ФЕДЕРАЛЬНОЕ АГЕНТСТВО

ПО ТЕХНИЧЕСКОМУ РЕГУЛИРОВАНИЮ И МЕТРОЛОГИИ

РЕКОМЕНДАЦИИ
ПО СТАНДАРТИЗАЦИИ

**Р 50.1.100—
2014**

Статистические методы

ТРИ ПОДХОДА К ИНТЕРПРЕТАЦИИ И ОЦЕНКЕ НЕОПРЕДЕЛЕННОСТИ ИЗМЕРЕНИЙ

**ISO/TR 13587:2012
(NEQ)**

Издание официальное

**Москва
Стандартинформ
2015**

gost.e-ecolog.ru

Предисловие

1 РАЗРАБОТАНЫ Открытым акционерным обществом «Научно-исследовательский центр контроля и диагностики технических систем» (АО «НИЦ КД»)

2 ВНЕСЕНЫ Техническим комитетом по стандартизации ТК 125 «Применение статистических методов»

3 УТВЕРЖДЕНЫ И ВВЕДЕНЫ В ДЕЙСТВИЕ Приказом Федерального агентства по техническому регулированию и метрологии от 11 ноября 2014 г. № 1579-ст

4 Настоящие рекомендации разработаны с учетом основных нормативных положений международного документа ISO/TR 13587:2012 «Три статистических подхода к оценке и интерпретации неопределенности измерений» (ISO/TR 13587:2012 «Three statistical approaches for the assessment and interpretation of measurement uncertainty», NEQ)

5 ВВЕДЕНЫ ВПЕРВЫЕ

*Правила применения настоящих рекомендаций установлены в ГОСТ Р 1.0—2012 (раздел 8). Информация об изменениях к настоящим рекомендациям публикуется в ежегодном (по состоянию на 1 января текущего года) информационном указателе «Национальные стандарты», а официальный текст изменений и поправок — в ежемесячном информационном указателе «Национальные стандарты». В случае пересмотра (замены) или отмены настоящих рекомендаций соответствующее уведомление будет опубликовано в ближайшем выпуске ежемесячного информационного указателя «Национальные стандарты». Соответствующая информация, уведомление и тексты размещаются также в информационной системе общего пользования — на официальном сайте Федерального агентства по техническому регулированию и метрологии в сети Интернет (*gost.e-ecolog.ru

© Стандартинформ, 2015

Настоящие рекомендации не могут быть полностью или частично воспроизведены, тиражированы и распространены в качестве официального издания без разрешения Федерального агентства по техническому регулированию и метрологии

Содержание

1 Область применения 1
2 Нормативные ссылки 1
3 Термины и определения 2
4 Обозначения и сокращения 2
5 Описание задачи 3
6 Статистические подходы 3
7 Примеры 5
8 Частотный подход 6
9 Байесовский подход 13
10 Фидуциальный подход 17
11 Обсуждение результатов 21
12 Заключение 26
Библиография 27

Введение

Принятие Руководства ИСО/МЭК 98-3 (GUM)[1)] привело к возрастающему признанию необходимости включать указание неопределенности в результаты измерений. Аккредитация лабораторий на основе ГОСТ ИСО/МЭК 17025[2)] ускорила этот процесс. Признавая, что указание неопределенности необходимо для принятия решений, метрологи в лабораториях всех типов (от национальных институтов метрологии до коммерческих лабораторий калибровки) проявляют значительные усилия по разработке соответствующих оценок неопределенности для различных типов измеряемых величин и методов, приведенных в GUM.

Некоторым преимуществом процедур, описанных и популяризированных в GUM, является стандартизированный подход к определению оценки неопределенности с адаптацией к источникам неопределенности, которая может быть статистической (тип А) или нестатистической (тип В), с акцентом на отчетах обо всех источниках рассматриваемой неопределенности. В основе подхода распространения неопределенности GUM лежит линейная аппроксимация функции измерений. Во многих практических ситуациях такой подход дает результаты, аналогичные полученным более формальными методами. Таким образом, принятие GUM, произвело революцию в оценке неопределенности.

