}(u)$. По формуле (6.2.10)

$$\sigma_\varphi^2 = \int_0^1 (\Phi^{-1}(u))^2\,du = \int x^2\,d\Phi(x) = 1.$$

Снова применяя второе выражение в формуле (6.2.19), находим

$$\begin{aligned}\sigma(l^{(1)}, \varphi(F)) &= \int_0^1 \operatorname{sign}(F^{-1}(u))\Phi^{-1}(u)\,du = 2\int_{1/2}^1 \Phi^{-1}(u)\,du \\ &= 2\int_0^\infty x\,d\Phi(x) = \frac{2}{\sqrt{2\pi}}\int_0^\infty e^{-x^2/2}\,d\Big(\frac{x^2}{2}\Big) = \sqrt{\frac{2}{\pi}}.\end{aligned} \tag{6.2.31}$$

По формуле (6.2.20) теперь имеем

$$\beta_{N,\mathcal{N}}(t) \to \Phi\bigg(\sqrt{\frac{2}{\pi}}\,t - z_\alpha\bigg).$$

Пользуясь тем же правилом, что АОЭ равна квадрату отношения коэффициентов при t в выражениях для предельной мощности, получаем

$$e^{DE}_{\mathcal{N},M} = \frac{2}{\pi} \approx 0.637, \quad e^{DE}_{W,M} = \frac{3}{4} = 0.75, \quad e^{DE}_{\mathcal{N},W} = \frac{2}{\pi}\cdot\frac{4}{3} = 0.849 \tag{6.2.32}$$

(DE (double-exponential) – сокр. от англ. названия распределения Лапласа).

Предлагаем читателю самостоятельно проверить, что медианный критерий в случае распределения Лапласа эквивалентен ЛАНМ критерию, основанному на статистике Z_N (см. (6.1.4) или (6.1.25)). Поэтому полученные выше АОЭ критерия нормальных меток и критерия Уилкоксона по отношению к медианному критерию представляют питменовскую эффективность этих критериев относительно асимптотически НМ критерия в этой задаче. Предоставляем читателю также проверить, что питменовская эффективность критерия Стьюдента, как и соответствующего одновыборочного критерия в п. 6.1.2, равна

$$e^{DE}_{T,M} = \frac{1}{2}. \tag{6.2.33}$$

В предыдущем пункте мы видели, что критерий нормальных меток асимптотически эффективен по отношению к критерию Стьюдента (см. (6.2.30)) в случае нормально распределенных наблюдений, когда критерий Стьюдента является «наилучшим» (ЛАНМ) критерием. Эта связь между двумя критериями, однако, пропадает при любом отклонении от нормальной модели. В случае распределения Лапласа критерий Стьюдента оказывается менее эффективным, чем каждый из трех рассматриваемых в данном примере ранговых критериев. Две общие теоремы на этот счет будут приведены (без доказательства) в п. 6.2.5.

6.2.4. Асимптотически эффективные ранговые критерии. До сих пор мы исследовали асимптотические свойства (мощность, эффективность) *заранее заданного* рангового критерия. В этом пункте мы ставим задачу построения асимптотически эффективного (или ЛАНМ) рангового критерия для *заданного семейства* распределений с плотностями $p_\theta(x) = p(x-\theta)$, $\theta \in \mathbb{R}$, определяемыми данной плотностью $p(x)$, $x \in \mathbb{R}$. Рассуждения, проведенные в начале предыдущего пункта, позволяют перейти от задачи проверки сложной гипотезы (6.2.1) против сложной альтернативы (6.2.2) к задаче проверки простой гипотезы (6.2.4) против простой альтернативы (6.2.8) с коэффициентами $c_{N,i}$, удовлетворяющими условиям (6.2.7).

В п. 5.3.3 мы видели, что ЛАНМ критерий в этой задаче может быть основан на статистике

$$Z_N = \sum_1^N c_{N,i} l^{(1)}(X_i) \tag{6.2.34}$$

(см. (5.3.30)). В предыдущем пункте было показано, что ранговая статистика S_N (6.2.3) приближается суммой T_N (6.2.15), если ее метки $a_N(i)$ построены по функции $\varphi(u)$, входящей в конструкцию суммы T_N. В настоящем пункте мы используем эту аппроксимацию в обратном направлении: для суммы Z_N нам нужно построить близкую к ней ранговую статистику S_N. Из построений предыдущего пункта видно, что для этого надо представить Z_N в форме (6.2.15), т.е. найти функцию φ такую, чтобы

$$l^{(1)}(X_i) = \varphi(F(X_i)), \tag{6.2.35}$$

а затем построить статистику S_N по этой функции $\varphi(u)$. Ясно, что для этого следует взять

$$\varphi(u) = l^{(1)}(F^{-1}(u)). \tag{6.2.36}$$

Читатель легко убедится, что по этому правилу построен критерий нормальных меток, являющийся ЛАНМ (асимптотически эффективным) в случае нормального семейства.

Пример 6.2.4. (*Логистическое распределение.*) Рассмотрим семейство $p_\theta(x) = p(x-\theta)$, $\theta \in \mathbb{R}$, определяемое плотностью

$$p(x) = \frac{e^{-x}}{(1+e^{-x})^2}.$$

Нетрудно проверить, что

$$F(x) = \int_{-\infty}^{x} p(y)\,dy = \frac{1}{1+e^{-x}},$$

а (обозначая штрихом производную по x)

$$l^{(1)}(x) = -\frac{p'(x)}{p(x)} = \frac{1-e^{-x}}{1+e^{-x}} = 2F(x) - 1 = \varphi(F(x)),$$

где $\varphi(u) = 2u - 1$, $0 \leqslant u \leqslant 1$. Построенная по такой функции $\varphi(u)$ ранговая статистика равна $2S_N$, где S_N (6.2.24) – статистика Уилкоксона, отвечающая функции $\varphi(u) = u - \frac{1}{2}$. Понятно, что множитель 2 не влияет на свойства соответствующего критерия. Таким образом, критерий Уилкоксона является ЛАНМ (асимптотически эффективным) ранговым критерием в случае логистического семейства распределений с параметром сдвига.

Заметим, что в практических приложениях распределение наблюдаемых данных обычно зависит еще от неизвестного параметра масштаба. При построении асимптотически эффективного рангового критерия наличие параметра масштаба, скажем, $\eta > 0$, т.е. рассмотрение плотности $(1/\eta)p(x/\eta)$ вместо $p(x)$, приводит к функции $\varphi(x)$, отличающейся от «стандартной» (отвечающей плотности $p(x)$) лишь постоянным множителем (зависящим от η). Соответствующая ранговая статистика отличается от «стандартной» тем же множителем. (Предлагаем читателю самостоятельно проверить это утверждение и убедиться в его справедливости на данном примере.) Тем самым одна и та же статистика S_N служит статистикой ЛАНМ критерия независимо от масштабного параметра η.

Кроме того, при построении ЛАНМ «параметрического» критерия приходится вводить в статистику критерия оценку этого параметра, компенсирующую зависимость от него распределения статистики при гипотезе H_0 (подобно введению оценки для σ^2 в знаменателе статистики Стьюдента). Привлекательным свойством ранговых критериев является независимость «нулевого» распределения ранговой статистики от параметра масштаба (и более того, от самого распределения наблюдений – свойство свободы от распределения).

6.2.5. Дальнейшие результаты об асимптотической эффективности ранговых критериев (обзор). Приведем два результата, касающихся АОЭ критериев нормальных меток и Уилкоксона по отношению к критерию Стьюдента в задаче сравнения двух выборок. Как и ранее в этом параграфе, предполагается, что наблюдения имеют распределения, принадлежащие семейству $p(x-\theta)$, $\theta \in \mathbb{R}$. Обозначим $F(x)$ функцию распределения, отвечающую плотности $p(x)$. Предполагается, что

$$\int x^2 p(x)\, dx < \infty.$$

Обозначим $e^F_{T,\mathcal{N}}$ АОЭ критерия Стьюдента по отношению к критерию нормальных меток. Тогда

$$e^F_{T,\mathcal{N}} \leqslant 1,$$

причем равенство достигается только когда F – нормальное распределение (Чернов и Сэвидж (1958) [20]).

Для АОЭ $e^F_{W,T}$ критерия Уилкоксона по отношению к критерию Стьюдента при любых F имеет место неравенство

$$e^F_{W,T} \geqslant 0.864$$

(Ходжес и Леман (1956) [22]).

Упомянем еще некоторые результаты, связанные с так называемым *дефектом* (англ.: *deficiency*) асимптотически эффективных критериев.

В статье Ходжеса и Лемана (1970) [23] был обнаружен неожиданный эффект. Напомним, что, как отмечено в п. 6.1.1, существует целый класс локально асимптотически наиболее мощных критериев, т.е. критериев, мощность которых при локальных альтернативах асимптотически совпадает с максимально достижимой мощностью. Из рассмотрений п. 6.1.2 видно, что все эти критерии имеют асимптотическую эффективность 1, и это свойство тем самым не позволяет судить о сравнительном качестве асимптотически эффективных критериев. Поэтому авторы цит. статьи предложили вместо *отношения* объемов выборок, необходимых для достижения заданных ошибок 1-го и 2-го рода при близких альтернативах, рассматривать их *разность*. Ожидалось, что разность объемов выборок, стремящихся к бесконечности, при стремлении их отношения к 1 будет также стремиться к бесконечности, только с меньшей скоростью. То есть, если $N^*(\theta)$ – необходимый объем выборки для наилучшего (НМ) критерия, а $N(\theta)$ – для некоторого асимптотически эффективного критерия (где θ характеризует отклонение от гипотезы H_0), то $N^*(\theta)/N(\theta) \to 1$ при $\theta \to 0$, но при этом естественно ожидать, что $N(\theta) - N^*(\theta) \to \infty$, например, со скоростью порядка $\sqrt{N(\theta)}$. Однако в цит. статье Ходжес и Леман рассмотрели ряд простых, но характерных примеров асимптотически эффективных критериев и оценок и показали, что во всех рассмотренных примерах эта разность стремится к конечному пределу. Эту разность они назвали дефектом критерия, а ее предел – асимптотическим дефектом. Величина дефекта показывает, *на сколько* больше наблюдений требуется данному асимптотически эффективному критерию по сравнению с минимально возможным числом наблюдений. В заключительном разделе статьи авторы высказали предположение, что этот эффект (конечность асимптотического дефекта) носит достаточно общий характер, и поставили задачу исследовать дефекты различных классов статистических процедур.

В 1970-х – начале 1980-х гг. эта задача подверглась детальному исследованию. Выяснилось, что при естественных условиях регулярности конечность асимптотического дефекта является посуществу общим свойством асимптотически эффективных процедур. Соответствующие результаты, относящиеся к ранговым критериям, были получены Алберсом, Бикелом и Ван Цветом [17] для одновыборочных ранговых статистик и Бикелом и Ван Цветом [19] для двухвыборочных ПЛРС, рассматриваемых в настоящем курсе. Оказалось, что ранговые статистики, рассмотренные в этих работах, имеют конечный асимптотический дефект, за одним исключением: критерий нормальных меток (как двухвыборочный, так и его одновыборочный аналог) представляет особый случай и его дефект d_N растет при $N \to \infty$, хотя и очень медленно:

$$d_N \sim \frac{1}{2} \log \log N.$$

Эти результаты показали, что ранговые критерии очень мало уступают «параметрическим» критериям: применение асимптотически эффективного рангового критерия вместо НМ «параметрического» критерия требует для получения той же мощности увеличения объема выборки на конечное число наблюдений (при том, что и в указанном особом случае $\frac{1}{2} \log \log N = 0.97$ при $N = 1000$, т.е. и в этом случае в большинстве практических ситуаций достаточно одного дополнительного наблюдения).

Отметим, что даже для критериев достаточно простой структуры формулы для асимптотического дефекта возникали в результате весьма громоздких выкладок. Это, в частности, не позволяло придать этим формулам наглядный характер. В начале 1980-х гг. автором настоящих записок была предложена явная формула для асимптотического дефекта с достаточно простым эвристическим выводом [14], [21]. Эта формула объясняла структуру асимптотического дефекта и позволяла легко выписать его в конкретных задачах. Обоснование этой формулы требовало, однако, достаточно сложной техники, и при этом различной для различных классов статистик критериев. Автором такое обоснование было дано в работе 1985 г. для широкого класса статистик «параметрических» критериев. Для одновыборочных ранговых статистик обоснование этой формулы было дано В. Е. Бенингом [18].

Глава 7. Условия локальной асимптотической нормальности

Цель настоящей главы – доказать теоремы 5.2.1 и 5.2.2, сформулированные в п. 5.2.3. Первая из них доказывается в § 7.1, вторая – в § 7.2.

§ 7.1. Дифференцируемость корня из плотности в среднем квадратичном

7.1.1. Введение и формулировка теоремы. Для удобства читателя повторим постановку задачи и формулировки теорем. Формулируемая в настоящем пункте теорема 7.1.1 в точности повторяет формулировку теоремы 5.2.1, а теорема 7.2.1 (п. 7.2.2) повторяет теорему 5.2.2.

Мы рассматриваем постановку задачи, введенную в п. 5.2.1, и используем обозначения, введенные в п. 5.2.2. Имеются независимые наблюдения $X_1, \ldots, X_N$ с плотностями распределения, принадлежащими семейству $p(x, \theta)$, $\theta \in \Theta$, где Θ – множество, содержащее интервал вида $\Delta = [-\delta, \delta]$, $\delta > 0$. Наблюдение X_i имеет плотность $p(x, \theta_i)$, $\theta_i \in \Theta$. Относительно значений θ_i (точнее было бы писать $\theta_{N,i}$) рассматриваются две гипотезы: $H_0 \colon \theta_1 = \cdots = \theta_N = 0$ и $H_1 \colon \theta_i = c_{Ni} t$, $i = 1, \ldots, N$. Коэффициенты $c_{N1}, \ldots, c_{NN}$ предполагаются удовлетворяющими условию (5.2.1) нормировки,

$$\sum_{i=1}^{N} c_{Ni}^2 = 1, \tag{7.1.1}$$

и условию (5.2.2) равномерной малости

$$\max_i |c_{N,i}| \to 0 \text{ при } N \to \infty. \tag{7.1.2}$$

Теорема 7.1.1. *Предположим, что в описанной выше схеме коэффициенты* $c_{N,1}, \ldots, c_{N,N}$ *удовлетворяют условиям* (7.1.1) *и* (7.1.2), *а плотности* $p_\theta(x)$, $\theta \in \Theta$, – *условию*

$$\int \left(\frac{\sqrt{p_\theta(x)} - \sqrt{p_0(x)}}{\theta} - g(x) \right)^2 dx \to 0 \tag{7.1.3}$$

для некоторой функции $g(x)$ *такой, что* $\int g^2(x)\,dx < \infty$ *и* $g(x) = 0$ *на множестве* $\{x \colon p_0(x) = 0\}$. *Тогда для любой ограниченной последовательности* $t_N \in \mathbb{R}$

$$\Lambda_{N,t_N} = t_N Z_N - \frac{1}{2} t_N^2 I + \zeta_{N,t_N}, \tag{7.1.4}$$

где Z_N *и* I *определены в* (5.2.20), $Eh = 0$, *и* $\zeta_{N,t_N} \xrightarrow{p} 0$ *относительно* $P_{N,0}$. *Как следствие,*

$$\mathcal{L}(Z_N \mid P_{N,0}) \to \mathcal{N}(0, I). \tag{7.1.5}$$

Приведем для удобства формулы (5.2.20), определяющие Z_N и I:

$$I = 4 \int g^2(x)\,dx, \quad Z_N = \sum_1^N c_{N,i} h(X_i), \quad \text{где} \quad h(x) = \frac{2g(x)}{\sqrt{p(x)}}. \tag{7.1.6}$$

7.1.2. Дифференцируемость отношения правдоподобия в среднем квадратичном. В этом пункте в качестве предварительного результата на пути к доказательству теоремы 7.1.1 мы рассмотрим случай, когда носитель плотности $\{x \colon p(x,\theta) > 0\}$ не зависит от θ, и покажем, что ЛАН выполняется, если отношение правдоподобия дифференцируемо в среднем квадратичном, т.е.

$$\int \left(\frac{p(x,\theta)}{p(x,0)} - 1 - \theta h(x) \right)^2 p(x,0)\,dx = o(\theta^2) \tag{7.1.7}$$

с некоторой функцией $h(x)$ такой, что

$$I := \int h^2(x) p(x,0)\,dx < \infty. \tag{7.1.8}$$

(В регулярном случае, когда существуют частные производные по θ в обычном смысле, $h(x) = p^{(1)}(x)/p(x) = l^{(1)}(x)$. Мы здесь пользуемся обозначениями, введенными в п. 5.2.2, в частности, отсутствие второго аргумента означает, что $\theta = 0$.) Напомним, что свойство ЛАН, которое мы хотим доказать, состоит в выполнении соотношений (7.1.4) и (7.1.5), в которых Z_N определяется формулой, входящей в (7.1.6), а остаточный член $\zeta_{N,t}$ стремится к нулю по $P_{N,0}$-вероятности.

В дальнейшем мы для краткости, как правило, опускаем аргумент x, а аргумент θ пишем в виде индекса, который опускается при $\theta = 0$, например, $p_\theta = p(x, \theta)$, $p = p(x, 0)$. Обозначим r_θ выражение в круглых скобках под знаком интеграла в (7.1.7). Тогда, пользуясь введенной системой обозначений, (7.1.7) можно записать в виде:

$$\frac{p_\theta}{p} = 1 + \theta h + r_\theta, \quad \text{где} \quad Er_\theta^2 = o(\theta^2). \tag{7.1.9}$$

Здесь и далее E обозначает математическое ожидание по плотности p $(= p(x, 0))$, например, $Er_\theta^2 = \int r_\theta^2(x) p(x)\, dx$. Кроме того, P будет обозначать вероятность, отвечающую плотности p, т.е. $P(A) = \int_A p(x)\, dx$ для произвольного борелевского множества A. Дополнение множества A обозначаем A^c.

Покажем прежде всего, что из условия (7.1.7) (или, что то же, из условия на Er_θ^2 в (7.1.9)) следует, что $Eh = 0$. Действительно, учитывая, что $E(p_\theta/p) = \int p_\theta\, dx = 1$, и беря математические ожидания от обеих частей равенства в (7.1.9), получаем $\theta Eh = -Er_\theta$. Но $|Er_\theta| \leqslant \sqrt{Er_\theta^2} = o(\theta)$, следовательно, $Eh = o(1)$, и таким образом $Eh = 0$, поскольку Eh – фиксированное число, не зависящее от θ.

Кроме того, снова беря математические ожидания от обеих частей равенства (7.1.9), получаем

$$Er_\theta = 0. \tag{7.1.10}$$

Для некоторого $0 < a < 1/2$ введем множества $A_{1\theta} = \{x\colon |r_\theta| \leqslant a\}$. Из условия на r_θ в (7.1.9) по неравенству Чебышева[1] имеем $P(A_{1\theta}^c) = o(\theta^2)$. Кроме того, введем множества $A_{2\theta} = \{x\colon \theta|h| \leqslant a\}$. Покажем, что $A_{2\theta}$ обладают тем же свойством:

$$P(A_{2\theta}^c) = o(\theta^2). \tag{7.1.11}$$

Непосредственное применение неравенства Чебышева дает:

$$P(A_{2\theta}^c) = P\left(|h| > \frac{a}{\theta}\right) \leqslant \frac{\theta^2}{a^2} Eh^2, \tag{7.1.12}$$

[1]Используется неравенство $P(|X| > a) \leqslant EX^2/a^2$.