Конечно, необходимо много усилий для улучшения оценки неопределенности в практических ситуациях. Совместный комитет по руководствам в метрологии (JCGM), ответственный за GUM с 2000 года, закончил Дополнение 1 к GUM, а именно «Неопределенность измерения. Часть 3. Руководство по выражению неопределенности измерения. Дополнение 1. Трансформирование распределений с использованием метода Монте-Карло» (называемый GUMS1). В настоящее время JCGM разрабатывает также другие дополнения к GUM в таких направлениях, как моделирование и модели с любым количеством выходных величин.

Применительно к широкому кругу измерительных задач в Руководстве ИСО/МЭК 99:2007 (см. [4]) приведено достаточно общее определение неопределенности измерения как неотрицательного параметра, характеризующего разброс значений, приписываемых измеряемой величине, на основе используемой информации. В результате определение и понимание функций различных статистических величин при определении оценки неопределенности, даже в хорошо понятных применениях измерений особенно интересны как статистикам, так и метрологам.

Ранее проводились исследования этих проблем с метрологической точки зрения. Некоторые авторы исследовали статистические свойства процедур, установленных в GUM. В [5] показано, что к этим процедурам непосредственно не применимы байесовская и фидуциальная интерпретация. В [6] предложено несколько модифицированных процедур GUM, которые дают результаты, более согласованные с интерпретацией Байеса в некоторых случаях. В [7] рассмотрено соотношение между процедурами определения оценки неопределенности, предложенной в GUMS1 (см. [3]) и результатами байесовского анализа для моделей особого вида. В [8] рассмотрены возможные вероятностные интерпретации интервалов охвата и даны рекомендации по аппроксимации апостериорного распределения для этого класса байесовского анализа распределений вероятностей семейства распределений Пирсона.

В [9] приведено сопоставление частотного и байесовского подходов для определения оценки неопределенности. Однако исследование выполнено только для измерительных систем, причем для всех источников неопределенности могут быть использованы оценки типа А. Напротив, в настоящих рекомендациях рассмотрены и иллюстрированы несколькими примерами измерительные системы с источниками неопределенности, для которых использованы оценки типа А и В.

Статистики потратили много сил на использование методов определения оценок неопределенности, имеющих вероятностное обоснование или интерпретацию. В результате их работы (часто вне метрологии) было разработано несколько подходов, относящихся к оценке неопределенности. В настоящих рекомендациях представлены некоторые из этих подходов и со статистической точки зрения рассмотрена их связь с методами, используемыми в настоящее время в метрологии. Статистическими подходами, для которых описаны различные методы определения оценки неопределенности, являются частотный, байесовский и фидуциальный подходы, рассмотренные в настоящих рекомендациях.

[1)] Национальный стандарт ГОСТ Р 54500.3—2011/Руководство ИСО/МЭК 98-3:2008 «Неопределенность измерения. Часть 3. Руководство по выражению неопределенности измерения» идентичен ISO/IEC Guide 98-3:2008 (см. [1]).

[2)] Национальный стандарт ГОСТ ИСО/МЭК 17025—2009 «Общие требования к компетентности испытательных и калибровочных лабораторий» идентичен ISO/IEC 17025:2005 (см. [2]).

РЕКОМЕНДАЦИИ ПО СТАНДАРТИЗАЦИИ

Статистические методы

ТРИ ПОДХОДА К ИНТЕРПРЕТАЦИИ И ОЦЕНКЕ НЕОПРЕДЕЛЕННОСТИ ИЗМЕРЕНИЙ

Statistical methods. Three approaches for the interpretation and assessment of measurement uncertainty

Дата введения — 2015—12—01

1 Область применения

В настоящих рекомендациях рассмотрены три основных статистических подхода к интерпретации и оценке неопределенности измерений: частотный подход, байесовский подход и фидуциальный подход. Общая черта этих подходов — четкая вероятностная интерпретация интервалов неопределенности. Для каждого подхода описаны основной метод, предположения и вероятностная интерпретация неопределенности. В настоящих рекомендациях также рассмотрено соотношение этих статистических подходов с методами, предложенными в ГОСТ Р 54500.3—2011 (далее GUM).