т.е. оценку порядка $O(\theta^2)$, но не $o(\theta^2)$. Но вспомним доказательство неравенства Чебышева:

$$P\left(|h| > \frac{a}{\theta}\right) = \int_{\{|h|>a/\theta\}} p\,dx \leqslant \frac{\theta^2}{a^2}\int_{\{|h|>a/\theta\}} h^2 p\,dx. \tag{7.1.13}$$

Далее неравенство (7.1.12) получается распространением области интегрирования в последнем интеграле на все пространство. Этим достигается универсальность неравенства Чебышева: оно представляет точное неравенство, выполняющееся равномерно по классу всех случайных величин с данным математическим ожиданием квадрата (или с данной дисперсией, если случайная величина предварительно центрирована своим математическим ожиданием). Но если для конкретной функции h рассмотреть интеграл в правой части (7.1.13), то видно, что этот интеграл берется по внешности бесконечно расширяющегося интервала и следовательно, стремится к нулю при $\theta \to 0$. Поэтому из (7.1.13) следует (7.1.11).

Положим $A_\theta = A_{1\theta} \cap A_{2\theta}$. Из (7.1.11) и аналогичного свойства $A_{1\theta}$ следует

$$P(A_\theta^c) \leqslant P(A_{1\theta}^c) + P(A_{2\theta}^c) = o(\theta^2). \tag{7.1.14}$$

Пользуясь обозначениями, введенными в п. 5.2.2, имеем

$$\Lambda_{N,t} = \log \prod_{i=1}^{N} \frac{p_{\theta_i}(X_i)}{p(X_i)} = \sum_{i=1}^{N} \log \frac{p_{\theta_i}(X_i)}{p(X_i)}, \tag{7.1.15}$$

где $\theta_i = c_{Ni}t$ с константами c_{Ni}, удовлетворяющими (7.1.1) и (7.1.2). Подставим в эту формулу представление для p_θ/p из (7.1.9):

$$\Lambda_{N,t} = \sum_{i=1}^{N} \log(1 + \theta_i h + r_{\theta_i}).$$

Положим $h'_\theta = h\mathbf{1}_{A_\theta}$ и $r'_\theta = r_\theta \mathbf{1}_{A_\theta}$ и определим

$$\Lambda'_{N,t} = \sum_{i=1}^{N} \log(1 + \theta_i h'_{\theta_i} + r'_{\theta_i}). \tag{7.1.16}$$

Введем множество

$$A_{N,t} = \times_{i=1}^{N} A_{\theta_i} = \{\mathbf{x} = (x_1, \ldots, x_N) \colon x_i \in A_{\theta_i}\}. \tag{7.1.17}$$

Множество $A_{N,t}$ состоит из точек $\mathbf{x} = (x_1, \ldots, x_N)$, каждая координата x_i которых принадлежит соответствующему множеству A_{θ_i}, а его дополнение $A^c_{N,t}$ это множество точек, у которых хотя бы одна координата лежит в дополнении $A^c_{\theta_i}$. Заметим, что в силу (7.1.14) и (7.1.1)

$$
\begin{aligned}
P(A^c_{N,t}) &= P\big(X_i \in A^c_{\theta_i} \ \text{хотя бы при одном} \ i = 1, \ldots, N\big) \\
&\leqslant \sum P(X_i \in A^c_{\theta_i}) = \sum o(\theta_i^2) = o(1).
\end{aligned}
\tag{7.1.18}
$$

Равенство $\sum o(\theta_i^2) = o(1)$, которое будет неоднократно встречаться в дальнейшем, требует пояснения. Если про какую-то функцию, скажем, $\xi(\theta)$, известно, что $\xi(\theta) = o(\theta^2)$, это значит, что $\xi(\theta) = \theta^2\varepsilon(\theta)$, где $\varepsilon(\theta) \to 0$ при $\theta \to 0$. Положим $\bar\varepsilon(\theta) = \sup\{|\varepsilon(\tau)|\colon 0 < |\tau| < |\theta|\}$. Тогда $|\xi(\theta)| \leqslant \theta^2\bar\varepsilon(\theta)$ и $\bar\varepsilon(\theta) \to 0$ при $\theta \to 0$, причем $\bar\varepsilon(\theta_1) \leqslant \bar\varepsilon(\theta_2)$, если $|\theta_1| < |\theta_2|$. Теперь, вспоминая, что $\theta_i = c_{N,i}t_N$, где последовательность t_N ограничена, скажем, $|t_N| \leqslant C$, а $c_{N,i}$ удовлетворяют условиям (7.1.1) и (7.1.2), получаем, что

$$
\sum \xi(\theta_i) \leqslant C^2 \sum c^2_{N,i}\bar\varepsilon(\theta_i) \leqslant C^2\bar\varepsilon(C\max_i |c_{N,i}|) \sum c^2_{N,i} \to 0,
$$

поскольку $\sum c^2_{N,i} = 1$ в силу (7.1.1), а $C\max_i |c_{N,i}| \to 0$ в силу (7.1.2). В дальнейшем мы будем использовать соотношение

$$
\sum o(\theta_i^2) = o(1),
$$

имея в виду проведенное выше рассуждение.

Разность $\Lambda_{N,t} - \Lambda'_{N,t}$ отлична от нуля на множестве $A^c_{N,t}$ и из $P(A^c_{N,t}) \to 0$ следует, что $\Lambda_{N,t} - \Lambda'_{N,t} \xrightarrow{p} 0$. Таким образом, (7.1.4) можно доказывать для $\Lambda'_{N,t}$, включая отличие от $\Lambda_{N,t}$ в остаточный член $\zeta_{N,t}$.

Воспользуемся формулой Тейлора для функции $\log(1+x)$:

$$
\log(1+x) = x - \frac{1}{2}x^2 + \frac{x^3}{3(1+\lambda x)^3}, \quad 0 < \lambda < 1, \quad x > -1 \tag{7.1.19}
$$

(последний член представляет произведение $x^3/3!$ на 3-ю производную функции $\log(1+x)$, взятую в промежуточной точке). Применяя эту формулу к представлению (7.1.16), получим

$$
\Lambda'_{N,t} = \sum_{i=1}^{N} \left[(\theta_i h'_{\theta_i} + r'_{\theta_i}) - \frac{1}{2}(\theta_i h'_{\theta_i} + r'_{\theta_i})^2 \right] + \eta_{N,t}, \tag{7.1.20}
$$

где

$$\eta_{N,t} = \sum_{i=1}^{N} \frac{(\theta_i h'_{\theta_i} + r'_{\theta_i})^3}{(1+\lambda_i(\theta_i h'_{\theta_i} + r'_{\theta_i}))^3}, \quad 0 < \lambda_i < 1. \tag{7.1.21}$$

Рассмотрим суммы членов, входящих в (7.1.20). Все суммирования далее производятся по i от 1 до N, поэтому пределы суммирования при знаках суммы опускаем.

Учитывая, что $\theta_i = c_{N,i}t$, сумму слагаемых $\theta_i h'_{\theta_i}$ обозначим tZ'_N, где

$$Z'_N = \sum c_{N,i} h'_{\theta_i}(X_i). \tag{7.1.22}$$

Положим

$$Z_N = \sum c_{N,i} h(X_i). \tag{7.1.23}$$

Разность $Z'_N - Z_N$ отлична от нуля, когда хотя бы одна из величин X_i принадлежит $A^c_{\theta_i}$, т.е. когда $\mathbf{X} = (X_1, \ldots, X_N) \in A^c_{N,t}$, следовательно,

$$tZ'_N = tZ_N + \zeta^{(1)}_{N,t}, \quad \text{где} \quad \zeta^{(1)}_{N,t} \xrightarrow{p} 0. \tag{7.1.24}$$

Далее, обозначим $\zeta^{(2)\prime}_{N,t} = \sum_{i=1}^{N} r'_{\theta_i}$ и $\zeta^{(2)}_{N,t} = \sum_{i=1}^{N} r_{\theta_i}$. Из тех же соображений, что и выше, имеем $P\big(\zeta^{(2)\prime}_{N,t} \neq \zeta^{(2)}_{N,t}\big) \to 0$. Поэтому из $\zeta^{(2)}_{N,t} \xrightarrow{p} 0$ будет следовать $\zeta^{(2)'}_{N,t} \xrightarrow{p} 0$. Учитывая, что слагаемые в сумме $\zeta^{(2)}_{N,t}$ независимы и $Er_\theta = 0$, а $Er^2_\theta = o(\theta^2)$, получаем

$$E(\zeta^{(2)}_{N,t})^2 = \sum Er^2_{\theta_i} = \sum o(\theta_i^2) = o(1), \ \text{следовательно,} \ \zeta^{(2)}_{N,t} \xrightarrow{p} 0. \tag{7.1.25}$$

Отсюда

$$\zeta^{(2)\prime}_{N,t} \xrightarrow{p} 0. \tag{7.1.26}$$

Рассмотрим теперь $\sum \theta_i^2 h'^2_{\theta_i}$. Снова заменяем h'_{θ_i} на h с погрешностью, стремящейся к нулю при $N \to \infty$, и к полученной сумме $t^2 \sum c^2_{N,i} h^2(X_i)$ применяем теорему 8.3.5, согласно которой

$$\sum c^2_{N,i} h^2(X_i) \xrightarrow{p} Eh^2 = I.$$

Таким образом, получаем

$$-\frac{1}{2} \sum \theta_i^2 h'^2_{\theta_i} = -\frac{1}{2} t^2 I + \zeta^{(3)}_{N,t}, \quad \zeta^{(3)}_{N,t} \xrightarrow{p} 0. \tag{7.1.27}$$

Положим теперь $\zeta^{(4)}_{N,t} = \sum \theta_i h'_{\theta_i} r'_{\theta_i}$. По неравенству Коши–Буняковского

$$\sum \theta_i h'_{\theta_i} r'_{\theta_i} \leqslant \left(\sum (\theta_i h'_{\theta_i})^2 \right)^{1/2} \left(\sum (r'_{\theta_i})^2 \right)^{1/2}.$$

Как было только что показано, $\sum(\theta_i h'_{\theta_i})^2 \xrightarrow{p} t^2 I$, а аналогично доказательству (7.1.25) видим, что $\sum(r'_{\theta_i})^2 \xrightarrow{p} 0$. Следовательно,

$$\zeta^{(4)}_{N,t} \xrightarrow{p} 0. \tag{7.1.28}$$

Точно так же

$$\zeta^{(5)}_{N,t} = \sum (r'_{\theta_i})^2 \xrightarrow{p} 0. \tag{7.1.29}$$

Остается показать, что остаточный член $\eta_{N,t}$ (см. (7.1.20), (7.1.21)) стремится к нулю по вероятности. Тогда из $\Lambda_{N,t} - \Lambda'_{N,t} \xrightarrow{p} 0$ и (7.1.20)–(7.1.29) будет следовать искомое соотношение (7.1.4), где Z_N определяется формулой (7.1.23).

Заметим прежде всего, что $0 < a < 1/2$ и по определению $\theta_i h'_{\theta_i} < a$ и $r'_{\theta_i} < a$. Поэтому знаменатель каждого слагаемого в (7.1.21) не меньше $(1-2a)^3$, следовательно

$$|\eta_{N,t}| \leqslant \frac{1}{(1-2a)^3} \sum_{i=1}^{N} |\theta_i h'_{\theta_i} + r'_{\theta_i}|^3. \tag{7.1.30}$$

Для произвольных случайных величин X и Y имеет место неравенство $E|X+Y|^3 \leqslant 4(E|X|^3 + E|Y|^3)$ (см. «c_r-неравенство»[2] на с. 168 книги Лоэва [8]). Покажем, что

$$E|\theta h'_\theta|^3 = o(\theta^2), \quad E|r'_\theta|^3 = o(\theta^2). \tag{7.1.31}$$

Тогда из (7.1.30) следует $\eta_{N,t} \xrightarrow{p} 0$.

[2]Упомянутое «c_r-неравенство» выводится в книге Лоэва из неравенства

$$|a+b|^r \leqslant c_r|a|^r + c_r|b|^r, \quad r > 0, \quad c_r = \begin{cases} 1, & r \leqslant 1, \\ 2^{r-1}, & r > 1, \end{cases}$$

которое названо автором «элементарным» и приведено без доказательства. В интересующем нас случае $r > 1$ это неравенство следует из выпуклости функции x^r:

$$\left| \frac{a+b}{2} \right|^r \leqslant \frac{|a|^r + |b|^r}{2}.$$

Второе соотношение в (7.1.31) следует из $|r'_\theta| < a$, откуда $E|r'_\theta|^3 \leqslant aE|r'_\theta|^2 = o(\theta^2)$. Рассмотрим первое соотношение. Имеем

$$\begin{aligned} E|\theta h'_\theta|^3 &= |\theta|^3 \int |h'_\theta(x)|^3 p(x)\,dx \\ &\leqslant |\theta|^3 \int_{|h'_\theta(x)|\leqslant|\theta|^{-1/6}} |h'_\theta(x)|^3 p(x)\,dx \\ &\quad + |\theta|^3 \int_{|h'_\theta(x)|>|\theta|^{-1/6}} |h'_\theta(x)|^3 p(x)\,dx. \end{aligned} \tag{7.1.32}$$

Первое слагаемое в правой части не превосходит $|\theta|^3 \cdot |\theta|^{-1/2} = |\theta|^{5/2} = o(\theta^2)$. Во втором слагаемом пользуемся оценкой $|\theta h'_\theta| < a$, поэтому оно не превосходит

$$a\theta^2 \int_{|h'_\theta(x)|>|\theta|^{-1/6}} (h'_\theta(x))^2 p(x)\,dx = o(\theta^2),$$

поскольку интеграл здесь берется по внешности бесконечно расширяющейся области и потому стремится к нулю. Из полученных оценок следует первое соотношение в (7.1.31). Следовательно мы доказали, что $\eta_{N,t} \to 0$.

Этим наше утверждение, что дифференцируемость отношения правдоподобия в среднем квадратическом влечет свойство ЛАН, полностью доказано. □

Доказанное условие налагает, однако, излишне сильные требования на семейство $p(x, \theta)$. В следующем пункте мы получим условия для ЛАН, одним из которых будет несколько ослабленное условие данного пункта, и покажем, что совокупность этих условий вытекает из условия квадратичной дифференцируемости корня из плотности.

7.1.3. Доказательство теоремы 7.1.1 (= теоремы 5.2.1). Мы рассматриваем постановку задачи, введенную в п. 7.1.1, не предполагая, в частности, что носитель плотности не зависит от параметра θ. Напомним, что P – распределение вероятностей на $(\mathbb{R}, \mathcal{B})$ ($\mathcal{B}$ – σ-алгебра борелевских множеств на прямой) с плотностью $p = p(x) = p(x, 0)$ и E – соответствующее математическое ожидание, т.е. для функции $f = f(x)$

$$Ef = \int f(x)p(x)\,dx.$$

Введем также обозначение: для $A \in \mathcal{B}$

$$E[f; A] = Ef\mathbf{1}_A = \int_A f(x)p(x)\,dx.$$

Условие D**.** Существует семейство борелевских множеств $\{A_\theta\}$, $\theta \in \Theta$, таких, что $P(A_\theta^c) = o(\theta^2)$ и $p(x,0) > 0$ при $x \in A_\theta$, и существует функция $h(x)$ такая, что $I := Eh^2 < \infty$, для которых

$$\int_{A_\theta} \left(\frac{p(x,\theta)}{p(x,0)} - 1 - \theta h(x)\right)^2 p(x,0)\,dx = o(\theta^2). \tag{7.1.33}$$

Условие D представляет собой модификацию условий предыдущего пункта, состоящую в ограничении области интегрирования в (7.1.7) множеством A_θ. Читатель может проверить, что в применении к гамма-плотности, рассмотренной в примере 5.2.1 (п. 5.2.3), интеграл (7.1.33) по области $A_\theta = (\theta, \infty)$ сходится при $\alpha > 2$, при этом, как показано в примере, $P(A_\theta^c) = o(\theta^2)$ при $\alpha > 2$. Условие $p(x,0) > 0$ при $x \in A_\theta$ исключает возможность неопределенности, связанной с появлением нулевых значений p в знаменателе.

При соответствующих изменениях в доказательстве предыдущего пункта (например, при определении множества A_θ (см. фразу, предшествующую (7.1.14)) следует добавить пересечение с множеством A_θ из условия D) это доказательство проходит при условии D за исключением того, что из условия D не следует, что $Eh = 0$ и $Er_\theta = 0$. Последнее равенство использовалось при получении оценки (7.1.25) и последующих аналогичных оценок. Нужные свойства будут получены при некотором дополнительном условии (условие E), и тогда условие D в совокупности с этим условием будут достаточными условиями для ЛАН. В дальнейшем будет показано, что условия D и E следуют из условий теоремы 7.1.1.

Обозначая, как и в предыдущем пункте, r_θ выражение в круглых скобках в (7.1.33), получаем аналог (7.1.9):

$$\frac{p_\theta}{p} = 1 + \theta h + r_\theta, \quad \text{где} \quad E[r_\theta^2; A_\theta] = o(\theta^2). \tag{7.1.34}$$

Для произвольного $0 < a < 1/2$ введем множества $A_{1\theta} = \{x\colon |r_\theta| < a\}$ и $A_{2\theta} = \{x\colon |\theta h| < a\}$. Так же, как в п. 7.1.2 (см. (7.1.11) и сопутствующие рассуждения), получаем

$$P(A_{1\theta}^c \cap A_\theta) = o(\theta^2), \quad P(A_{2\theta}^c) = o(\theta^2). \tag{7.1.35}$$

Поэтому, переопределяя множество A_θ как $A_\theta := A_\theta \cap A_{1\theta} \cap A_{2\theta}$, мы можем считать, что $|r_\theta| < a$ и $|\theta h| < a$ на множестве A_θ. Как следствие, из (7.1.34) имеем $|(p_\theta/p) - 1| < 2a$ на множестве A_θ.

Введем упомянутое выше дополнительное условие.

Условие E**.** Для всякого семейства множеств $\{B_\theta\}$ такого, что $P(B_\theta^c) = o(\theta^2)$, имеет место также $P_\theta(B_\theta^c) = o(\theta^2)$.

Лемма 7.1.1. *Если выполнены условия D и E, то*
1) $Eh = 0$,
2) $E[r_\theta; A_\theta] = o(\theta^2)$.

Доказательство. По условию D (см. (7.1.34)) на множестве A_θ

$$p_\theta = p(1 + \theta h + r_\theta), \quad E[r_\theta^2; A_\theta] = o(\theta^2). \tag{7.1.36}$$

Кроме того, по условию D $P(A_\theta) = 1 + o(\theta^2)$. Тогда по условию E также $P_\theta(A_\theta) = 1 + o(\theta^2)$. С другой стороны,

$$P_\theta(A_\theta) = \int_{A_\theta} (1 + \theta h + r_\theta) p\, dx = P(A_\theta) + \theta \int_{A_\theta} hp\, dx + \int_{A_\theta} r_\theta p\, dx.$$

Следовательно,

$$\theta \int_{A_\theta} hp\, dx + \int_{A_\theta} r_\theta p\, dx = o(\theta^2).$$

Рассмотрим каждое из двух слагаемых в левой части. Имеем $\int_{A_\theta} hp\, dx = Eh - \int_{A_\theta^c} hp\, dx$ и по неравенству Коши–Буняковского

$$\left| \int_{A_\theta^c} hp\, dx \right| \leqslant \left(\int_{A_\theta^c} h^2 p\, dx \right)^{1/2} \cdot \left(\int_{A_\theta^c} p\, dx \right)^{1/2} = o(\theta), \tag{7.1.37}$$

поскольку второй сомножитель в правой части есть $\sqrt{P(A_\theta^c)} = o(\theta)$, а первый сомножитель стремится к нулю. Далее,

$$\left| \int_{A_\theta} r_\theta p\, dx \right| = |E[r_\theta; A_\theta]| \leqslant \sqrt{E[r_\theta^2; A_\theta]} = o(\theta).$$

Подставляя две последние оценки в предыдущую формулу, получаем

$$\theta Eh + E[r_\theta; A_\theta] = o(\theta^2), \tag{7.1.38}$$

откуда, с учетом последней оценки, следует $Eh = o(1)$, а значит $Eh = 0$. Теперь из (7.1.38) получаем второе утверждение леммы, $E[r_\theta; A_\theta] = o(\theta^2)$. $\square$

Покажем, что из второго утверждения леммы 7.1.1 следует оценка (7.1.26). Как и в предыдущем пункте, обозначим $r'_\theta = r_\theta \mathbf{1}_{A_\theta}$. В доказательстве того пункта оценка (7.1.26) получалась переходом от $\zeta^{(2)\prime}_{N,t} = \sum_{i=1}^N r'_{\theta_i}$ к $\zeta^{(2)}_{N,t} = \sum_{i=1}^N r_{\theta_i}$. К последней сумме применима оценка (7.1.25) благодаря тому, что $Er_{\theta_i} r_{\theta_j} = Er_{\theta_i}(X_i) Er_{\theta_j}(X_j) = 0$ при $i \neq j$ ввиду независимости X_i и X_j и свойства $Er_\theta = 0$. Теперь же, пользуясь свойством $E[r_\theta; A_\theta] = o(\theta^2)$, мы покажем, что $E(\zeta^{(2)\prime}_{N,t})^2 \to 0$, откуда будет следовать требуемое соотношение (7.1.26). Воспользуемся известным равенством

$$E\left(\sum r'_{\theta_i}\right)^2 = \operatorname{var}\left(\sum r'_{\theta_i}\right) + \left(E\sum r'_{\theta_i}\right)^2.$$

Оценим каждое из слагаемых в правой части этого равенства. Из условия $E(r'_\theta)^2 = E[r_\theta^2; A_\theta] = o(\theta^2)$ получаем

$$\operatorname{var}\left(\sum r'_{\theta_i}\right) = \sum \operatorname{var}(r'_{\theta_i}) \leqslant \sum E(r'_{\theta_i})^2 = \sum o(\theta_i^2) = o(1).$$

В силу п. 2 леммы 7.1.1

$$E\sum r'_{\theta_i} = \sum Er'_{\theta_i} = \sum o(\theta_i^2) = o(1).$$

Из полученных оценок следует $E(\zeta^{(2)\prime}_{N,t})^2 \to 0$ и, как следствие, (7.1.26).

Нетрудно проверить, что если выполнено условие D, оценка (7.1.26) и $Eh = 0$, то повторение доказательства предыдущего пункта приводит к требуемому результату, т.е. из условий D и E следует свойство ЛАН. Наше доказательство теоремы 7.1.1 завершают две леммы, утверждающие, что из условий этой теоремы следуют условия D и E.

Напомним формулу, выражающую основное предположение теоремы 7.1.1:

$$\int\left(\frac{\sqrt{p_\theta} - \sqrt{p}}{\theta} - g(x)\right)^2 dx \to 0. \tag{7.1.39}$$

Обозначим $h = 2g/\sqrt{p}$. Предположение теоремы $\int g^2(x)\,dx < \infty$ равносильно условию $Eh^2 < \infty$. (В случае дифференцируемости плотности в обычном смысле $g(x) = p^{(1)}/(2\sqrt{p})$, $h = p^{(1)}/p = l^{(1)}$.) Из (7.1.39) следует

$$E\left(\frac{1}{\theta}\left(\sqrt{\frac{p_\theta}{p}}-1\right)-\frac{1}{2}h\right)^2 \to 0. \tag{7.1.40}$$

(Отметим, что это соотношение не эквивалентно (7.1.39): оно равносильно (7.1.39) с интегралом в левой части, взятым по области $\{x\colon p(x) > 0\}$.)

Положим

$$X_0 = \frac{1}{2}h, \quad X_\theta = \frac{1}{\theta}\left(\sqrt{\frac{p_\theta}{p}}-1\right).$$

Тогда (7.1.40) можно записать как

$$E(X_\theta - X_0)^2 \to 0.$$

Для произвольного $0 < a < 1/2$ положим $A_{1\theta} = \{x\colon |\theta(X_\theta - X_0)| < a\}$, $A_{2\theta} = \{x\colon |\theta X_0| < a\}$ и $A_\theta = A_{1\theta} \cap A_{2\theta}$. Аналогично (7.1.35) имеем $P(A_\theta^c) = o(\theta^2)$. Очевидно, $\theta|X_\theta| < 2a$ на множестве A_θ.

Лемма 7.1.2. *При условиях теоремы 7.1.1 выполнено условие D с определенным выше семейством $\{A_\theta\}$.*

Доказательство. Напомним, что для доказательства леммы нужно показать, что

$$E\left[\left(\frac{1}{\theta}\left(\frac{p_\theta}{p}-1\right)-h\right)^2; A_\theta\right] \to 0, \tag{7.1.41}$$

см. (7.1.33). Легко проверяется равенство

$$\frac{1}{\theta}\left(\frac{p_\theta}{p}-1\right)-h = 2(X_\theta - X_0)+\theta X_\theta^2 = 2(X_\theta - X_0)+\theta X_0^2+\theta(X_\theta^2 - X_0^2).$$

Для доказательства (7.1.41) достаточно проверить выполнение соответствующего соотношения для каждого слагаемого в правой части. Для первого слагаемого это, как было сказано, эквивалент формулы (7.1.40) (сужение на множество A_θ только уменьшает математическое ожидание). Рассмотрим последнее слагаемое. Имеем $\theta|X_\theta^2 - X_0^2| = \theta|X_\theta + X_0||X_\theta - X_0| \leqslant 3a|X_\theta - X_0|$ на множестве A_θ, следовательно

$$E[\theta^2(X_\theta^2 - X_0^2)^2; A_\theta] \leqslant 9a^2 E[(X_\theta - X_0)^2; A_\theta] \to 0.$$

Остается показать, что $\theta^2 E[X_0^4; A_\theta] \to 0$. Учитывая определение $X_0 = h/2$ и неравенство $|\theta X_0| < a$ на A_θ, получаем $\theta^2 E[X_0^4; A_\theta] \leqslant (a/8)|\theta|E[|h|^3; A_\theta]$ и требуемое соотношение $|\theta|E[|h|^3; A_\theta] \to 0$ доказывается так же, как соотношение $E|\theta h'_\theta|^3 = o(\theta^2)$ в предыдущем пункте (см. (7.1.32)). (Напомним, что мы обозначали там $h'_\theta = h\mathbf{1}_{A_\theta}$.) Этим лемма 7.1.2 доказана. □

Лемма 7.1.3. *При условиях теоремы* 7.1.1 *выполнено условие* E.

Доказательство. Пусть B_θ – произвольная система множеств такая, что $P(B_\theta^c) = o(\theta^2)$. Требуется доказать, что $P_\theta(B_\theta^c) = o(\theta^2)$.

Обозначим $f_\theta = \sqrt{p} + \frac{1}{2}\theta h\sqrt{p}$ и $\tau_\theta = \sqrt{p_\theta} - f_\theta$. В этих обозначениях наша цель – доказать, что

$$\int_{B_\theta^c} (\tau_\theta + f_\theta)^2\, dx = o(\theta^2).$$

Пользуясь неравенством $(a+b)^2 \leqslant 2(a^2+b^2)$, покажем для этого, что

$$\int_{B_\theta^c} \tau_\theta^2\, dx = o(\theta^2), \quad \int_{B_\theta^c} f_\theta^2\, dx = o(\theta^2).$$

Используя введенное ранее обозначение $h = 2g/\sqrt{p}$, (7.1.39) можно переписать как $\int \tau_\theta^2\, dx = o(\theta^2)$, откуда следует первое из этих соотношений. В силу того же неравенства второе соотношение следует из

$$\int_{B_\theta^c} p\, dx = o(\theta^2), \quad \theta^2 \int_{B_\theta^c} h^2 p\, dx = o(\theta^2).$$

Первое из этих соотношений выражает условие $P(B_\theta^c) = o(\theta^2)$. Второе следует из того, что интеграл от интегрируемой функции по множеству бесконечно малой меры стремится к нулю. Лемма 7.1.3 доказана. □

Таким образом, при условиях теоремы 7.1.1 выполняются условия D и E, из которых следует свойство ЛАН. Этим доказательство теоремы 7.1.1 завершено. □

§ 7.2. Конечность информации Фишера

В этом параграфе мы покажем, что из условий теоремы 5.2.2 следует выполнение условий теоремы 7.1.1 и тем самым свойство ЛАН.

7.2.1. Лемма. Предварительно мы докажем лемму, которая потребуется для доказательства теоремы.

Эта лемма дает условие, при котором из абсолютной непрерывности функции $g(y) \geqslant 0$ следует абсолютная непрерывность функции $\sqrt{g(y)}$.

Прежде всего напомним определение абсолютной непрерывности и его основное следствие (см. Колмогоров и Фомин [6], глава VI, § 4). Функция $f(y)$, $y \in [a, b]$, называется абсолютно непрерывной, если для любого $\varepsilon > 0$ найдется $\delta > 0$ такое, что для любого конечного или счетного набора непересекающихся интервалов, принадлежащих отрезку $[a, b]$, суммарной длины, меньшей δ, суммарное приращение функции $f(y)$ на этих интервалах меньше ε, т.е. из $(a_i, b_i) \subset [a, b]$, $i = 1, 2, \ldots$, $(a_i, b_i) \cap (a_j, b_j) = \varnothing$ для любых $i \neq j$ и $\sum_i (b_i - a_i) < \delta$ следует $\sum_i |f(b_i) - f(a_i)| < \varepsilon$. Абсолютно непрерывная функция $f(y)$ дифференцируема почти всюду (по мере Лебега) на $[a, b]$, т.е. при почти всех $y \in [a, b]$ существует производная $f'(y)$, и приращение $f(y)$ на любом интервале $[\alpha, \beta] \subset [a, b]$ равно интегралу от ее производной, т.е.

$$f(\beta) - f(\alpha) = \int_\alpha^\beta f'(y)\, dy.$$

Абсолютная непрерывность является необходимым и достаточным условием выполнения этого свойства.

Прокомментируем следующее ниже доказательство леммы. Из определения абсолютной непрерывности сразу следует, что если функция $h(\cdot)$ определена и дифференцируема всюду на области значений функции $f(y)$, $y \in [a, b]$, и производная h' ограничена в этой области, $h' < C$, то функция $h(f(y))$ также абсолютно непрерывна (поскольку $|h(f(b_i)) - h(f(a_i))| \leqslant C|f(b_i) - f(a_i)|$). Следовательно, если $g(y)$ ограничена снизу положительной константой на каком-нибудь интервале, то функция $\sqrt{g(y)}$ тоже абсолютно непрерывна на этом интервале. Поэтому «опасность» представляют малые значения $g(y)$, т.е. значения вблизи точек, где $g(y) = 0$. Ключевым утверждением в доказательстве является формула (7.2.4), говорящая, что если g положительна внутри интервала (α, β), то интеграл от производной функции $\sqrt{g(y)}$ по этому интервалу равен приращению этой функции на данном интервале. В частности, если g обращается в нуль в точках α и β, то интеграл от производной функции $\sqrt{g(y)}$ по (α, β) равен нулю.

Лемма 7.2.1. *Пусть функция* $g(y) \geqslant 0$ *абсолютно непрерывна на интервале* $[a, b]$ *и*

$$\int_a^b \frac{|g'(y)|}{2\sqrt{g(y)}}\, dy < \infty \tag{7.2.1}$$

(*подынтегральная функция считается равной нулю, если* $g(y) = 0$). *Тогда* $\sqrt{g(y)}$ *тоже абсолютно непрерывна на* $[a, b]$.

Доказательство. Необходимым и достаточным условием абсолютной непрерывности функции является то, что она равна интегралу от своей производной. В силу этого нам нужно показать, что для любого $x \in [a, b]$

$$\sqrt{g(x)} - \sqrt{g(a)} = \int_a^x \frac{g'(y)}{2\sqrt{g(y)}}\, dy. \tag{7.2.2}$$

Возьмем точку $y_1 \in (a, x)$ такую, что $g(y_1) > 0$. Функция $g(y)$ непрерывна на интервале $[a, b]$ (как следствие абсолютной непрерывности), поэтому найдется интервал $[\alpha, \beta]$, $a \leqslant \alpha < y_1 < \beta \leqslant x$, такой, что $g(y) \geqslant c > 0$ при $\alpha \leqslant y \leqslant \beta$. Как отмечалось выше, функция $\sqrt{g(y)}$ абсолютно непрерывна в области, где $g(y)$ ограничена снизу положительной константой. Поэтому $\sqrt{g(y)}$ абсолютно непрерывна на $[\alpha, \beta]$ и

$$\sqrt{g(\beta)} - \sqrt{g(\alpha)} = \int_\alpha^\beta \frac{g'(y)}{2\sqrt{g(y)}}\, dy. \tag{7.2.3}$$

Будем теперь увеличивать β пока $g(y)$ остается положительной на $[\alpha, \beta]$. Тогда есть две возможности: либо $g(y) > 0$ при $\alpha < y < x$ (и $g(x) \geqslant 0$), либо имеется точка $\beta_1 < x$, где g впервые обращается в нуль, т.е. $g(\beta_1) = 0$ и $g(y) > 0$ при $\alpha < y < \beta_1$. В первом случае полагаем $\beta_1 = x$. Точно так же, уменьшая α, приходим к точке α_1, которая либо равна a и тогда $g(y) > 0$ при $a < y < \beta_1$ (и $g(a) \geqslant 0$), либо $\alpha_1 > a$, $g(\alpha_1) = 0$ и $g(y) > 0$ при $\alpha_1 < y < \beta_1$. Заметим теперь, что $\sqrt{g(\beta)} - \sqrt{g(\alpha)} \to \sqrt{g(\beta_1)} - \sqrt{g(\alpha_1)}$ при $\alpha \downarrow \alpha_1$ и $\beta \uparrow \beta_1$ в силу непрерывности функции $\sqrt{g(y)}$. Покажем, что интеграл в правой части (7.2.3) при этом предельном переходе сходится к интегралу по $[\alpha_1, \beta_1]$. Обозначим для краткости подынтегральную функцию в правой части (7.2.3) через $f(y)$ и запишем этот интеграл как $\int_{\alpha_1}^{\beta_1} f(y)\mathbf{1}_{[\alpha,\beta]}(y)\,dy$. Тогда $f(y)\mathbf{1}_{[\alpha,\beta]}(y) \to f(y)$ для всех $y \in (\alpha_1, \beta_1)$ при $\alpha \downarrow \alpha_1$, $\beta \uparrow \beta_1$ и функции $|f(y)|\mathbf{1}_{[\alpha,\beta]}(y)$ ограничены сверху подынтегральной функцией в (7.2.1). Следовательно, рассматриваемый интеграл сходится к $\int_{\alpha_1}^{\beta_1} f(y)\,dy$ по теореме Лебега о мажорированной сходимости. Отсюда следует, что

$$\sqrt{g(\beta_1)} - \sqrt{g(\alpha_1)} = \int_{\alpha_1}^{\beta_1} \frac{g'(y)}{2\sqrt{g(y)}}\,dy. \tag{7.2.4}$$

Если $\alpha_1 = a$ и $\beta_1 = x$, т.е. $g(y) > 0$ при всех $a < y < x$, то формула (7.2.2) доказана. Предположим поэтому, что либо α_1, либо β_1 не совпадают с соответствующими концами интервала $[a, x]$. Заметим, что тогда

$$\sqrt{g(\beta_1)} - \sqrt{g(\alpha_1)} = \begin{cases} -\sqrt{g(a)}, & \text{если } \alpha_1 = a, \\ \sqrt{g(x)}, & \text{если } \beta_1 = x, \\ 0, & \text{если } \alpha_1 \neq a, \ \beta_1 \neq x. \end{cases}$$

Пользуясь этой формулой, нетрудно проверить, что если $g(y) \equiv 0$ при $y \in [a, x] \setminus [\alpha_1, \beta_1]$, то также выполняется (7.2.2). Поэтому рассмотрим случай, когда найдется точка $y_2 \in [a, x] \setminus [\alpha_1, \beta_1]$, для которой $g(y_2) > 0$. Повторяя предыдущие рассуждения, найдем интервал $[\alpha_2, \beta_2]$ с теми же свойствами. Продолжая таким же образом, получим, что множество $\{y \in (a, x)\colon g(y) > 0\}$ представляется как конечное или счетное объединение непересекающихся интервалов (α_i, β_i) таких, что $g(\alpha_i) = g(\beta_i) = 0$ за исключением самое большее одного интервала с $\alpha_i = a$ и одного интервала с $\beta_j = x$. Поэтому

$$\int_a^x \frac{g'(y)}{2\sqrt{g(y)}}\,dy = \sum_i \int_{\alpha_i}^{\beta_i} \frac{g'(y)}{2\sqrt{g(y)}}\,dy = \sqrt{g(x)} - \sqrt{g(a)},$$

что и требовалось доказать. □

7.2.2. Теорема об информации Фишера. Теперь мы переходим к доказательству теоремы 5.2.2, формулировка которой повторяется ниже как теорема 7.2.1. Мы рассматриваем постановку задачи, описанную в п. 7.1.1. Напомним, что независимые наблюдения $X_1, \dots, X_N$ имеют плотности распределения, принадлежащие семейству $p_\theta(x)$, $\theta \in \Theta$, где Θ – множество на прямой, содержащее интервал вида $\Delta = [-\delta, \delta]$, $\delta > 0$.

Теорема 7.2.1. *Предположим, что плотность $p_\theta(x)$ удовлетворяет следующим условиям*:

(A) *При каждом $x \in \mathbb{R}$ плотность $p_\theta(x)$ абсолютно непрерывна по θ в некоторой окрестности точки $\theta = 0$;*

(B) *При каждом θ из этой окрестности производная $p_\theta^{(1)}(x) = \frac{\partial}{\partial\theta} p_\theta(x)$ существует при почти всех (по мере Лебега) $x \in \mathbb{R}$;*

(C) *Функция*

$$I(\theta) = E_\theta \left(\frac{p_\theta^{(1)}}{p_\theta} \right)^2 < \infty,$$

положительна и непрерывна в этой окрестности.