2 Нормативные ссылки

В настоящих рекомендациях использованы нормативные ссылки на следующие стандарты:

ГОСТ ИСО/МЭК 17025—2009 Общие требования к компетентности испытательных и калибровочных лабораторий

ГОСТ Р 50779.10—2000 Статистические методы. Вероятность и основы статистики. Термины и определения

ГОСТ Р 50779.11—2000 Статистические методы. Статистическое управление качеством. Термины и определения

ГОСТ Р 54500.3—2011/Руководство ИСО/МЭК 98-3:2008 Неопределенность измерения. Часть 3. Руководство по выражению неопределенности измерения

ГОСТ Р 54500.3.1—2011/Руководство ИСО/МЭК 98-3:2008 Дополнение 1:2008 Неопределенность измерения. Часть 3. Руководство по выражению неопределенности измерения по выражению неопределенности измерения. Дополнение 1. Трансформирование распределений с использованием метода Монте-Карло

Примечание — При пользовании настоящими рекомендациями целесообразно проверить действие ссылочных стандартов в информационной системе общего пользования — на официальном сайте Федерального агентства по техническому регулированию и метрологии в сети Интернет или по ежегодному информационному указателю «Национальные стандарты, который опубликован по состоянию на 1 января текущего года, и по выпускам ежемесячного информационного указателя «Национальные стандарты» за текущий год. Если заменен ссылочный стандарт, на который дана недатированная ссылка, то рекомендуется использовать действующую версию этого стандарта с учетом всех внесенных в данную версию изменений. Если заменен ссылочный стандарт, на который дана датированная ссылка, то рекомендуется использовать версию этого стандарта с указанным выше годом утверждения (принятия). Если после утверждения настоящих рекомендаций в ссылочный стандарт, на который дана датированная ссылка, внесено изменение, затрагивающее положение, на которое дана ссылка, то это положение рекомендуется применять без учета данного изменения. Если ссылочный стандарт отменен без замены, то положение, в котором дана ссылка на него, рекомендуется применять в части, не затрагивающей эту ссылку.

Издание официальное

3 Термины и определения

В настоящих рекомендациях применены термины по ГОСТ Р 50779.10, ГОСТ Р 50779.11, а также следующие термины с соответствующими определениями.

3.1 **эмпирическая функция распределения, эмпирическая интегральная (кумулятивная) функция распределения** (empirical distribution function, empirical cumulative distribution function): Функция распределения, присваивающая вероятность $1/n$ каждому из n элементов случайной выборки и представляющая собой ступенчатую функцию вида

$$F_n(x) = \frac{|\{x_i \le x\}|}{n},$$

где $\{x_1, ..., x_n\}$ — выборка, и $|A|$ — количество элементов, удовлетворяющих условию A.

3.2 **байесовский анализ** (Bayesian sensitivity analysis): Анализ влияния априорных распределений параметров статистической модели на апостериорное распределение измеряемой величины.

3.3 **достаточная статистика** (sufficient statistic): Функция выборки $X_1, ..., X_n$ из распределения, функция плотности вероятностей которой зависит от параметра θ, а условное распределение $X_1, ..., X_n$ при заданном значении статистики не зависит от θ.

Примечание — Достаточная статистика содержит всю информацию о параметре θ, как функции $X_1, ..., X_n$.

3.4 **модель наблюдений** (observation model): Математическая связь результатов измерений (наблюдений) измеряемой величины и соответствующей случайной ошибки погрешности измерений.

3.5 **структурное уравнение** (structural equation): Статистическая модель, связывающая наблюдаемую случайную величину с неизвестными параметрами и ненаблюдаемой случайной величиной, распределение которой известно и не зависит от неизвестных параметров.

3.6 **нецентральное хи-квадрат распределение** (non-central chi-squared distribution): Нецентральное распределение вероятностей, представляющее собой обобщение центрального χ^2-распределения.