Тогда выполнены условия теоремы 7.1.1 с $g(x) = p_0^{(1)}(x) / \bigl(2 \times \sqrt{p_0(x)}\,\bigr)$ на множестве $\{x\colon p_0(x) > 0\}$ и $g(x) = 0$, если $p_0(x) = 0$.

Доказательство. Нам нужно показать, что при условиях (A)–(C) выполнено соотношение (7.1.3) с функцией $g(x)$, определенной в формулировке теоремы. Для удобства напомним это соотношение:

$$\int \left(\frac{\sqrt{p(x,\theta)} - \sqrt{p(x,0)}}{\theta} - g(x) \right)^2 dx \to 0 \ \text{при} \ \theta \to 0. \tag{7.2.5}$$

Мы докажем сначала, что

$$E \left[\frac{1}{\theta} \left(\sqrt{\frac{p_\theta}{p}} - 1 \right) - \frac{1}{2} \frac{p^{(1)}}{p} \right]^2 \to 0. \tag{7.2.6}$$

Это соотношение эквивалентно (7.2.5) с интегралом, взятым по области $\{x\colon p_0(x) > 0\}$. Далее мы покажем, что

$$\frac{1}{\theta^2} \int_{\{x\colon p_0(x) = 0\}} p_\theta \, dx \to 0. \tag{7.2.7}$$

Левая часть (7.2.5) равна сумме левых частей (7.2.6) и (7.2.7) и тем самым (7.2.5) будет доказано.

Обозначим для краткости

$$\zeta(\theta) = \sqrt{\frac{p_\theta}{p}} - 1, \quad h = \frac{p^{(1)}}{p}.$$

В этих обозначениях (7.2.6) запишется как

$$E\left(\frac{1}{\theta}\zeta(\theta)-\frac{1}{2}h\right)^2 \to 0. \tag{7.2.8}$$

По теореме Витали (теорема 8.2.7) с $r=2$ для (7.2.8) достаточно выполнения двух условий:

$$\limsup_{\theta\to 0}\frac{1}{\theta^2}E\zeta^2(\theta) \leqslant \frac{1}{4}Eh^2 = \frac{1}{4}I(0), \tag{7.2.9}$$

$$\frac{1}{\theta}\zeta(\theta) \xrightarrow{p} \frac{1}{2}h \tag{7.2.10}$$

(из последнего соотношения, как легко проверить, следует условие (B) теоремы 8.2.7).

По условию (B) доказываемой теоремы функция $p_\theta(x)$ дифференцируема по θ в точке $\theta=0$ при почти всех x. Мы рассматриваем область, где $p_0(x)>0$, поэтому функция $\sqrt{p_\theta(x)}$ также дифференцируема по θ в точке $\theta=0$ при почти всех x в этой области. Это значит, что

$$\sqrt{p}\frac{1}{\theta}\zeta(\theta) = \frac{\sqrt{p_\theta}-\sqrt{p}}{\theta} \to \frac{p^{(1)}}{2\sqrt{p}} = \frac{\sqrt{p}}{2}h$$

при почти всех x, откуда следует (7.2.10).

Переходя к доказательству (7.2.9), покажем предварительно, что функция $\sqrt{p_\theta(x)}$ абсолютно непрерывна по $\theta\in\Delta$ при почти всех $x\in\mathbb{R}$, где $\Delta=(-\delta,\delta)$ – окрестность точки $\theta=0$, фигурирующая в условиях теоремы 7.2.1.

По условию (C) (конечность и непрерывность $I(\theta)$)

$$\int_\Delta I(\theta)\,d\theta < \infty.$$

Подставляя выражение для $I(\theta)$ и меняя порядок интегрирования, получаем

$$\int_\Delta\left(\int_{\mathbb{R}}\left(\frac{p_\theta^{(1)}}{p_\theta}\right)^2 p_\theta\,dx\right)d\theta = \int_{\mathbb{R}}\left(\int_\Delta \frac{(p_\theta^{(1)})^2}{p_\theta}\,d\theta\right)dx < \infty.$$

Отсюда следует, что

$$\int_\Delta \frac{(p_\theta^{(1)})^2}{p_\theta}\,d\theta < \infty$$

при почти всех $x \in \mathbb{R}$ (по мере Лебега). Следовательно,

$$\int_{\Delta} \frac{|p_\theta^{(1)}|}{\sqrt{p_\theta}}\, d\theta < \infty$$

при почти всех $x \in \mathbb{R}$. Тогда по лемме 7.2.1 функция $\sqrt{p_\theta(x)}$ абсолютно непрерывна по θ при почти всех $x \in \mathbb{R}$.

Перейдем к доказательству (7.2.9). Имеем

$$\begin{aligned} \frac{1}{\theta^2} E\zeta^2(\theta) &= \frac{1}{\theta^2} \int_{p\neq 0} \left(\frac{\sqrt{p_\theta} - \sqrt{p}}{\sqrt{p}} \right)^2 p\, dx = \frac{1}{\theta^2} \int_{p\neq 0} (\sqrt{p_\theta} - \sqrt{p})^2\, dx \\ &\leqslant \frac{1}{\theta^2} \int_{\mathbb{R}} (\sqrt{p_\theta} - \sqrt{p})^2\, dx. \end{aligned} \tag{7.2.11}$$

Отметим, что разность между правой частью (7.2.11) и предыдущим выражением равна левой части (7.2.7). Это замечание будет использовано для доказательства (7.2.7).

Воспользуемся доказанной выше абсолютной непрерывностью $\sqrt{p_\theta}$ при почти всех x, вследствие которой

$$\sqrt{p_\theta} - \sqrt{p} = \int_0^\theta \frac{p_\tau^{(1)}}{2\sqrt{p_\tau}}\, d\tau$$

при почти всех x. Следовательно, правая часть в (7.2.11) равна

$$\int \left(\int_0^\theta \frac{p_\tau^{(1)}}{2\sqrt{p_\tau}} \frac{d\tau}{\theta} \right)^2 dx \leqslant \int \left(\int_0^\theta \left(\frac{p_\tau^{(1)}}{2\sqrt{p_\tau}} \right)^2 \frac{d\tau}{\theta} \right) dx.$$

Неравенство объясняется тем, что внутренний интеграл в левой части можно рассматривать как усреднение (математическое ожидание) функции $\frac{p_\tau^{(1)}}{2\sqrt{p_\tau}}$ по равномерному распределению на отрезке $[0,\theta]$, так что подынтегральная функция представляет собой квадрат математического ожидания, который не превосходит математического ожидания квадрата, стоящего в правой части. Меняя порядок интегрирования, получаем, что последний интеграл равен

$$\frac{1}{4\theta} \int_0^\theta I(\tau)\, d\tau \to \frac{1}{4} I(0) \tag{7.2.12}$$

вследствие непрерывности функции $I(\theta)$ (условие (C)). Из полученных оценок следует (7.2.9). Таким образом, мы проверили условия (7.2.9) и (7.2.10), из которых по теореме Витали следует (7.2.8), или, что то же, (7.2.6).

Как уже отмечалось, левая часть (7.2.7) равна разности интегралов во 2-й и 1-й строках (7.2.11). Но из доказанного нами соотношения (7.2.8) следует, что $\lim_{\theta\to 0} \frac{1}{\theta^2} E\zeta^2(\theta) = \frac{1}{4}Eh^2 = \frac{1}{4}I(0)$, поэтому все промежуточные выражения в цепочке неравенств между (7.2.11) и (7.2.12) стремятся к тому же пределу, а разности между ними стремятся к нулю. Отсюда следует (7.2.7). Этим доказательство теоремы 7.2.1 (= теоремы 5.2.2) завершено. $\square$

Глава 8. Приложение

В данном Приложении мы напомним и проиллюстрируем основные понятия, связанные с задачами статистической проверки гипотез, а также приведем некоторые вспомогательные результаты.

§ 8.1. Задача проверки гипотез

8.1.1. Постановка задачи. Начнем с задачи проверки простой гипотезы против простой альтернативы. Пусть имеется измеримое пространство $(\mathbb{X}, \mathcal{A})$, т.е. множество $\mathbb{X}$ с σ-алгеброй $\mathcal{A}$ его подмножеств, и на этом пространстве заданы две вероятностных меры P и Q. В распоряжении статистика имеется наблюдение случайного элемента X пространства $\mathbb{X}$, подчиняющегося одному из распределений вероятностей P или Q. (В задачах, рассматриваемых в связи с применением ранговых статистик, обычно $\mathbb{X} = \mathbb{R}^N$ и X это N-мерный вектор наблюдений.) По наблюдению X требуется проверить гипотезу H_0 о том, что распределение случайного элемента X есть P, против альтернативной гипотезы H_1, что это распределение есть Q.

Формально эта задача ставится следующим образом. Считается заданным некоторое (малое) число $\alpha > 0$, называемое уровнем значимости или размером критерия, и выбирается *критическое множество*[1] $S \subset \mathbb{X}$ такое, что $P(X \in S) \leqslant \alpha$[2]. Тогда критерий (решающее правило), определяемый критическим множеством S, состоит в том, что если результат наблюдений $X \in S$, то гипотеза H_0 отвергается, т.е. принимается решение, что X имеет распределение Q. В противном случае, если $X \in \mathbb{X} \setminus S$, принимается

[1] Мы излагаем задачу проверки гипотез в несколько упрощенном виде, в частности, мы не рассматриваем рандомизированные критерии, а также приводим лемму Неймана–Пирсона в упрощенном варианте. Полное изложение можно найти в книге Э. Лемана [7].

[2] Допуская некоторую вольность в обозначениях, мы пишем вероятность события A как $P(A)$ и как $P(X \in A)$ (вероятность попадания случайного элемента X в множество A).

гипотеза H_0. Если имеет место гипотеза H_0 и происходит событие $\{X \in S\}$, то согласно нашему правилу мы отвергаем H_0, совершая тем самым ошибку, называемую *ошибкой 1-го рода.* Условие $P(X \in S) \leqslant \alpha$ ограничивает сверху вероятность этой ошибки заданным уровнем α. Если же верна H_1, то исход $X \in \mathbb{X} \setminus S$ приводит к принятию H_0 и в таком случае мы совершаем *ошибку 2-го рода*[3]. Таким образом, вероятность ошибки 2-го рода равна $Q(X \in \mathbb{X} \setminus S)$ и задача состоит в выборе множества S, минимизирующего эту вероятность в классе множеств, удовлетворяющих условию на вероятность ошибки 1-го рода.

Часто, однако, вместо ошибки 2-го рода рассматривают дополнительное событие $X \in S$, которое приводит к решению принять H_1, т.е. к принятию правильного решения, когда верна H_1. Вероятность такого исхода при гипотезе H_1 равна

$$Q(X \in S) = 1 - Q(X \in \mathbb{X} \setminus S),$$

т.е. равна дополнению вероятности ошибки 2-го рода до 1. Эту вероятность называют *мощностью* критерия. Ясно, что задача минимизации вероятности ошибки 2-го рода равносильна задаче максимизации мощности. Мы используем для мощности обозначение β, так что $\beta = Q(X \in S)$, при необходимости снабжая этот символ индексами или аргументами.

8.1.2. Лемма Неймана–Пирсона. Решение поставленной выше задачи дается известной леммой Неймана–Пирсона, которую мы напомним здесь в несколько упрощенных предположениях. Мы будем предполагать, что распределения P и Q имеют плотности p и q относительно некоторой σ-конечной меры ν на $(\mathbb{X}, \mathcal{A})$, т.е. $P(A) = \int_A p\, d\nu$, $Q(A) = \int_A q\, d\nu$ для всякого множества $A \in \mathcal{A}$. Кроме того, будем предполагать, что p и q отличны от нуля на одном и том же множестве, т.е. $\{x \in \mathbb{X}\colon p(x) > 0\} = \{x \in \mathbb{X}\colon q(x) > 0\}$ (это значит, что меры P и Q взаимно абсолютно непрерывны). Тогда на этом множестве определено отношение правдоподобия (ОП)

$$L(x) = \frac{q(x)}{p(x)}. \tag{8.1.1}$$

Еще одно упрощающее допущение будет состоять в предположении, что ОП имеет непрерывную функцию распределения $F(z) =$

[3] В определенных приложениях гипотезы H_0 и H_1 соответствуют отсутствию и наличию цели и тогда ошибка 1-го рода это «ложная тревога», а ошибка 2-го рода – «пропуск цели».

$P(L(X) < z) = P\{x\colon L(x) < z\}$, т.е. $P\{x\colon L(x) = z\} = 0$ для всякого $z \in \mathbb{R}$. (В силу взаимной абсолютной непрерывности P и Q тем же свойством обладает функция распределения L относительно Q.) Определим число c_α равенством

$$F(c_\alpha) = 1 - \alpha. \tag{8.1.2}$$

(Такое c_α существует в силу предположения о непрерывности $F(z)$.)[4] Определим множество

$$S_\alpha = \{x\colon L(x) > c_\alpha\}. \tag{8.1.3}$$

Из (8.1.2) следует, что $P(S_\alpha) = \alpha$. Лемма Неймана–Пирсона утверждает, что множество S_α задает наиболее мощный критерий уровня α.

Лемма 8.1.1 (Неймана–Пирсона). *Пусть S – произвольное измеримое множество такое, что $P(S) \leqslant \alpha$. Тогда*

$$Q(S) \leqslant Q(S_\alpha).$$

Доказательство. Учитывая равенства $Q(S) = Q(S \cap S_\alpha) + Q(S \setminus S_\alpha)$, $Q(S_\alpha) = Q(S \cap S_\alpha) + Q(S_\alpha \setminus S)$ и равенство $q = Lp$ (см. (8.1.1)), запишем разность этих вероятностей как

$$\begin{aligned} Q(S_\alpha) - Q(S) &= Q(S_\alpha \setminus S) - Q(S \setminus S_\alpha) \\ &= \left(\int_{S_\alpha \setminus S} - \int_{S \setminus S_\alpha}\right) L(x) p(x)\, d\nu. \end{aligned} \tag{8.1.4}$$

Точно так же из условий $P(S_\alpha) = \alpha$ и $P(S) \leqslant \alpha$ имеем

$$\left(\int_{S_\alpha \setminus S} - \int_{S \setminus S_\alpha}\right) p(x)\, d\nu \geqslant 0. \tag{8.1.5}$$

Умножим (8.1.5) на c_α и вычтем из (8.1.4). Получим

$$Q(S_\alpha) - Q(S) \geqslant \left(\int_{S_\alpha \setminus S} - \int_{S \setminus S_\alpha}\right) (L(x) - c_\alpha) p(x)\, d\nu.$$

[4] Значение c_α может определяться неоднозначно, если $F(z)$ имеет горизонтальный участок, т.е. $F(z) = 1 - \alpha$ при $a < z < b$. Но тогда вероятность попадания $L(X)$ в интервал $[a, b]$ равна нулю как по мере P, так и по мере Q, и выбор c_α в интервале $[a, b]$ не влияет на вероятности рассматриваемых далее событий.

Из определения (8.1.3) следует, что $L(x) - c_\alpha \geqslant 0$ на множестве $S_\alpha \setminus S$ и $L(x) - c_\alpha \leqslant 0$ на множестве $S \setminus S_\alpha$. Таким образом, $Q(S_\alpha) - Q(S)$ представляется как разность двух интегралов, из которых первый – неотрицательный, а второй – неположительный, откуда следует утверждение леммы. □

Замечание 8.1.1. Алгоритм, предлагаемый леммой Неймана–Пирсона, состоит из двух этапов: на первом строится семейство множеств $\{L(x) > c\}$, из которого следует выбирать искомую критическую область, а второй состоит в выборе множества из этого семейства, имеющего требуемую P-вероятность α. При практическом применении леммы Неймана–Пирсона используется тот факт, что *семейство* множеств $\{L(x) > c\}$, $c \in \mathbb{R}$, не меняется при переходе от $L(x)$ к $f(L(x))$, где $f(\cdot)$ – непрерывная строго возрастающая функция. Например, множества $\{\log L(x) > c\}$, $c \in \mathbb{R}$, образуют то же *семейство*, что и исходное (хотя при каждом конкретном c множества $\{L(x) > c\}$ и $\{\log L(x) > c\}$, разумеется, различны). Поэтому в конкретных задачах к ОП $L(x)$ применяются монотонные преобразования $f(L(x))$ с целью привести неравенство к наиболее удобному виду для вычисления его вероятности.

8.1.3. Пример 1: параметр сдвига нормального распределения. Проиллюстрируем применение леммы Неймана–Пирсона на простейшем примере семейства нормальных распределений, зависящих от параметра сдвига. Предположим, что по выборке $X_1, \ldots, X_N$, состоящей из независимых одинаково распределенных (н.о.р.) случайных величин, имеющих нормальное распределение $\mathcal{N}(\mu, \sigma^2)$ с известным σ^2, проверяется гипотеза $H_0\colon \mu = \mu_0$, где $\mu_0 \in \mathbb{R}$ – некоторое заданное значение, против односторонних альтернатив $H_1\colon \mu > \mu_0$. Будем обозначать через P_μ распределение вероятностей вектора $\mathbf{X} = (X_1, \ldots, X_N)$ и через E_μ соответствующее математическое ожидание. Кроме того, символом $\mathcal{L}_\mu(T)$ будет обозначаться распределение случайной величины $T = T(\mathbf{X})$.

В данном случае альтернатива H_1 *сложная*, т.е. она определяет не единственное значение параметра, а *множество* значений $\{\mu > \mu_0\}$. Зафиксируем некоторое значение μ_1 из этого множества и рассмотрим проверку простой гипотезы H_0 против простой альтернативы $H_1'\colon \mu = \mu_1$. Для этого применим лемму Неймана–Пирсона (см. п. 8.1.2).

Выпишем явно плотности p, q и отношение правдоподобия L, фигурирующие в этой лемме. Распределению P соответствует в нашем случае распределение P_{μ_0} с плотностью

$$p(\mathbf{x}) = \varphi_N(\mathbf{x}; \mu_0, \sigma^2) = \prod_{i=1}^{N} \varphi(x_i; \mu_0, \sigma^2), \tag{8.1.6}$$

где

$$\varphi(x; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi}}\, e^{-\frac{(x-\mu)^2}{2\sigma^2}} \tag{8.1.7}$$

обозначает плотность нормального распределения $\mathcal{N}(\mu, \sigma^2)$. Распределению Q леммы 8.1.1 отвечает распределение P_{μ_1} с плотностью

$$q(\mathbf{x}) = \varphi_N(\mathbf{x}; \mu_1, \sigma^2) = \prod_{i=1}^{N} \varphi(x_i; \mu_1, \sigma^2). \tag{8.1.8}$$

По лемме Неймана–Пирсона критическая область наиболее мощного критерия проверки H_0 против H_1 имеет вид $\{L(\mathbf{x}) > c\}$, где $L(\mathbf{x}) = q(x)/p(\mathbf{x})$ и c надлежит выбрать так, чтобы P_{μ_0}-вероятность этой области равнялась заданному $\alpha > 0$. Введем $\Lambda(\mathbf{x}) = \log L(\mathbf{x})$. Как уже говорилось, мы можем искать критическую область в виде $\{\Lambda(\mathbf{x}) > c\}$ с тем же условием на (неопределенную пока) константу c. Запишем $\Lambda(\mathbf{x})$ в виде

$$\Lambda = \sum_{i=1}^{N} \big(\log \varphi(x_i; \mu_1, \sigma^2) - \log \varphi(x_i; \mu_0, \sigma^2)\big). \tag{8.1.9}$$

Введем нормированный параметр $t = (\mu_1 - \mu_0)\sqrt{N}$. Используя выражение (8.1.7), запишем разность логарифмов в виде

$$\log \varphi(x_i; \mu_1, \sigma^2) - \log \varphi(x_i; \mu_0, \sigma^2) = \frac{t}{\sigma^2\sqrt{N}}(x_i - \mu_0) - \frac{t^2}{2\sigma^2 N}. \tag{8.1.10}$$

Мы будем рассматривать L и Λ как случайные величины, определенные на пространстве $\mathbb{R}^N$, и для наглядности записывать их аргумент в виде случайной точки $\mathbf{X} = (X_1, \ldots, X_N)$, подчиняющейся распределению P_{μ_0} или P_{μ_1}. Тогда

$$\Lambda = \frac{t}{\sigma^2\sqrt{N}} \sum_{1}^{N} (X_i - \mu_0) - \frac{1}{2\sigma^2} t^2. \tag{8.1.11}$$

Видно, что

$$\mathcal{L}(\Lambda \mid P_{\mu_0}) = \mathcal{N}\left(-\frac{1}{2\sigma^2}t^2, \frac{1}{\sigma^2}t^2\right). \tag{8.1.12}$$

Чтобы подчеркнуть, что последующие выводы полностью определяются значениями параметров этого нормального распределения, а не их конкретными выражениями, обозначим $A = t/\sigma$. Тогда (8.1.12) перепишется как

$$\mathcal{L}(\Lambda \mid P_{\mu_0}) = \mathcal{N}\big(-\frac{1}{2}A^2, A^2\big). \tag{8.1.13}$$

То, что между средним $\mu = -\frac{1}{2}A^2$ и дисперсией $\sigma^2 = A^2$ этого нормального распределения выполняется соотношение $\mu = -\sigma^2/2$, – неслучайно. Легко проверить путем несложных преобразований квадратичных форм в показателе степени нормальной плотности, что

$$Y \overset{d}{=} \mathcal{N}(\mu, \sigma^2) \Longrightarrow Ee^Y = \exp(\mu + \sigma^2/2). \tag{8.1.14}$$

По определению, $e^\Lambda = L = p_{\mu_1}/p_{\mu_0}$, поэтому $E_{\mu_0}e^\Lambda = \int Lp_{\mu_0}\,d\mathbf{x} = \int p_{\mu_1}\,d\mathbf{x} = 1$, что и влечет указанное соотношение между параметрами распределения Λ.

Как уже говорилось, критическая область наиболее мощного критерия имеет вид $\{\Lambda(\mathbf{x}) > c\}$. Обозначим $S = \big(\Lambda + \frac{1}{2}A^2\big)/A$. Тогда

$$\{\Lambda(\mathbf{x}) > c\} = \{S > c_1\}, \quad c_1 = \frac{1}{A}\left(c + \frac{1}{2}A^2\right). \tag{8.1.15}$$

Из (8.1.13) следует, что $S \overset{d}{=} \mathcal{N}(0,1)$ при гипотезе H_0. Обозначим функцию стандартного нормального распределения $\mathcal{N}(0,1)$ через $\Phi(x)$ и определим z_α уравнением $\Phi(z_\alpha) = 1-\alpha$ (т.е., z_α есть $(1-\alpha)$-квантиль стандартного нормального распределения). Иначе говоря, если $S \overset{d}{=} \mathcal{N}(0,1)$, то $P(S > z_\alpha) = \alpha$. Поэтому искомая критическая область имеет вид

$$\{S > z_\alpha\}. \tag{8.1.16}$$

Из формул (8.1.11) и $A = t/\sigma$ получаем, что

$$S = \frac{1}{\sigma\sqrt{N}}\sum_{i=1}^{N}(X_i - \mu_0). \tag{8.1.17}$$

Тот факт, что S не зависит от выбора μ_1, означает, что один и тот же критерий с критической областью (8.1.16) является наиболее мощным критерием для проверки H_0 против любой простой альтернативы из H_1 и тем самым он является равномерно наиболее мощным (РНМ) критерием для проверки H_0 против сложной альтернативы H_1. РНМ критерии существуют в исключительных случаях (как правило, при наличии достаточной статистики). В общем случае (см. § 6.1) удается построить локально асимптотически наиболее мощный критерий при стремлении к бесконечности объема выборки.

Выпишем теперь мощность критерия (8.1.16). Будем обозначать ее $\beta(t)$, чтобы указать на зависимость от альтернативы $\mu_1 = \mu_0 + t/\sqrt{N}$. Таким образом (см. (8.1.15), (8.1.16))

$$\beta(t) = P_{\mu_1}(\Lambda > c) = P_{\mu_1}(S > z_\alpha).$$

Найдем распределение Λ при альтернативе μ_1. Перепишем формулу (8.1.11) в виде

$$\begin{aligned}\Lambda &= \frac{t}{\sigma^2\sqrt{N}}\sum_1^N (X_i - \mu_1) + \frac{t}{\sigma^2\sqrt{N}} N(\mu_1 - \mu_0) - \frac{t^2}{2\sigma^2} \\ &= \frac{t}{\sigma^2\sqrt{N}}\sum_1^N (X_i - \mu_1) + \frac{t^2}{2\sigma^2}.\end{aligned}$$

Пользуясь обозначением $A = t/\sigma$, получаем

$$\mathcal{L}(\Lambda \mid P_{\mu_1}) = \mathcal{N}\left(\frac{1}{2}A^2, A^2\right). \tag{8.1.18}$$

Сравнивая это выражение с (8.1.13), видим, что распределение Λ при альтернативе μ_1 отличается от распределения при μ_0 только знаком математического ожидания. Мы вывели этот факт, пользуясь конкретным выражением (8.1.11), но асимптотический аналог этого свойства выполняется для логарифма ОП распределений, обладающих свойством локальной асимптотической нормальности, см. п. 5.3.2.

Из (8.1.15) и (8.1.16) видно, что

$$c = Ac_1 - \frac{1}{2}A^2, \quad \text{где} \quad c_1 = z_\alpha.$$

Поэтому критическая область (8.1.16) в терминах Λ имеет вид

$$\Lambda > Az_\alpha - \frac{1}{2}A^2.$$

Из (8.1.18) следует, что $\left(\Lambda - \frac{1}{2}A^2\right)/A \stackrel{d}{=} \mathcal{N}(0,1)$ относительно P_{μ_1}. Следовательно, мощность НМ критерия равна

$$P_{\mu_1}\left(\Lambda > Az_\alpha - \frac{1}{2}A^2\right) = P_{\mu_1}\left(\frac{\Lambda - \frac{1}{2}A^2}{A} > z_\alpha - A\right)$$
$$= 1 - \Phi(z_\alpha - A) = \Phi(A - z_\alpha). \qquad (8.1.19)$$

Замечание 8.1.2. Обратим внимание на смысл параметра A в этой формуле. Выражение (8.1.11) для Λ состояло из двух слагаемых, первое из которых случайно (зависит от наблюдений), а второе – неслучайная константа. Фактически мы обозначили через A^2 дисперсию первого слагаемого и таким образом параметр A в выражении для мощности НМ критерия – это стандартное отклонение этого слагаемого.

В нашем случае, когда $A = t/\sigma$, получаем

$$\beta(t) = \Phi(t/\sigma - z_\alpha). \qquad (8.1.20)$$

В случае сложной альтернативы H_1 (как в данном примере) мощность зависит от конкретного распределения, входящего в H_1, и следовательно представляет собой функцию от соответствующего параметра (в данном случае от μ_1 или t). Такая функция называется *функцией мощности*.

В данном примере логарифм ОП Λ имел точное нормальное распределение, благодаря чему мы получили точные выражения для критического значения и функции мощности в терминах нормального распределения. В общем случае подобные точные выражения не имеют места, но предельные критические значения и функции мощности, даваемые асимптотической теорией, имеют вид, аналогичный полученному в рассмотренном примере (см. § 6.1).

8.1.4. Пример 2: Сравнение средних в двух нормальных выборках. Задача, рассмотренная в предыдущем пункте, была приведена в иллюстративных целях. Построенный там критерий не имеет аналогов в классе простых линейных ранговых статистик (ПЛРС), рассматриваемых в данном курсе. (Для нее используются ранговые статистики другого типа – так называемые знаково-ранговые статистики, см. [3]; читателя, заинтересованного в приложениях, отсылаем к книге [13].) Здесь мы рассмотрим задачу сравнения двух выборок, описанную во Введении, и построение в ней наилучшего «параметрического» критерия в случае нормально распределенных наблюдений. Итак, мы имеем две выборки, $X_1, \ldots, X_{n_1}$ и $X_{n_1+1}, \ldots, X_N$, $N = n_1+n_2$, элементы которых независимы и распределены нормально $\mathcal{N}(\mu_1, \sigma^2)$ и $\mathcal{N}(\mu_2, \sigma^2)$ соответственно. Проверяется гипотеза $H_0\colon \mu_1 = \mu_2$ против альтернативы $H_1\colon \mu_1 > \mu_2$. Хорошо известно, что в этой постановке применяется критерий Стьюдента, основанный на статистике[5]

$$t = \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{s^2}} \sqrt{\frac{n_1 n_2}{N}}, \tag{8.1.21}$$

где s^2 – оценка дисперсии σ^2. Чтобы упростить изложение и сосредоточиться на том аспекте вопроса, который представляет для нас первостепенный интерес, предположим, что дисперсия σ^2 известна. Тогда вместо статистики t применяется статистика, получаемая из нее заменой s^2 на σ^2, которая эквивалентна просто разности $Z = \overline{X}_1 - \overline{X}_2$. Посмотрим, какими свойствами обладает эта статистика. Прежде всего, при гипотезе H_0 она имеет распределение, не зависящее от неизвестного параметра μ (а именно, $\mathcal{N}\big(0, \sigma^2 \frac{N}{n_1 n_2}\big)$). Это важное свойство, позволяющее строить критерий с заданным уровнем значимости $\alpha > 0$, не зависящим от μ. Но критериев с таким свойством много, в частности, таковы ранговые критерии, обсуждавшиеся во Введении. Критерий же, основанный на статистике $Z = \overline{X}_1 - \overline{X}_2$, является наиболее мощным в классе критериев с уровнем значимости $\alpha(\mu) \leqslant \alpha$, в частности, $\alpha(\mu) \equiv \alpha$.

Докажем это свойство. Зафиксируем на время значение $\mu = \mu_0$ и значения $\mu_1 > \mu_2$ и рассмотрим задачу проверки *простой* гипотезы H_0: все X_i, $i = 1, \ldots, N$, имеют распределение $\mathcal{N}(\mu_0, \sigma^2)$ против *простой* альтернативы H_1: элементы первой (соответственно второй) выборки имеют распределение $\mathcal{N}(\mu_1, \sigma^2)$ (соответственно $\mathcal{N}(\mu_2, \sigma^2)$). Пользуясь леммой Неймана–Пирсона, построим наиболее мощный критерий для проверки H_0 против H_1. Аналогично

[5] Буква t традиционно используется для обозначения статистики Стьюдента. В данном пункте она используется в этом смысле только в текущем абзаце, за пределами которого она, как и в предыдущем пункте, обозначает нормированное отклонение альтернативы от гипотезы.

предыдущему пункту выпишем плотности распределения вектора наблюдений $\mathbf{X} = (X_1, \dots, X_N)$, отвечающие H_0 и H_1. При гипотезе H_0 вектор $\mathbf{X}$ имеет плотность $p(\mathbf{x})$, даваемую равенством (8.1.6), а при гипотезе H_1 – плотность

$$q(\mathbf{x}) = \varphi_{n_1,n_2}(\mathbf{x}; \mu_1, \mu_2, \sigma^2) = \prod_{i=1}^{n_1} \varphi(x_i; \mu_1, \sigma^2) \prod_{i=n_1+1}^{N} \varphi(x_i; \mu_2, \sigma^2).$$

Повторяя рассуждения предыдущего пункта, следующие за (8.1.8), получаем

$$\begin{aligned}\Lambda = \sum_{i=1}^{n_1} \bigl(\log \varphi(x_i; \mu_1, \sigma^2) - \log \varphi(x_i; \mu_0, \sigma^2)\bigr) \\ + \sum_{i=n_1+1}^{N} \bigl(\log \varphi(x_i; \mu_2, \sigma^2) - \log \varphi(x_i; \mu_0, \sigma^2)\bigr). \end{aligned} \tag{8.1.22}$$

Используем формулу (8.1.10), подставляя в ней $t = (\mu_1 - \mu_0)\sqrt{N}$ в первой сумме в (8.1.22) и $t = (\mu_2 - \mu_0)\sqrt{N}$ во второй. Обозначим

$$t(\mu_0, \mu_1, \mu_2) = \sqrt{n_1(\mu_1 - \mu_0)^2 + n_2(\mu_2 - \mu_0)^2}. \tag{8.1.23}$$

Тогда

$$\begin{aligned}\Lambda = \frac{1}{\sigma^2}\left((\mu_1 - \mu_0)\sum_{i=1}^{n_1}(X_i - \mu_0) + (\mu_2 - \mu_0)\sum_{i=n_1+1}^{N}(X_i - \mu_0)\right) \\ - \frac{t^2(\mu_0, \mu_1, \mu_2)}{2\sigma^2}. \end{aligned} \tag{8.1.24}$$

Аналогично (8.1.12) имеем

$$\mathcal{L}(\Lambda \mid P_{\mu_0}) = \mathcal{N}\left(-\frac{1}{2\sigma^2}t^2(\mu_0, \mu_1, \mu_2), \frac{1}{\sigma^2}t^2(\mu_0, \mu_1, \mu_2)\right). \tag{8.1.25}$$

Обозначим $\beta(\mu_0, \mu_1, \mu_2)$ мощность критерия уровня α, основанного на статистике Λ. С помощью рассуждений, аналогичных использованным при выводе формулы (8.1.20), получаем

$$\beta(\mu_0, \mu_1, \mu_2) = \Phi\bigl(t(\mu_0, \mu_1, \mu_2)/\sigma - z_\alpha\bigr). \tag{8.1.26}$$

Теперь, оставляя фиксированными μ_1 и μ_2, найдем «наименее благоприятное» значение $\bar{\mu}_0$, при котором эта мощность минимальна, т.е. найдем точку минимума $t^2(\mu_0, \mu_1, \mu_2)$ (см. (8.1.23) по μ_0. Дифференцируя, получаем

$$\bar{\mu}_0 = \frac{n_1\mu_1 + n_2\mu_2}{N}. \tag{8.1.27}$$

Подставим это значение в (8.1.24). Отбрасывая неслучайные слагаемые и неслучайные общие множители, получаем, что статистика Λ эквивалентна статистике

$$n_2 \sum_{i=1}^{n_1} X_i - n_1 \sum_{i=n_1+1}^{N} X_i,$$

из которой, деля на $n_1 n_2$, приходим к статистике $Z = \overline{X}_1 - \overline{X}_2$. При гипотезе H_0 (т.е. когда $X_1, \ldots, X_N$ одинаково распределены $\mathcal{N}(\mu, \sigma^2)$) эта статистика, как уже отмечалось, имеет распределение $\mathcal{N}\big(0, \sigma^2 \frac{N}{n_1 n_2}\big)$, не зависящее от μ. Поэтому для заданного $\alpha > 0$, исходя из этого распределения, можно построить критическую область $Z > c_\alpha$ так, чтобы

$$P_0(Z > c_\alpha \mid \mu) \equiv \alpha \ \text{ при всех } \ \mu \in \mathbb{R}.$$

Кроме того, по построению, этот критерий является наиболее мощным критерием уровня α для проверки простой гипотезы $H_0\colon \mu = \bar{\mu}_0$ против простой альтернативы, определяемой средними μ_1 и μ_2 для первой и второй выборок. Рассмотрим теперь произвольный критерий, имеющий уровень значимости α (или $\leqslant \alpha$) при всех $\mu \in \mathbb{R}$ (и, в частности, при $\mu = \bar{\mu}_0$), и применим его для проверки той же простой гипотезы $H_0\colon \mu = \bar{\mu}_0$ против той же простой альтернативы (μ_1, μ_2). Его мощность в этой задаче не превосходит мощности наиболее мощного критерия, основанного на статистике Z. Поскольку это свойство выполняется при любой простой альтернативе (μ_1, μ_2) с условием $\mu_1 > \mu_2$, а статистика Z не зависит от выбора этой альтернативы, критерий, основанный на статистике Z, является *равномерно наиболее мощным* (РНМ) в классе всех критериев уровня α (или $\leqslant \alpha$).

Вопрос о свойствах оптимальности критерия Стьюдента (8.1.21), применяемого в практически важном случае неизвестной дисперсии, более сложен. Он является РНМ несмещенным (т.е. он несмещенный и является РНМ в классе несмещенных критериев) уровня α, см. [7], глава 5, § 2.

Выпишем функцию мощности построенного выше РНМ критерия. Она дается формулой (8.1.26) с $\mu_0 = \bar{\mu}_0$ (8.1.27), и, таким образом, равна

$$\beta(t) = \Phi(t/\sigma - z_\alpha), \tag{8.1.28}$$

где, как показывают простые выкладки,

$$t = t_{\min}(\mu_0, \mu_1, \mu_2) = t(\bar{\mu}_0, \mu_1, \mu_2) = \sqrt{\frac{n_1 n_2}{N}}(\mu_1 - \mu_2). \tag{8.1.29}$$

Напомним, что μ_1 и μ_2 это значения параметра μ для наблюдений 1-й и 2-й выборок, и формулы (8.1.28) и (8.1.29) показывают, что мощность критерия, основанного на $Z = \overline{X}_1 - \overline{X}_2$, зависит только от их разности.

Видно, что выражение (8.1.28) совпадает с (8.1.20). Это происходит благодаря общему принципу выбора нормированного параметра t. А именно, пусть $X_i \overset{d}{=} \mathcal{N}(\mu_i, \sigma^2)$, $i = 1, \ldots, N$. Тогда как в предыдущем, так и в настоящем пунктах $\beta(t)$ есть функция мощности НМ критерия в задаче проверки гипотезы $H_0\colon \mu_i = \mu_0$, $i = 1, \ldots, N$, против альтернативы $H_1\colon \mu_i = \mu_0 + tc_i$, где $\sum_{i=1}^N c_i^2 = 1$. Действительно, в п. 8.1.3 альтернатива состояла в том, что $\mu_i = \mu_1$ при всех i, и мы вводили параметр $t = (\mu_1 - \mu_0)\sqrt{N}$. Тогда альтернатива H_1 имеет вид $\mu_i = \mu_0 + t/\sqrt{N}$, $i = 1, \ldots, N$, и $c_1 = \cdots = c_N = 1/\sqrt{N}$ очевидно удовлетворяют условию $\sum c_i^2 = 1$. В настоящем пункте в роли μ_0 выступает $\bar{\mu}_0$ и из (8.1.27) находим, что

$$\mu_1 - \bar{\mu}_0 = \frac{n_2}{N}(\mu_1 - \mu_2), \quad \mu_2 - \bar{\mu}_0 = -\frac{n_1}{N}(\mu_1 - \mu_2),$$

откуда, сравнивая с (8.1.29), видим, что $\mu_i - \bar{\mu}_0 = tc_i$, где

$$c_i = \begin{cases} \sqrt{n_2/Nn_1}, & i = 1, \ldots, n_1, \\ -\sqrt{n_1/Nn_2}, & i = n_1 + 1, \ldots, N. \end{cases} \tag{8.1.30}$$

Снова очевидно, что $\sum c_i^2 = 1$. Такая нормировка постоянно используется в главах 5 и 6, что позволяет сравнивать между собой получаемые в них формулы для предельной мощности, а в частном случае нормально распределенных наблюдений – сравнивать их с полученными выше выражениями.