Примечание 1 — Для k нормально распределенных случайных величин X_i со средним μ_i и дисперсией σ_i^2 случайная величина $X = \sum_{i=1}^{k} (X_i / \sigma_i)^2$ имеет нецентральное χ^2-распределение. Нецентральное χ^2-распределение имеет два параметра: k — число степеней свободы (количество X_i) и λ, который зависит от средних случайных величин $X_i \left(\lambda = \sum_{i=1}^{k} (\mu_i / \sigma_i)^2 \right)$ и называется параметром нецентральности.

Примечание 2 — Плотность χ^2-распределения представляет собой смесь плотностей центральных распределений:

$$g_X(\xi) = \sum_{i=0}^{\infty} \frac{e^{-\lambda/2}(\lambda/2)^i}{i!} g_{Y_{k+2i}}(\xi) = \frac{e^{\frac{-(\xi+\lambda)}{2}}}{2^{k/2}} \sum_{i=0}^{\infty} \frac{\xi^{k/2+i-1}\lambda^i}{\Gamma(k/2+i)2^{2i}i!},$$

где Y_q подчиняется χ^2-распределению с q степенями свободы.

4 Обозначения и сокращения

В настоящих рекомендациях использованы для обозначений греческие и латинские буквы. Греческие буквы использованы для обозначения параметров статистической модели (например, измеряемых величин), которые могут быть и случайными величинами и постоянными величинами в зависимости от используемого статистического подхода и модели. Прописные латинские буквы использованы для обозначения случайных величин, которые могут принимать различные значения при наблюдениях, и строчные латинские буквы для обозначения наблюдаемых значений величин (например, результатов измерений). В некоторых случаях использованы другие обозначения. Однако в этом случае смысл обозначений ясен из контекста.

5 Описание задачи

5.1 В настоящих рекомендациях рассмотрена модель измерений:

$$\theta = f(\mu_1, ..., \mu_p), \quad (1)$$

где $\mu_1, ..., \mu_p$ — входные величины;

θ — выходная величина;

f — функция измерений.

Функция f определена математически в виде формулы или алгоритма вычислений. В GUM (примечание 1, 4.1) те же самые функциональные зависимости определены соотношением

$$Y = f(X_1, ..., X_p), \quad (2)$$

которое сложно отличить от функции измерений, определяющей зависимость случайной величины от результатов наблюдений входной величины.

В соответствии с процедурой, рекомендованной GUM, для p неизвестных величин определяют оценки $\mu_1, ..., \mu_p$ по значениям $x_1, ..., x_p$, полученным при выполнении измерений или из других источников. Соответствующие стандартные неопределенности также получают по имеющимся данным с помощью статистических методов или плотностей вероятностей, построенных на основе экспертных знаний о переменных. В GUM (см. также п. 4.5 в [11]) модель измерений, связывающую измеряемую величину θ с входными величинами $\mu_1, ..., \mu_p$, рекомендовано использовать также для вычисления функции, описывающей зависимость y от $x_1, ..., x_p$. Таким образом, результат измерений (или оценка) y для θ имеют вид

$$y = f(x_1, ..., x_p), \quad (3)$$

т. е. оценка Y, $y = f(x_1, ..., x_p)$ представляет собой результат измерений θ. Оценки y, $x_1, ..., x_p$ являются реализациями случайных величин Y, $X_1, ..., X_p$ соответственно.

5.2 В настоящих рекомендациях приведено три статистических подхода, обеспечивающих определение:

(а) наилучшей оценки y для θ,

(б) соответствующей стандартной неопределенности $u(y)$,

(в) доверительного интервала или интервала охвата для θ с заданной вероятностью охвата (обычно 95 %).

5.3 Необходимо различать оценки стандартной неопределенности, соответствующие оценкам различных величин и соответствующие теоретические значения стандартной неопределенности. Теоретические значения стандартных неопределенностей обозначены соответственно σ_μ или σ_x, их оценки до и после наблюдений обозначены S_X и s_x соответственно.

6 Статистические подходы

6.1 Частотный подход

6.1.1 Статистический подход, позволяющий определить вероятностную оценку неопределенности, называют частотным. Этот подход иногда называют «классическим» или «общепринятым». Однако в силу особенностей неопределенности в метрологических задачах методы этого семейства для определения частотного интервала неопределенности в реальных условиях часто требуют адаптации.

6.1.2 При использовании частотного подхода входные значения $\mu_1, ..., \mu_p$ модели измерений (1) и выходную величину θ рассматривают как неизвестные постоянные величины. Полученные для каждой величины μ_i данные используют для определения оценки θ с помощью модели измерений или других статистических моделей. Для определения оценки θ с помощью использования одного из математических методов (наименьших квадратов, максимального правдоподобия или бутстреп-метода) определяют доверительные интервалы с заданным уровнем доверия.

6.1.3 Поскольку θ рассматривают как постоянную величину, вероятностное утверждение, относящееся к доверительному интервалу для θ, не является прямым утверждением относительно значения θ. Это утверждение лишь указывает, как часто доверительный интервал, полученный с применением данной процедуры, накрывает измеряемую величину при многократном повторении процедуры. Повторение процедуры означает, что определение оценки неопределенности повторяют много раз с использованием различных данных, взятых из одних и тех же распределений. Частотный подход

обеспечивает выполнение вероятностного утверждения о свойствах процедуры построения интервала неопределенности в конкретных условиях процесса измерений на достаточно большом количестве повторений процедуры.

6.1.4 В большинстве практических метрологических задач интервалы неопределенности должны учитывать как неопределенность, соответствующую оценкам величин, полученным с использованием результатов измерений, так и неопределенность, соответствующую экспертным оценкам. Для получения интервала неопределенности, аналогичного доверительному интервалу, оценки величин, не основанные на результатах измерений, рассматривают как случайные величины с распределениями вероятностей (величины, оценки которых могут быть получены с использованием статистических данных, рассматривают как неизвестные постоянные величины).

6.1.5 Традиционная частотная процедура построения доверительного интервала может быть модифицирована для обеспечения заданного уровня доверия после усреднения по возможным значениям величин, имеющих экспертные оценки [5]. Это позволяет использовать вероятностные утверждения, аналогичные утверждениям в случае доверительных интервалов для величин, которые не были измерены.

6.1.6 В таблице 1 приведено краткое описание частотного, байесовского и фидуциального подходов к оценке неопределенности.

Таблица 1 — Интерпретации частотного, байесовского и фидуциального подходов

Наименование подхода	Характеристика величин модели измерений $\theta = f(\mu_1,...,\mu_p)$	Интервал неопределенности для выходной величины θ	Примечания
Частотный	θ и μ_i — неизвестные постоянные величины	Доверительный интервал накрывает θ с заданной вероятностью, при длительном повторении процедуры	Классический частотный подход применяют для объединения неопределенностей, которые не являются статистическими оценками
Байесовский	θ и μ_i — случайные величины, распределения вероятностей которых основаны на предварительной информации о значениях входных и выходных величин	Интервал охвата для θ рассчитывают на основе апостериорного распределения θ	Возможна неоднозначность интервала, обусловленная выбором априорных распределений
Фидуциальный	μ_i — случайные величины, распределения которых получены на основе предположений о наблюдаемых данных, использованных для определения оценок μ_i и экспертных знаниях о μ_i	Интервал охвата для θ рассчитывают на основе фидуциального распределения θ	Не единственность интервала, обусловленная выбором структурного уравнения

6.2 Байесовский подход

Второй подход называют байесовским подходом в честь фундаментальной теоремы Байеса [12], на которой он основан. В этом подходе параметры модели измерений (1) μ_1, ..., μ_p рассматривают как случайные величины с соответствующими распределениями вероятностей. Теорема Байеса позволяет получить распределение вероятностей на основе данных наблюдений и параметров, определенных в соответствии с функцией *f* или эквивалентными статистическими моделями. Полученное распределение вероятностей учитывает знания о распределении и информацию о наблюдаемых данных. Из этого распределения могут быть получены интервалы неопределенности, которые накрывают θ с заданной вероятностью. Поскольку знания о параметрах заданы в виде распределений вероятностей, байесовский метод обеспечивает возможность прямых вероятностных утверждений о значениях θ и других параметров, используя определение вероятности, как меры уверенности.