Отметим, что мощность критерия Стьюдента, основанного на статистике (8.1.21), сходится при $N \to \infty$ к $\beta(t)$ (8.1.28). Действительно, при фиксированном σ^2 статистика

$$Z' = \frac{\overline{X}_1 - \overline{X}_2}{\sigma} \sqrt{\frac{n_1 n_2}{N}},$$

получаемая из (8.1.21) заменой s^2 на σ^2, отличается от статистики $Z = \overline{X}_1 - \overline{X}_2$ неслучайным множителем, поэтому в рассмотренной выше задаче критерий, основанный на статистике Z', имеет ту же мощность $\beta(t)$, что и критерий, основанный на статистике Z. Статистика Стьюдента (8.1.21) отличается от Z' множителем $\sqrt{\sigma^2/s^2}$, который при гипотезе H_0 сходится по вероятности к 1. Из результатов гл. 5 следует, что эта сходимость имеет место и при любом фиксированном t (8.1.29). Поэтому распределение статистики Стьюдента сходится к распределению статистики Z' как при гипотезе H_0 (когда $Z' \stackrel{d}{=} \mathcal{N}(0,1)$), так и при любой последовательности альтернатив, при которой μ_1 и μ_2 сближаются так, что их нормированная разность (8.1.29) остается постоянной. Как следствие, мощность критерия Стьюдента сходится к $\beta(t)$.

Для статистики Z в п. 6.2.4 строится ранговый аналог, дающий асимптотически наиболее мощный критерий в рассмотренной выше задаче.

8.1.5. Пример 3: Проверка гипотезы об отсутствии тренда в нормальной регрессии. Рассмотрим еще одну задачу проверки гипотез для нормально распределенных наблюдений, в которой, как и в предыдущем пункте, строится равномерно наиболее мощный критерий, также допускающий асимптотически наиболее мощный ранговый аналог.

А именно, рассмотрим модель линейной регрессии, в которой наблюдения Y_i, $i = 1, \ldots, N$, независимы и распределены нормально $\mathcal{N}(\mu_i, \sigma^2)$ с математическими ожиданиями $\mu_i = \theta + \eta x_i$, где θ и η – неизвестные параметры, а $x_1, \ldots, x_N$ – заданные константы, не все равные между собой[6]. Дисперсия $\sigma^2 > 0$ снова предполагается известной. Иначе говоря, обозначая «ошиб-

[6] В данном пункте мы пишем Y_i вместо принятого во всем курсе обозначения наблюдений X_i в соответствии с общепринятым в теории линейной регрессии обозначением для «зависимой переменной» («отклика»). Однако мы избегаем употребляемого обычно обозначения α, β для коэффициентов линии регрессии, поскольку эти буквы используются для обозначения уровня значимости и мощности критериев.

ки» (отклонения Y-ов от линии регрессии) через $e_i = Y_i - \mu_i$, $i = 1, \ldots, N$, мы можем записать наши наблюдения в виде

$$Y_i = \theta + \eta x_i + e_i, \quad e_i \stackrel{d}{=} \mathcal{N}(0, \sigma^2), \quad i = 1, \ldots, N. \tag{8.1.31}$$

В этой схеме рассмотрим задачу проверки гипотезы $H_0\colon \eta = 0$ против альтернативы $H_1\colon \eta > 0$. Заметим, что гипотеза H_0 здесь точно та же, что и в предыдущем пункте (с точностью до обозначений), и что в случае $x_1 = \cdots = x_{n_1}$, $x_{n_1+1} = \cdots = x_N$, $1 < n_1 < N$, $x_1 \neq x_N$, данная задача переходит в задачу сравнения двух выборок предыдущего пункта.

Мы будем строить наиболее мощный критерий по той же схеме, что и в предыдущем пункте. Зафиксируем некоторые θ_0, θ_1 и η_1 и рассмотрим задачу проверки простой гипотезы $H_0'\colon \mu_i = \theta_0$, $i = 1, \ldots, N$, против простой альтернативы $H_1'\colon \mu_i = \theta_1 + \eta_1 x_i$, $i = 1, \ldots, N$.

При гипотезе H_0' вектор наблюдений $\mathbf{Y} = (Y_1, \ldots, Y_N)$ имеет (аналогично (8.1.6)) плотность распределения

$$p(\mathbf{y}) = \prod_{i=1}^{N} \varphi(y_i; \theta_0, \sigma^2), \quad \mathbf{y} = (y_1, \ldots, y_N), \tag{8.1.32}$$

а при H_1' – плотность распределения

$$q(\mathbf{y}) = \prod_{i=1}^{N} \varphi(y_i; \theta_1 + \eta_1 x_i, \sigma^2). \tag{8.1.33}$$

Соответствующие распределения вектора $\mathbf{Y}$ будем обозначать P и Q. Как и раньше, полагаем $\Lambda = \Lambda(\mathbf{Y}) = \log[q(\mathbf{Y})/p(\mathbf{Y})]$. По лемме Неймана–Пирсона НМ критерий уровня α отвергает гипотезу H_0' при $\Lambda > c_\alpha$, где c_α выбрано так, чтобы $P(\Lambda > c_\alpha) = \alpha$. Чтобы получить выражение для Λ, выпишем сначала выражение для логарифма отношения i-х сомножителей в (8.1.32) и (8.1.33):

$$\log \frac{\varphi(y_i; \theta_1 + \eta_1 x_i, \sigma^2)}{\varphi(y_i; \theta_0, \sigma^2)} = y_i((\theta_1 + \eta_1 x_i) - \theta_0) + c(\theta_0, \theta_1, \eta_1, \sigma^2), \tag{8.1.34}$$

где $c(\theta_0, \theta_1, \eta_1, \sigma^2)$ – некоторое выражение, зависящее от указанных параметров, но не от y_i.

Суммируя эти выражения по i, получаем, что

$$\Lambda(\mathbf{Y}) = Z + C, \quad Z = \sum_{i=1}^{N} \mathbf{Y}_i((\theta_1 + \eta_1 x_i) - \theta_0), \tag{8.1.35}$$

где C – неслучайное число, зависящее от фиксированных (заданных) параметров задачи. Поэтому критическую область $\{\Lambda > c_\alpha\}$ можно представить как $\{Z > c'_\alpha\}$, где $P(Z > c'_\alpha) = \alpha$.

Представление (8.1.35) отличается от (8.1.11) тем, что $E(Z \mid H'_0) \neq 0$, но будет полным аналогом (8.1.11), если в формуле для Z написать $\mathbf{Y}_i - \theta_0$ вместо $\mathbf{Y}_i$ с соответствующим изменением константы C. Однако отличие на неслучайную постоянную не меняет дисперсии Z, равной при гипотезе H'_0

$$\sigma_Z^2 = \sigma^2 \sum_{i=1}^{N} ((\theta_1 + \eta_1 x_i) - \theta_0)^2, \tag{8.1.36}$$

и для получения мощности НМ критерия мы воспользуемся замечанием 8.1.2, согласно которому эта мощность равна

$$\beta_Z = \Phi(\sigma_Z - z_\alpha).$$

Теперь нам нужно найти θ_0^* – точку минимума σ_Z^2. Как известно, для произвольных $a_1, \dots, a_N$, $\min_\theta \sum_{i=1}^N (a_i - \theta)^2$ достигается при $\theta = \bar{a} = \sum_{i=1}^N a_i/N$. Поэтому (см. (8.1.36))

$$\theta_0^* = \theta_1 + \eta_1 \bar{x}, \quad \bar{x} = \frac{1}{N} \sum_{i=1}^{N} x_i, \tag{8.1.37}$$

и

$$\bar{\sigma}_Z^2 = \min_{\theta_0} \sigma_Z^2 = \eta_1^2 (\sigma^*)^2, \quad \text{где} \quad (\sigma^*)^2 = \sigma^2 \sum_{i=1}^{N} (x_i - \bar{x})^2. \tag{8.1.38}$$

Положим $Z^* = Z(\theta_0^*)/\eta_1$, где $Z(\theta_0^*)$ – статистика, получаемая при подстановке $\theta_0 = \theta_0^*$ в выражение для Z, см. (8.1.35). Тогда

$$Z^* = \sum_{i=1}^{N} \mathbf{Y}_i (x_i - \bar{x}). \tag{8.1.39}$$

По построению, критерий вида $Z^* > c^*$ является НМ критерием для проверки простой гипотезы $H^*\colon \mathbf{Y}_i \stackrel{d}{=} \mathcal{N}(\theta_0^*, \sigma^2)$, $i = 1, \ldots, N$, против простой альтернативы $H_1'\colon \mathbf{Y}_i \stackrel{d}{=} \mathcal{N}(\theta_1 + \eta_1 x_i, \sigma^2)$, $i = 1, \ldots, N$. Как всегда, критическое значение c^* выбирается так, чтобы $P(Z^* > c^*) = \alpha$ при гипотезе H^*. Легко видеть, однако, что если $\mathbf{Y}_i \stackrel{d}{=} \mathcal{N}(\theta, \sigma^2)$, $i = 1, \ldots, N$, то $Z^* \stackrel{d}{=} \mathcal{N}(0, (\sigma^*)^2)$ для любого $\theta \in \mathbb{R}$, см. (8.1.38) (т.е. значение θ_0^*, отвечающее гипотезе H^*, не играет никакой особой роли при определении c^*). Это значит, что с помощью статистики Z^* можно проверять *сложную* гипотезу H_0 (см. (8.1.31) и далее), при которой все $\mathbf{Y}_i$, $i = 1, \ldots, N$, имеют одно и то же распределение $\mathcal{N}(\theta, \sigma^2)$ с произвольным $\theta \in \mathbb{R}$. А именно, критерий, отвергающий гипотезу H_0 при $Z^* > \sigma^* z_\alpha$, имеет уровень значимости α при любом $\theta \in \mathbb{R}$.

Покажем, что этот критерий является равномерно наиболее мощным (РНМ) критерием проверки H_0 против альтернативы

$$H_1\colon \mathbf{Y}_i \stackrel{d}{=} \mathcal{N}(\theta + \eta x_i, \sigma^2), \quad \theta \in \mathbb{R}, \quad \eta > 0, \quad i = 1, \ldots, N$$

(это лишь иная формулировка альтернативы H_1, введенной после (8.1.31)). В терминах параметров (θ, η) H_1 представляется полуплоскостью $\{(\theta, \eta)\colon \theta \in \mathbb{R},\ \eta > 0\}$, а H_0 – прямой $\{(\theta, \eta)\colon \theta \in \mathbb{R},\ \eta = 0\}$. Рассмотрим какой-нибудь другой критерий, имеющий уровень значимости $\alpha(\theta) \leqslant \alpha$ (в частности, $\alpha(\theta) \equiv \alpha$) и сравним его мощность (зависящую от (θ, η)) с мощностью критерия, основанного на Z^*. Возьмем произвольные $\theta_1 \in \mathbb{R}$, $\eta_1 > 0$, определим θ^* по формуле (8.1.37) и рассмотрим задачу проверки простой гипотезы H^* против простой альтернативы H_1', введенных выше. Наш критерий, отвергающий гипотезу H^* при $Z^* > \sigma^* z_\alpha$, является НМ критерием в этой задаче. Это значит, что его мощность не меньше мощности любого другого критерия уровня α (или $\leqslant \alpha$) при гипотезе H^*. Взятый нами конкурирующий критерий удовлетворяет этому условию и следовательно, его мощность при альтернативе H_1' не превосходит мощности критерия, основанного на Z^*. Но альтернатива H_1' представляет собой произвольно выбранную точку (θ_1, η_1) из множества точек (θ, η), образующих сложную альтернативу H_1, т.е. мощность нашего критерия не меньше мощности конкурируюшего критерия в любой точке (θ, η) из H_1. Это означает, что критерий $Z^* > \sigma^* z_\alpha$ является *равномерно наиболее мощным* критерием уровня α в задаче проверки сложной гипотезы H_0 против сложной альтернативы H_1.

Предлагаем читателю самостоятельно выписать функцию мощности построенного критерия и убедиться, что при выборе параметра t, описанном в предыдущем пункте (см. абзац после (8.1.29)), она дается тем же выражением (8.1.20) (или (8.1.28)).

Если дисперсия σ^2 неизвестна, то делением статистики Z^* на корень из оценки для σ^2 строится статистика, имеющая распределение Стьюдента с $N-2$ степенями свободы. (Эта конструкция имеется во всех руководствах по математической статистике, где рассматриваются задачи регрессионного анализа.) Свойства оптимальности соответствующего критерия можно найти в книге Лемана [7], глава 5, § 6.

§ 8.2. Сходимость распределений и интегралов

Здесь мы сформулируем некоторые вспомогательные результаты, которые используются при доказательстве предельных теорем. Их можно найти в книгах А. Н. Ширяева [15] и М. Лоэва [8]. Мы приводим их здесь для удобства ссылок.

8.2.1. Сходимость распределений на прямой. Мы обозначаем через $\mathbb{R}$ числовую прямую и через $\mathcal{B}$ сигма-алгебру борелевских подмножеств $\mathbb{R}$ (в дальнейшем та же буква $\mathcal{B}$ будет обозначать борелевскую σ-алгебру пространства, рассматриваемого в данном контексте; для уточнения будем писать, например, $\mathcal{B}(\mathbb{R})$ или $\mathcal{B}(\mathbb{R}^N)$ и т.п.). Пусть P, $P_1, P_2, \ldots$ – распределения вероятностей (вероятностные меры) на $(\mathbb{R}, \mathcal{B})$ и $F(x)$, $F_1(x)$, $F_2(x), \ldots$ – соответствующие им функции распределения.

Определение 8.2.1. Последовательность распределений P_N (ф.р. F_N), $N = 1, 2, \ldots$, слабо сходится к распределению P (к ф.р. F) при $N \to \infty$ (обозначается $P_N \xrightarrow{w} P$, соотв. $F_N \xrightarrow{w} F$), если для любой непрерывной ограниченной функции f на $\mathbb{R}$

$$\int f(x)\, dF_N(x) \to \int f(x)\, dF(x).$$

(Мы не указываем область интегрирования, если интегрирование производится по всему пространству, рассматриваемому в данном контексте, в данном случае – по всей прямой.)

Теорема 8.2.1. *$F_N \xrightarrow{w} F$ тогда и только тогда, когда $F_N(x) \to F(x)$ в каждой точке непрерывности $F(x)$.*

8.2.2. Сходимость распределений на плоскости. Это единственный пункт, где рассматриваются вероятностные меры на конечномерном пространстве $\mathbb{R}^k$, конкретнее, мы ограничиваемся случаем $k = 2$, достаточным для целей данного курса. Приводимые здесь утверждения с полными доказательствами можно найти в гл. 1 книги Биллингсли [2] (хотя в этой книге рассматриваются вероятностные меры и на более общих пространствах, нужные нам результаты, относящиеся к конечномерным евклидовым пространствам $\mathbb{R}^k$ и, в частности, $\mathbb{R}^2$, как правило, приводятся в качестве примеров или частных случаев).

Пусть P, $P_1, P_2, \ldots$ – вероятностные меры на плоскости $(\mathbb{R}^2, \mathcal{B}(\mathbb{R}^2))$. Как и выше, имеется несколько эквивалентных определений слабой сходимости $P_N \to P$. Мы принимаем за определение аналог свойства, фигурирующего в теореме 8.2.1.

Определение 8.2.2. Последовательность распределений P_N, $N = 1, 2, \ldots$, на плоскости слабо сходится к распределению P, если для любого борелевского множества $A \in \mathbb{R}^2$ такого, что $P(\partial A) = 0$, где ∂A – граница множества A,

$$P_N(A) \to P(A). \tag{8.2.1}$$

Обозначим $A(x, y) = \{(x_1, x_2)\colon x_1 < x, x_2 < y\}$ (левый нижний квадрант по отношению к точке (x, y)).

Теорема 8.2.2. *Последовательность распределений P_N, $N = 1, 2, \ldots$, на плоскости слабо сходится к распределению P тогда и только тогда, когда для любой точки (x, y) такой, что $P(\partial A(x, y)) = 0$,*

$$P_N(A(x, y)) \to P(A(x, y)). \tag{8.2.2}$$

Понятно, что $F_N(x, y) := P_N(A(x, y))$ и $F(x, y) := P(A(x, y))$ представляют собой функции распределения случайных векторов, имеющих распределения P_N и P. Поэтому теорема 8.2.2 является двумерным аналогом теоремы 8.2.1, утверждающим, что для слабой сходимости $P_N \xrightarrow{w} P$ необходимо и достаточно, чтобы соответствующие функции распределения сходились в каждой точке непрерывности $F(x, y)$.

В одну сторону теорема 8.2.2 очевидным образом следует из определения, поскольку всякое множество $A(x, y)$ – борелевское. Нетривиальной частью утверждения теоремы является то, что из сходимости (8.2.2) на бедном классе множеств (квадрантов) следует сходимость (8.2.1) на классе всех борелевских множеств. Доказательство теоремы см. в [2], с. 30–31.

Прием Крамера–Уолда. Этот хорошо известный теперь прием сводит доказательство сходимости последовательности распределений случайных (конечномерных) векторов к доказательству сходимости одномерных распределений. Этот прием используется в п.п. 6.1.1 и 6.1.3 для доказательства сходимости совместных распределений двух сумм независимых случайных величин к двумерному нормальному распределению. Поэтому мы сформулируем его для двумерного случая, хотя формулировка очевидным образом переносится на случай произвольной размерности.

Итак, мы рассматриваем последовательность двумерных случайных векторов $S_N = (S_{N1}, S_{N2})$, $N = 1, 2, \ldots$, и нас интересует их сходимость по распределению к случайному вектору $S = (S_1, S_2)$,

$$S_N \xrightarrow{d} S, \quad N \to \infty. \tag{8.2.3}$$

Иначе говоря, нас интересует сходимость их распределений P_1, $P_2, \ldots$ на $(\mathbb{R}^2, \mathcal{B}(\mathbb{R}^2))$ к распределению P вектора S,

$$P_N \xrightarrow{w} P, \quad N \to \infty. \tag{8.2.4}$$

Рассмотрим линейные комбинации $t \cdot S_N := t_1 S_{N1} + t_2 S_{N2}$ для всевозможных $t = (t_1, t_2) \in \mathbb{R}^2$. Если $S_N \xrightarrow{d} S$, то $t \cdot S_N \xrightarrow{d} t \cdot S$ для любого $t \in \mathbb{R}^2$ (почему?). Оказывается, что и обратно, из сходимости распределений указанных линейных комбинаций (одномерных случайных величин) следует сходимость распределений двумерных случайных векторов S_N к распределению вектора S. Именно, имеет место теорема (см. [2], с. 76).