6.3 Фидуциальный подход

6.3.1 Фидуциальный подход разработан Р. Фишером [13] в 1930-ых годах. В этом подходе распределение вероятностей для θ, названное фидуциальным распределением, является условным (по данным) и получено на основе взаимосвязи θ и μ_p описанной функцией *f*, предположениями о рас-

пределении данных, используемых для определения оценки μ_f. Фидуциальное распределение может быть использовано для определения интервалов неопределенности, которые содержат θ с заданной вероятностью.

6.3.2 Обоснование процесса определения фидуциального распределения иллюстрирует следующий пример. Предположим, что величину Y можно описать уравнением $Y = \mu + Z$, где μ — измеряемая величина, Z — случайная величина, подчиняющаяся нормированному нормальному распределению. Если y — реализация случайной величины Y, а z — реализация случайной величины Z, то $\mu = y - z$. Знание распределения z позволяет определить совокупность возможных значений μ. Распределение вероятностей Z может быть использовано для вывода распределения вероятностей μ. Процесс преобразования соотношения $\mu = y - z$ в соотношение $\mu = y - Z$ и есть суть фидуциального подхода. Фидуциальное распределение μ представляет собой распределение вероятностей случайной величины $(y - Z)$ при фиксированном y.

7 Примеры

7.1 Общие положения

Примеры связаны с корректировкой некоторой физической величины на фоне помех. В таблице 2 приведено описание и обозначение используемых величин, в 7.2—7.4 приведены примеры определения оценок.

Примечание — Описание величин, несущественных для целей настоящего примера, не приведено.

Таблица 2 — Пояснения к примеру

Величина	Обозначение
Исследуемая физическая величина (измеряемая)	θ
Величина, обнаруженная методом измерений при измерении фонового шума (т. е. среднее B)	β
Исследуемая физическая величина, обнаруженная методом измерений (т. е. среднее Y)	$y = \theta + \beta$
Стандартное отклонение метода измерений при измерении исследуемой физической величины (стандартное отклонение Y)	σ_Y
Стандартное отклонение метода измерений при измерении фонового шума (стандартное отклонение B)	σ_B

7.2 Пример а)

Наблюдаемой величиной является композиция сигнала и фонового шума. В результате измерений получено пять независимых значений. Предполагается, что каждое значение y является реализацией случайной величины Y, подчиняющейся нормальному распределению со средним $y = \theta + \beta$ и стандартным отклонением σ_Y. Результаты измерений y составили:

3,738; 3 442; 2 994; 3 637; 3 874.

Выборочные среднее и стандартное отклонение равны $\bar{y} = 3{,}537$ и $s_y = 0{,}342$.

Аналогично определено пять результатов измерений фонового шума. Эти значения, как предполагается, являются реализацией случайной величины B, подчиняющейся нормальному распределению со средним β и стандартным отклонением σ_B. Наблюдаемые значения фонового шума составили:

1,410; 1,085; 1,306; 1,137; 1,200.

Поскольку имеются результаты измерений для каждой величины, которая является источником неопределенности, то на основе данного примера может быть показана статистическая интерпретация каждого подхода.

7.3 Пример б)

Пример б) идентичен примеру а), но оценки параметров фонового шума определяют не на основе экспериментальных данных, а на основе предыдущего опыта или экспертных данных. В этом случае величина β подчиняется равномерному распределению на интервале с конечными точками 1,126 и 1,329.

Поскольку использована экспертная оценка, неопределенность, соответствующая фоновому шуму, получена с использованием оценки типа В. Пример б) ближе к реальной ситуации, чем пример а).

7.4 Пример в)

Пример в) идентичен примеру б) за исключением того, что сигнал θ ближе по характеристикам к фоновому шуму. Наблюдаемые данные «сигнал плюс фоновый шум» в этом случае составили

1,340; 1,078; 1,114; 1,256; 1,192.

Для сигнала, почти совпадающего с фоновым шумом, в примере в) показано, как физические ограничения могут быть использованы при определении оценки неопределенности в каждом подходе.