Теорема 8.2.3. *Для сходимости* (8.2.3) (*или, эквивалентно*, (8.2.4)) *необходимо и достаточно, чтобы*

$$t \cdot S_N \xrightarrow{d} t \cdot S \quad \textit{для любого} \quad t \in \mathbb{R}^2.$$

Эта теорема является непосредственным следствием теоремы непрерывности для характеристических функций (см. [2], с. 72).

8.2.3. Относительная компактность и плотность. Пусть снова $\mathcal{P} = \{P_\theta, \theta \in \Theta\}$ – некоторое семейство вероятностных мер на $(\mathbb{R}, \mathcal{B})$.

Определение 8.2.3. Семейство $\mathcal{P}$ *относительно компактно*, если любая последовательность мер из $\mathcal{P}$ содержит подпоследовательность, слабо сходящуюся к некоторой вероятностной мере.

Определение 8.2.4. Семейство $\mathcal{P}$ называется *плотным*[7], если для любого $\varepsilon > 0$ найдется компакт $K \in \mathbb{R}$ такой, что

$$\sup_\theta P_\theta(\mathbb{R} \setminus K) < \varepsilon. \tag{8.2.5}$$

Теорема 8.2.4 (теорема Прохорова). *Семейство $\mathcal{P}$ относительно компактно тогда и только тогда, когда оно плотно.*

Замечание 8.2.1. Эта теорема была доказана Ю. В. Прохоровым (1956) [10] для вероятностных мер на произвольном полном метрическом пространстве. Легко видеть, что определения 8.2.1–8.2.4 переносятся на этот общий случай почти буквально, с заменой лишь пространства $(\mathbb{R}, \mathcal{B})$ на соответствующее метрическое пространство с борелевской сигма-алгеброй его подмножеств. В рассматриваемом случае распределений на прямой в определении можно вместо компакта K говорить о замкнутом интервале вида $[-A, A]$, $A > 0$. Подробное изложение общей теории сходимости вероятностных мер можно найти в книге Биллингсли [2].

8.2.4. Математические ожидания и теорема о предельном переходе. Предположим при каждом $N = 1, 2, \dots$ задано измеримое пространство $(\mathbb{X}_N, \mathcal{A}_N)$, т.е. множество $\mathbb{X}_N$ с сигма-алгеброй его подмножеств $\mathcal{A}_N$, на которой определена вероятностная мера P_N. Точки пространства $\mathbb{X}_N$ будут обозначаться через $\mathbf{x}$. Рассматривая $P_N(A)$, $A \in \mathcal{A}_N$, как вероятность попадания точки $\mathbf{x}$ в множество A, мы можем считать $\mathbf{x}$ случайной точкой с распределением P_N. В дальнейшем мы будем часто иметь дело с математическими ожиданиями функций, определенных на $\mathbb{X}_N$. Пусть при каждом N задана функция $T_N(\mathbf{x})$, принимающая действительные значения. Ее математическое ожидание

[7] Англ. термин: tight.

будем обозначать $E_N(T_N)$. Отметим важное свойство математических ожиданий. По определению,

$$E_N(T_N) = \int_{\mathbb{X}_N} T_N(\mathbf{x})\, P_N(d\mathbf{x}).$$

С другой стороны, если мы введем функцию распределения случайной величины T_N,

$$F_N(t) = P_N\{\mathbf{x}\colon T_N(\mathbf{x}) < t\},$$

то

$$E_N(T_N) = \int_{\mathbb{R}} t\, dF_N(t).$$

Точно так же, если требуется найти математическое ожидание функции от T_N, скажем, $g(T_N)$, то

$$E_N(g(T_N)) = \int_{\mathbb{X}_N} g(T_N(\mathbf{x}))\, P_N(d\mathbf{x}) = \int_{\mathbb{R}} g(t)\, dF_N(t). \tag{8.2.6}$$

Приведем теорему, отвечающую на вопрос, будет ли

$$\int g(t)\, dF_N(t) \to \int g(t)\, dF(t), \text{ если } F_N(t) \to F(t).$$

Относительно сходимости функций распределения мы примем еще одно определение. Дело в том, что последовательность ф.р. может сходиться к «несобственной» ф.р., т.е. к монотонно неубывающей функции, полное изменение которой меньше 1. Крайним примером такого рода служит последовательность $F_N(t) = G(t - c_N)$, где $G(t)$ – произвольная ф.р., а $c_N \to \infty$ при $N \to \infty$. В этом случае $F_N(t) \to F(t) \equiv 0$. В определении 8.2.1 и теореме 8.2.1 мы заранее предполагали, что предельная ф.р. «собственная». Не делая такого предположения, будем говорить, что последовательность ф.р. $F_N(t)$ *сходится вполне* к функции $F(t)$, если $F_N(t) \to F(t)$ в каждой точке непрерывности $F(t)$. (Этот термин принят в книге Ширяева [15], в книге Лоэва [8] так определяется слабая сходимость.) Заметим, что $0 \leqslant F(t) \leqslant 1$, $t \in \mathbb{R}$, поскольку для любой точки t непрерывности функции F значение $F(t)$ является пределом числовой последовательности $F_N(t)$, заключенной в тех же пределах.

Определение 8.2.5. Пусть $F_N(t)$, $N = 1, 2, \ldots$, – последовательность ф.р. и $g(t)$, $t \in \mathbb{R}$, измеримая функция. Говорят, что g равномерно интегрируема по F_N, если для любого $\varepsilon > 0$ найдется $A > 0$ такое, что

$$\int_{|t|>A} |g(t)|\, dF_N(t) < \varepsilon.$$

Теперь мы можем сформулировать нужную нам теорему о сходимости (см. Лоэв [8], с. 196).

Теорема 8.2.5. *Предположим, что последовательность ф.р.* $F_N(t)$, $N = 1, 2, \ldots$, *сходится вполне к* $F(t)$ *и пусть* $g(t)$, $t \in \mathbb{R}$, – *измеримая функция. Тогда*:

1. $\liminf \int |g(t)|\, dF_N(t) \geqslant \int |g(t)|\, dF(t)$;
2. *если* g *равномерно интегрируема по* F_N, *то*

$$\int g(t)\, dF_N(t) \to \int g(t)\, dF(t);$$

3. $\int |g(t)|\, dF_N(t) \to \int |g(t)|\, dF(t) < \infty$ *тогда и только тогда, когда* $g(t)$ *равномерно интегрируема по* F_N.

Доказательство теоремы читатель найдет в цитированной книге Лоэва. Здесь мы приведем пример, иллюстрирующий утверждения теоремы и, в частности, показывающий роль условия неотрицательности подынтегральной функции в п. 3.

Пример 8.2.1. Рассмотрим отрезок $[0, 1]$ с мерой Лебега на нем и пусть $f(x) \equiv 1$, $x \in [0, 1]$. Обозначим меру Лебега через λ, и пусть $f_N(x)$ – последовательность функций на $[0, 1]$ такая, что

(1) $\int_0^1 f_N(x)\, d\lambda = \int_0^1 f(x)\, d\lambda = 1$ при всех N и

(2) $f_N \to f$ по мере, т.е. $\lambda\{x\colon |f_N(x) - f(x)| > \varepsilon\} \to 0\ \forall \varepsilon > 0$.

Можно ли утверждать, что тогда $\int_0^1 |f_N(x) - f(x)|\, d\lambda \to 0$? Конечно, нет. Например, возьмем последовательности $a_N \to \infty$ и $\delta_N \to 0$ и положим

$$f_N(x) = \begin{cases} 1 - a_N, & 0 \leqslant x \leqslant \delta_N, \\ 1 + a_N, & \delta_N < x \leqslant 2\delta_N, \\ 1 & 2\delta_N < x \leqslant 1. \end{cases}$$

Тогда условия (1) и (2) выполнены, но $\int_0^1 |f_N(x) - f(x)|\, d\lambda = 2a_N\delta_N$. Видно, что a_N и δ_N можно подобрать так, чтобы этот интеграл стремился к любому заданному числу или к $+\infty$. Заметим, что $f_N(x)$ отрицательна при $0 \leqslant x \leqslant \delta_N$ для достаточно больших N. Но если потребовать, чтобы $f_N(x) \geqslant 0$, такой контрпример не проходит. И действительно, если к (1) и (2) добавить условие неотрицательности $f_N(x)$, то в самом деле $\int_0^1 |f_N(x) - f(x)|\, d\lambda \to 0$. (Предоставляем читателю самостоятельно доказать это утверждение.) Легко видеть, что в условии (1) вместо равенства можно потребовать сходимость интегралов в левой части к интегралу от $f(x)$, что является аналогом условия сходимости интегралов в п. 3 теоремы.

Теперь переформулируем этот пример в терминах функций распределения. Будем рассматривать отрезок $[0, 1]$ с мерой Лебега на нем как вероятностное пространство и функции $f(x)$, $f_N(x)$ как случайные величины на этом пространстве. Обозначим через $F(t)$ и $F_N(t)$, $t \in \mathbb{R}$, соответствующие функции распределения. Тогда $F(t)$ – функция единичного скачка в точке $t = 1$, а $F_N(t)$ – ступенчатая функция со скачками δ_N в точках $\pm a_N$ и скачком $1 - 2\delta_N$ в точке $t = 1$. В силу свойств математических ожиданий, упомянутых выше,

$$Ef = \int_0^1 f(x)\, d\lambda = \int_{\mathbb{R}} t\, dF(t) = 1,$$
$$Ef_N = \int_0^1 f_N(x)\, d\lambda = \int_{\mathbb{R}} t\, dF_N(t) = 1.$$

Предоставляем читателю самостоятельно проанализировать на этом примере утверждения теоремы 8.2.5, при этом полезно рассмотреть также видоизменения функций $f_N(x)$, полагая, например, δ_N фиксированным (при этом $F_N(t)$ будут сходиться к несобственной ф.р.) или определяя $f_N(x)$ на интервале $[0, \delta_N]$ как и выше, но полагая $f_N(x) = 1$ на $(\delta_N, 1]$ (в нашем примере нет равномерной интегрируемости, но интегралы не только сходятся, но даже равны интегралу по предельной ф.р. в силу симметрии $F_N(t)$ относительно точки $t = 1$; указанное изменение приведет к отсутствию сходимости интегралов).

8.2.5. Лемма Шеффе и теорема Витали. В данном пункте мы доказываем два утверждения, которые находят широкое применение в асимптотическом анализе. Первое из них, лемма Шеффе, используется при доказательстве второго утверждения, теоремы Витали. Последняя применяется в настоящем курсе в п. 7.2.2.

Мы рассматриваем произвольное вероятностное пространство $(\mathcal{X}, \mathcal{A}, P)$, точки $x \in \mathcal{X}$ которого играют роль элементарных событий, а измеримые функции $f = f(x)$ – случайных величин. (В настоящем курсе в качестве $(\mathcal{X}, \mathcal{A})$ рассматривается обычно числовая прямая с сигма-алгеброй борелевских множеств.) Как обычно, буква E обозначает математическое ожидание, $Ef = \int f(x)\, P(dx)$. Кроме того, для множества $A \in \mathcal{A}$ обозначаем $E[f; A] = \int_A f(x)\, P(dx)$.

Лемма 8.2.1 (лемма Шеффе). *Пусть на* $(\mathcal{X}, \mathcal{A}, P)$ *заданы неотрицательная функция* $f(x) \geqslant 0$ *и последовательность неотрицательных функций* $f_n(x) \geqslant 0$, $n = 1, 2, \ldots$, *такие, что*

(a) $\limsup_{n\to\infty} Ef_n \leqslant Ef < \infty$,

(b) $\forall \varepsilon > 0$, $P\{x\colon f_n(x) < f(x) - \varepsilon\} \to 0$ *при* $n \to \infty$.

Тогда

$$E|f_n - f| \to 0 \text{ \textit{при} } n \to \infty.$$

Замечание 8.2.2. В исходной формулировке леммы Шеффе предполагается сходимость $f_n \xrightarrow{p} f$. Условие (b) представляет собой ослабленный («односторонний») вариант этого предположения.

Доказательство. Будем использовать обозначение: для $a \in \mathbb{R}$,

$$a^+ = \begin{cases} a, & \text{если } a \geqslant 0, \\ 0, & \text{если } a < 0, \end{cases} \qquad a^- = \begin{cases} 0, & \text{если } a \geqslant 0, \\ -a, & \text{если } a < 0. \end{cases}$$

Заметим, что $a^+ - a^- = a$ и $a^+ + a^- = |a|$. Таким образом, нам надо доказать, что

$$E(f_n - f)^+ \to 0 \quad \text{и} \quad E(f_n - f)^- \to 0.$$

Обозначим A_ε событие, фигурирующее в условии (b),

$$A_\varepsilon = \{x\colon f_n(x) < f(x) - \varepsilon\}.$$

В силу неотрицательности f_n и f имеем $(f_n - f)^- < f$ на множестве A_ε, а на множестве A_ε^c имеем $(f_n - f)^- \leqslant \varepsilon$. Поэтому

$$E(f_n - f)^- = E[(f_n - f)^-; A_\varepsilon] + E[(f_n - f)^-; A_\varepsilon^c] \leqslant E[f; A_\varepsilon] + \varepsilon.$$

Поскольку $P(A_\varepsilon) \to 0$ по условию (b), получаем, что $E[f; A_\varepsilon] \to 0$, следовательно,
$$E(f_n - f)^- < 2\varepsilon$$
при всех достаточно больших n.

По условию (a) $E(f_n - f) < \varepsilon$ при всех достаточно больших n. Поскольку $(f_n - f)^+ - (f_n - f)^- = f_n - f$, получаем
$$E(f_n - f)^+ - E(f_n - f)^- = E(f_n - f) < \varepsilon,$$
и следовательно,
$$E(f_n - f)^+ < 3\varepsilon$$
при всех достаточно больших n. Из полученных оценок следует утверждение леммы. $\square$

Прежде чем переходить к следующей теореме, приведем «теорему о L_r-сходимости» из книги Лоэва [8], с. 174. Нам потребуется следующее определение (см. там же): для последовательности случайных величин X_n интегралы от X_n по мере P называются равномерно непрерывными, если для любого $\varepsilon > 0$ найдется $\delta > 0$, на зависящее от n, такое, что $\int_A |X_n|\, dP < \varepsilon$ для любого множества A с $P(A) < \delta$.

Теорема 8.2.6. *Пусть имеется последовательность случайных величин $X_n \in L_r$, т.е. $E|X_n|^r < \infty$, и случайная величина X.*

(I) $E|X_n - X|^r \to 0$ *тогда и только тогда, когда*

(II) $X_n \xrightarrow{p} X$ *и* $\int |X_n - X|^r\, dP$ *равномерно непрерывны или*

(III) $X_n \xrightarrow{p} X$ *и* $\int |X_n|^r\, dP$ *равномерно непрерывны.*

Теорема 8.2.7 (теорема Витали). *Пусть на $(\mathcal{X}, \mathcal{A}, P)$ заданы функции $f, f_1, f_2, \ldots$ такие, что для некоторого $r > 0$*

(A) $\limsup_{n\to\infty} E|f_n|^r \leqslant E|f|^r$,

(B) *для любого* $\varepsilon > 0$
$$P\{f(x) \geqslant 0, f_n(x) < f(x) - \varepsilon\} + P\{f(x) < 0, f_n(x) > f(x) - \varepsilon\} \to 0.$$
Тогда
$$E|f_n - f|^r \to 0 \ \text{при}\ n \to \infty.$$

Замечание 8.2.3. Лемма Шеффе следует из теоремы Витали при $r = 1$, но она доказывалась в качестве самостоятельного утверждения, поскольку она используется при доказательстве теоремы Витали.

Доказательство. Положим $g = |f|^r$, $g_n = |f_n|^r$. По условию (В) $P(g_n < g - \varepsilon) \to 0$ для любого ε. Следовательно, по лемме Шеффе $E|g_n - g| \to 0$, а тогда по теореме 8.2.6 $g_n \xrightarrow{p} g$ и интегралы от g_n равномерно непрерывны. По той же теореме из $f_n \xrightarrow{p} f$ и равномерной непрерывности интегралов от $g_n = |f_n|^r$ следует доказываемое утверждение $E|f_n - f|^r \to 0$. Таким образом, нам надо показать, что $f_n \xrightarrow{p} f$, т.е. $P(|f_n - f| > \varepsilon) \to 0$ для произвольного $\varepsilon > 0$.

Из полученного выше соотношения $E\big|\,|f_n|^r - |f|^r\,\big| \to 0$ следует $P\big(\big|\,|f_n|^r - |f|^r\,\big| > \varepsilon\big) \to 0$ для любого $\varepsilon > 0$, т.е. $|f_n|^r \xrightarrow{p} |f|^r$, а из последнего соотношения следует $|f_n| \xrightarrow{p} |f|$.

Представим событие $\{|f_n - f| > \varepsilon\}$ как

$$\{|f_n - f| > \varepsilon\} = \{|f_n - f| > \varepsilon, f \geqslant 0\} \cup \{|f_n - f| > \varepsilon, f < 0\}. \quad (8.2.7)$$

Обозначим $A_\varepsilon^+ = \{f(x) \geqslant 0, f_n(x) < f(x) - \varepsilon\}$ (первое событие в левой части условия (В)). Рассмотрим первое из событий в правой части (8.2.7):

$$\begin{aligned}\{|f_n - f| > \varepsilon, f \geqslant 0\} &= \{f_n > f + \varepsilon, f \geqslant 0\} \cup \{f_n < f - \varepsilon, f \geqslant 0\} \\ &\subset \{|f_n| > |f| + \varepsilon, f \geqslant 0\} \cup A_\varepsilon^+.\end{aligned}$$

Вероятность первого события в правой части не превосходит $P\big(\big|\,|f_n| - |f|\,\big| > \varepsilon\big) \to 0$ вследствие доказанной выше сходимости $|f_n| \xrightarrow{p} |f|$, а $P(A_\varepsilon^+) \to 0$ по условию (В). Таким образом, вероятность первого события в правой части (8.2.7) стремится к нулю. Точно так же доказывается стремление к нулю вероятности второго события. Тем самым мы доказали, что $f_n \xrightarrow{p} f$, что завершает доказательство теоремы. $\square$

§ 8.3. Ц.п.т. и з.б.ч. специального вида

В этом параграфе мы приводим теорему из книги Гаека и Шидака [3], содержащуюся в п. 1.2 «Специальный случай центральной предельной теоремы» главы 5 книги. Мы формулируем эту теорему в еще более специализированном виде и для полноты изложения приводим доказательство, состоящее в непосредственной проверке условия Линдеберга. Кроме того, мы доказываем специальный случай закона больших чисел, тесно связанный с упомянутой теоремой.

8.3.1. Ц.п.т. специального вида. Пусть $V_1, V_2, \ldots$ – независимые одинаково распределенные случайные величины с $EV_i = 0$, $EV_i^2 = \sigma^2 < \infty$, $i = 1, 2, \ldots$, и функцией распределения $F(x) = P(V_1 < x)$. Пусть, кроме того, имеется «треугольный массив» коэффициентов $c_{N,i}$, $i = 1, \ldots, N$, $N = 1, 2, \ldots$.

Рассмотрим последовательность сумм

$$S_N = \sum_{i=1}^{N} c_{N,i} V_i, \quad N = 1, 2, \ldots . \tag{8.3.1}$$

Случайные величины $c_{N,i} V_i$, $i = 1, \ldots, N$, $N = 1, 2, \ldots$, образуют треугольный массив[8] специального вида, в котором распределения входящих в него величин отличаются только параметрами масштаба.

Теорема 8.3.1. *Предположим, что выполнены условия*

$$\sum_{i=1}^{N} c_{N,i}^2 = 1, \quad \max_{1 \leqslant i \leqslant N} |c_{N,i}| \to 0 \ \text{при} \ N \to \infty. \tag{8.3.2}$$

Тогда распределение S_N асимптотически нормально,

$$S_N \xrightarrow{d} \mathcal{N}(0, \sigma^2) \ \text{при} \ N \to \infty.$$

Прежде чем приступить к доказательству теоремы, приведем формулировку центральной предельной теоремы при условии Линдеберга (см., например, Гнеденко [4], § 47). Предположим, что мы имеем треугольный массив случайных величин $Y_{N,i}$, $i = 1, \ldots, N$, $N = 1, 2, \ldots$, удовлетворяющий условиям

$$\sum_{i=1}^{N} EY_{N,i}^2 = 1, \quad EY_{N,i} = 0, \quad i = 1, \ldots, N, \quad N = 1, 2, \ldots \tag{8.3.3}$$

Теорема 8.3.2 (теорема Линдеберга–Феллера)[9]. *При выполнении условий (8.3.3) для сходимости распределений сумм*

[8] Распространеным названием для треугольного массива случайных величин является «схема серий»; в курсе Гнеденко [4] используется название «элементарная система».

[9] Линдебергом в 1922 г. была доказана достаточность условия (8.3.4), а Феллером в 1935 г. – его необходимость.

$S_N = \sum_{i=1}^N Y_{N,i}$ *к нормальному закону* $\mathcal{N}(0,1)$ *необходимо и достаточно, чтобы для любого* $\varepsilon > 0$

$$\sum_{i=1}^{N} EY_{N,i}^2 \mathbf{1}_{\{|Y_{N,i}|>\varepsilon\}} \to 0 \ \textit{при} \ N \to \infty. \tag{8.3.4}$$

Доказательство теоремы 8.3.1. Применим теорему 8.3.2, полагая $Y_{N,i} = c_{N,i}V_i/\sigma$. Это значит, что нам надо проверить условие

$$\sum_{i=1}^{N} c_{N,i}^2 EV_i^2 \mathbf{1}_{\{|c_{N,i}V_i|/\sigma>\varepsilon\}} \to 0 \ \text{при} \ N \to \infty. \tag{8.3.5}$$

Обозначим $C_N = \max_{1 \leqslant i \leqslant N} |c_{N,i}|/\sigma$. Тогда

$$\mathbf{1}_{\{|c_{N,i}V_i|/\sigma>\varepsilon\}} \leqslant \mathbf{1}_{\{C_N|V_i|>\varepsilon\}}, \quad i = 1, \dots, N.$$

Поскольку V_i одинаково распределены, сумма в (8.3.5) не превосходит

$$EV_1^2 \mathbf{1}_{\{C_N|V_1|>\varepsilon\}} \sum_{i=1}^{N} c_{N,i}^2 = EV_1^2 \mathbf{1}_{\{C_N|V_1|>\varepsilon\}} \tag{8.3.6}$$

по условию (8.3.2), при этом в силу того же условия $C_N \to 0$ при $N \to \infty$. Но

$$EV_1^2 \mathbf{1}_{\{C_N|V_1|>\varepsilon\}} = \int_{\{C_N|x|>\varepsilon\}} x^2 \, dF(x).$$

Этот интеграл стремится к нулю, поскольку интегрирование производится по внешности расширяющегося интервала с границами $\pm\varepsilon/C_N \to \pm\infty$, в то время как интеграл по всей прямой (равный EV_1^2) сходится. Этим (8.3.5) доказано, что завершает доказательство теоремы. □

8.3.2. З.б.ч. специального вида. Теперь мы установим закон больших чисел, относящийся к сумме квадратов $c_{N,i}^2 X_i^2$ элементов того же треугольного массива. Сначала сформулируем общую теорему.

Теорема 8.3.3. *Предположим, что случайные величины* $Y_{N,i}$, $i = 1, \dots, N$, $N = 1, 2, \dots$, *удовлетворяют условиям* (8.3.3) *и* $\max_{1\leqslant i\leqslant N} P(|Y_{N,i}| > \varepsilon) \to 0$ *при* $N \to \infty$ *для любого* $\varepsilon > 0$. *Тогда распределения сумм* $S_N = \sum_{i=1}^N Y_{N,i}$ *сходятся к нормальному* $\mathcal{N}(0,1)$ *в том и только том случае, если*

$$\sum_{i=1}^N Y_{N,i}^2 \xrightarrow{p} 1 \ \textit{при}\ N \to \infty.$$

Доказательство можно найти в книге Гнеденко и Колмогорова [5] (§ 28, теорема 4). Для иллюстрации этой теоремы рассмотрим простейший случай последовательности независимых одинаково распределенных случайных величин $X_1, X_2, \ldots$ с $EX_1 = 0$, $EX_1^2 = 1$. Тогда, как известно,

$$\mathcal{L}\Big(\frac{1}{\sqrt{N}}\sum_{i=1}^N X_i\Big) \to \mathcal{N}(0,1).$$

Суммы в левой части представляют собой суммы по строкам треугольного массива $\widetilde{X}_{N,i} = N^{-1/2}X_i$, $i = 1,\ldots,N$, $N = 1,2,\ldots$, при этом $\sum_{i=1}^N \widetilde{X}_{N,i}^2 = N^{-1}\sum_{i=1}^N X_i^2 \to EX_1^2 = 1$ по закону больших чисел.

Следующая теорема является непосредственным следствием теорем 8.3.1 и 8.3.3. В ней рассматриваются независимые одинаково распределенные случайные величины $X_1, X_2, \ldots$ и коэффициенты $\{c_{N,i}\}$, введенные в начале этого пункта.

Теорема 8.3.4. *Предположим, что* $X_1, X_2, \ldots$ – *независимые одинаково распределенные случайные величины с* $EX_1 = 0$, $EX_1^2 = \sigma^2 < \infty$, *а* $\{c_{N,i}\}$ *удовлетворяют условиям* (8.3.2). *Тогда*

$$\sum_{i=1}^N c_{N,i}^2 X_i^2 \xrightarrow{p} \sigma^2 \ \textit{при}\ N \to \infty.$$

Теорема 8.3.4 неоднократно используется в главе 7. Она получена как простое следствие теоремы 8.3.3, доказательство которой, однако, использует достаточно сложный аппарат общей теории суммирования независимых случайных величин. Читателю, незнакомому с этой теорией, приходится принять теорему 8.3.3 и ее следствие, теорему 8.3.4, на веру. Поэтому мы приведем самостоятельное, достаточно простое, доказательство теоремы 8.3.4. Оно являющееся модификацией доказательства теоремы Хинчина, содержащейся в § 32 курса Гнеденко [4].

В этом доказательстве тот факт, что в сумму входят квадраты случайных величин, не играет роли. Поэтому в следующей теореме 8.3.5 мы будем рассматривать произвольную последовательность независимых одинаково распределенных случайных величин $Y_1, Y_2, \ldots$ с конечным математическим ожиданием. Теорема 8.3.4 тогда получается как специальный случай теоремы 8.3.5 с $Y_i = X_i^2$.

Теорема 8.3.5. *Предположим, что $Y_1, Y_2, \ldots$ – независимые одинаково распределенные случайные величины с $b := E|Y_1| < \infty$, а $\{c_{N,i}\}$ удовлетворяют условиям (8.3.2) и все отличны от нуля. Обозначим $a = EY_1$. Тогда*

$$\sum_{i=1}^{N} c_{N,i}^2 Y_i \xrightarrow{p} a \ \text{ при } \ N \to \infty. \tag{8.3.7}$$

Доказательство. В силу условия $\sum c_{N,i}^2 = 1$ (см. (8.3.2)) доказываемое соотношение (8.3.7) равносильно соотношению

$$\sum_{i=1}^{N} c_{N,i}^2 (Y_i - a) \xrightarrow{p} 0.$$

Поэтому, переходя к величинам $Y_i - a$, мы будем считать, что $EY_1 = 0$, и доказывать, что

$$\sum_{i=1}^{N} c_{N,i}^2 Y_i \xrightarrow{p} 0 \ \text{ при } \ N \to \infty. \tag{8.3.8}$$

Определим случайные величины

$$Y'_{N,i} = Y_i \mathbf{1}_{\{|Y_i| < 1/c_{N,i}^2\}}, \quad Y''_{N,i} = Y_i \mathbf{1}_{\{|Y_i| \geqslant 1/c_{N,i}^2\}}, \quad i = 1, \ldots, N. \tag{8.3.9}$$

Очевидно, что $Y_i = Y'_{N,i} + Y''_{N,i}$, $i = 1, \ldots, N$. Величины $Y'_{N,i}$ имеют конечные математические ожидания и дисперсии:

$$a_{N,i} := EY'_{N,i} = \int_{-1/c_{N,i}^2}^{1/c_{N,i}^2} y \, dF(y), \tag{8.3.10}$$

$$\operatorname{var} Y'_{N,i} = b_{N,i} - a_{N,i}^2, \quad b_{N,i} = \int_{-1/c_{N,i}^2}^{1/c_{N,i}^2} y^2 \, dF(y), \tag{8.3.11}$$

где $F(y)$ обозначает функцию распределения Y_1.

При $N \to \infty$ имеем $a_{N,i} \to 0$ равномерно по i, т.е.

$$\max_{1 \leqslant i \leqslant N} |a_{N,i}| \to 0.$$

Действительно, положим $C_N = \max_{1 \leqslant i \leqslant N} |c_{N,i}|$. Тогда, учитывая, что $EY_1 = \int_{\mathbb{R}} y\, dF(y) = 0$, имеем

$$a_{N,i} = -\int_{\{|y| \geqslant 1/c_{N,i}^2\}} y\, dF(y)$$

и следовательно

$$|a_{N,i}| \leqslant \int_{\{|y| \geqslant 1/c_{N,i}^2\}} |y|\, dF(y) \leqslant \int_{\{|y| \geqslant 1/C_N^2\}} |y|\, dF(y) \to 0 \quad (8.3.12)$$

при $N \to \infty$ равномерно по i в силу условия $C_N \to 0$. Обозначим

$$a_N = E\left(\sum_{i=1}^N c_{N,i}^2 Y'_{N,i}\right) = \sum_{i=1}^N c_{N,i}^2 a_{N,i}.$$

В силу (8.3.12)

$$a_N \leqslant \max_i |a_{N,i}| \sum_{i=1}^N c_{N,i}^2 \to 0. \quad (8.3.13)$$

Оценим дисперсию суммы $\sum_{i=1}^N c_{N,i}^2 Y'_{N,i}$. Пользуясь (8.3.11), имеем

$$\operatorname{var}\left(\sum_{i=1}^N c_{N,i}^2 Y'_{N,i}\right) = \sum_{i=1}^N c_{N,i}^4 b_{N,i} - \sum_{i=1}^N c_{N,i}^4 a_{N,i}^2. \quad (8.3.14)$$

Вторая сумма не превосходит $C_N^2 \max_{1 \leqslant i \leqslant N} a_{N,i}^2 \sum_{i=1}^N c_{N,i}^2 \to 0$. Покажем, что первая сумма стремится к нулю. Разобьем интеграл, определяющий $b_{N,i}$ (см. (8.3.11)), на сумму интегралов по областям $\{|y| \leqslant 1/\sqrt{|c_{N,i}|}\}$ и $\{1/\sqrt{|c_{N,i}|} < |y| \leqslant 1/c_{N,i}^2\}$, которые обозначим $b'_{N,i}$ и $b''_{N,i}$, так что $b_{N,i} = b'_{N,i} + b''_{N,i}$. Имеем $b'_{N,i} \leqslant 1/|c_{N,i}|$, поэтому

$$\sum_{i=1}^N c_{N,i}^4 b'_{N,i} \leqslant C_N \sum_{i=1}^N c_{N,i}^2 \to 0.$$

Далее,

$$b''_{N,i} \leqslant \frac{1}{c^2_{N,i}} \int_{|y|>1/\sqrt{|c_{N,i}|}} |y|\, dF(y) \leqslant \frac{1}{c^2_{N,i}} \int_{|y|>1/\sqrt{C_N}} |y|\, dF(y) = \frac{1}{c^2_{N,i}} o(1).$$

Сдедовательно,

$$\sum_{i=1}^{N} c^4_{N,i} b''_{N,i} \leqslant \sum_{i=1}^{N} c^2_{N,i} \cdot o(1) \to 0.$$

Подставляя полученные оценки в (8.3.14), получаем

$$\operatorname{var}\left(\sum_{i=1}^{N} c^2_{N,i} Y'_{N,i}\right) \to 0.$$

С учетом полученной ранее сходимости к нулю математического ожидания a_N той же суммы (см. (8.3.13)) отсюда следует

$$\sum_{i=1}^{N} c^2_{N,i} Y'_{N,i} \xrightarrow{p} 0. \tag{8.3.15}$$

Остается рассмотреть $\sum_i c^2_{N,i} Y''_{N,i}$ (см. (8.3.9)). Оценим вероятность того, что $Y''_{N,i}$ отлична от нуля. Она не превосходит $P(|Y_i| \geqslant 1/c^2_{N,i})$. Имеем

$$P\big(|Y_i| \geqslant 1/c^2_{N,i}\big) \leqslant c^2_{N,i} \int_{|Y_i| \geqslant 1/c^2_{N,i}} |y|\, dF(y) \leqslant c^2_{N,i} \int_{|Y_i| \geqslant 1/C^2_N} |y|\, dF(y).$$

Последний интеграл стремится к нулю равномерно по $i = 1, \ldots, N$, следовательно, $\sum_i P(Y''_{N,i} \neq 0) \to 0$, а это значит, что вероятность наличия хотя бы одного ненулевого слагаемого в сумме $\sum_i c^2_{N,i} Y''_{N,i}$ стремится к нулю, следовательно эта сумма стремится к нулю по вероятности. Вместе с (8.3.15) это доказывает (8.3.8). Тем самым теорема доказана. □

Список литературы

[1] Андронов А. М., Копытов Е. А. и Гринглаз Л. Я., *Теория вероятностей и математическая статистика*, Питер, СПб, 2004.

[2] Биллингсли П., *Сходимость вероятностных мер*, Наука, Физматлит, М., 1977 MR 0494353.

[3] Гаек Я. и Шидак З., *Теория ранговых критериев*, Наука, М., 1971.

[4] Гнеденко Б. В., *Курс теории вероятностей*, 9-е изд., URSS, М., 2007 MR 0139186.

[5] Гнеденко Б. В., Колмогоров А. Н., *Предельные распределения для сумм независимых случайных величин*, Гостехиздат, М.–Л., 1949.

[6] Колмогоров А. Н., Фомин С. В., *Элементы теории функций и функционального анализа*, 6-е изд., Наука, Физматлит, М., 1989 MR 1025126.

[7] Леман Э., *Проверка статистических гипотез*, 2-е изд., Наука, Физматлит, М., 1979 MR 0555243.

[8] Лоэв М., *Теория вероятностей*, Изд-во иностр. лит., М., 1962 MR 0140129.

[9] Никитин Я. Ю., *Асимптотическая эффективность непараметрических критериев*, Наука, Физматлит, М., 1995 MR 1415127.

[10] Прохоров Ю. В., “Сходимость случайных процессов и предельные теоремы теории вероятностей”, *Теория вероятностей и ее примен.*, **1** (1956), 178–238.

[11] Русас Дж., *Контигуальность вероятностных мер*, Мир, М., 1975 MR 0375543.

[12] Хеттманспергер Т., *Статистические выводы, основанные на рангах*, Финансы и статистика, М., 1987 MR 0923535, Zbl 0665.62039.

[13] Холлендер М., Вулф Д., *Непараметрические методы статистики*, Финансы и статистика, М., 1983 Zbl 0541.62024.

[14] Чибисов Д. М., “Асимптотические разложения в задачах проверки гипотез. II”, *Известия АН УзССР, сер. физ.-мат. наук*, 1982, № 6, 23–30 MR 692126, Zbl 0532.62013.

[15] Ширяев А. Н., *Вероятность*, Кн. 1: *Элементарная теория вероятностей. Математические основания. Предельные теоремы*, МЦНМО, М., 2007.

[16] Ширяев А. Н., *Вероятность*, Кн. 2: *Суммы и последовательности случайных величин – стационарные, мартингалы, марковские цепи*, МЦНМО, М., 2007.

[17] Albers W., Bickel P. J., van Zwet W. R., “Asymptotic expansions for the power of distribution free tests in the one-sample problem”, *Ann. Statist.*, **4** (1976), 108–156 doi 10.1214/aos/1176343350, MR 391373, Zbl 0321.62049.

[18] Bening V. E., “A formula for deficiency: one sample L- and R-tests I, II”, *Math. Methods Statist.*, **4** (1995), 167–188 MR 1335153, Zbl 0833.62043; 274–293 MR 1355249, Zbl 0836.62036.

[19] Bickel P. J., van Zwet W. R., “Asymptotic expansions for the power of distribution free tests in the two-sample problem”, *Ann. Statist.*, **6** (1978), 937–1004 doi 10.1214/aos/1176344305, MR 499567, Zbl 0378.62047.

[20] Chernoff H., Savage I. R., “Asymptotic normality and efficiency of certain non-parametric test statistics”, *Ann. Math. Statist.*, **29** (1958), 972–994 doi 10.1214/aoms/1177706436, MR 100322.

[21] Chibisov D. M., “Asymptotic expansions and deficiencies of tests”, *Proc. Internat. Congress of Mathematicians*, v. 2 (Warszawa, 1984), 1063–1079 MR 0804758, Zbl 0561.62021.

[22] Hodges J. L., Jr., Lehmann E. L., “The efficiency of some nonparametric competitors of the t-test”, *Ann. Math. Statist.*, **27** (1956), 324–335 doi 10.1214/aoms/1177728261, MR 79383, Zbl 0075.29206.

[23] Hodges J. L., Lehmann E. L., “Deficiency”, *Ann. Math. Statist.*, **41** (1970), 783–801 doi 10.1214/aoms/1177696959, MR 272092, Zbl 0225.62063.

[24] Le Cam L., “Locally asymptotically normal families of distributions”, *Univ. of California Publ. in Statist.*, **3** (1960), 37–98 MR 0126903, Zbl 0104.12701.

[25] Mann H. B., Whitney D. R., “On a test whether one of two random variables is stochastically larger than the other”, *Ann. Math. Statist.*, **18** (1947), 50–60 doi 10.1214/aoms/1177730491, MR 0022058, Zbl 0041.26103.

[26] Wilcoxon F., “Individual comparisons by ranking methods”, *Biometrics Bull.*, **1** (1945), 80–83 doi 10.2307/3001968.

Научное издание

Лекционные курсы НОЦ

Выпуск 14

Чибисов Дмитрий Михайлович

Лекции по асимптотической теории ранговых критериев

Сдано в набор 16.11.2009. Подписано в печать 23.12.2009.
Формат 60×90/16. Усл. печ. л. 11. Тираж 200 экз.

Отпечатано в Математическом институте им. В. А. Стеклова РАН
Москва, 119991, ул. Губкина, 8.
`http://www.mi.ras.ru/noc/` e-mail: `pavlov@mi.ras.ru`