8 Частотный подход

8.1 Основной метод

8.1.1 При частотном подходе параметры рассматривают как неизвестные постоянные величины. Далее случайные переменные обозначены прописными буквами, а соответствующие им наблюдаемые значения — строчными. Доверительный интервал может быть получен на основе функции $W(Y, \theta)$ от Y и параметра θ, которая может быть многомерной. Распределение вероятностей параметра θ не имеет неизвестных параметров (если такое распределение может быть определено). Тогда доверительный интервал уровня 100 $(1-\alpha)$ % для θ может быть определен через нижнюю и верхнюю процентили ℓ_α и u_α, удовлетворяющие условию $P_\theta(\ell_\alpha \le W(Y, \theta) \le u_\alpha) = 1-\alpha$.

8.1.2 Например, если $Y = (Y_1, ..., Y_n)$ — случайные величины, подчиняющиеся нормальному распределению $N(\mu, \sigma^2)$, то $\overline{Y} = \sum_{i=1}^{n} Y_i / n$ — также случайная величина, подчиняющаяся нормальному распределению. Пусть необходимо определить оценку μ при известном значении σ. Величина $Z = \frac{\overline{Y}-\mu}{\sigma/\sqrt{n}}$ подчиняется $N(0, 1)$.

Тогда границы доверительного интервала для μ имеют вид

$$\overline{Y} \pm \frac{\sigma}{\sqrt{n}} z_{\alpha/2}, \tag{4}$$

где $z_{\alpha/2}$ — квантиль уровня $\alpha/2$ нормированного нормального распределения.

Если σ неизвестно, можно использовать в качестве его оценки выборочное стандартное отклонение

$$S = \sqrt{\frac{\sum_{j=1}^{n}(Y_j - \overline{Y})^2}{n-1}}.$$

Оценку μ получают, заменяя σ на S.

Величина

$$\frac{\overline{Y}-\mu}{S/\sqrt{n}} \in t(n-1) \tag{5}$$

подчиняется распределению Стьюдента с числом степеней свободы $(n-1)$. Доверительный интервал для μ определяют по формуле

$$\overline{Y} \pm \frac{S}{\sqrt{n}} t_{n-1,1-\alpha/2},$$

где $t_{n-1,1-\alpha/2}$ — квантиль распределения Стьюдента с $(n-1)$ степенями свободы.

8.1.3 Вместо точных оценок, которые можно получить только в простых ситуациях, обычно используют приближенные оценки. Для больших выборок приближенные доверительные интервалы могут быть получены на основе центральной предельной теоремы.

8.1.4 Дополнительные методы определения доверительных интервалов приведены в [14]. Некоторые из них упомянуты в примерах. При построении доверительного интервала для обратных величин

с неизвестными распределениями может быть использован бутстреп-метод. Процедура бутстреп-метода приведена в 8.2.

8.1.5 При получении приближенного доверительного интервала для измеряемой величины также могут быть использованы процедуры, рекомендуемые GUM, хотя они отличаются от методов построения доверительного интервала на основе частотного подхода. Такие доверительные интервалы основаны на аппроксимации распределения функции наблюдений в модели измерений (1) распределением Стьюдента (t-распределением). В соответствии с этой процедурой оценки неизвестных величин $\mu_1, ..., \mu_p$ определяют на основе значений $x_1, ..., x_p$, полученных в результате измерений или из других источников. Значения x_i могут быть выборочными средними или другими функциями данных, используемыми для оценки μ_i, $i = 1, ..., m$. Их суммарную стандартную неопределенность $u(x_i)$ также определяют с помощью статистических методов, как правило, используя выборочное стандартное отклонение или робастные ранговые процедуры. Такие методы позволяют определить оценки неопределенности типа А. Число степеней свободы ν_i, связанное с $u(x_i)$, зависит от объема выборки, используемой для оценки μ_i.

8.1.6 Так как физические измерения не всегда возможны или целесообразны для некоторых μ_i, оценки x_i параметров μ_i для некоторых i, например $i = m + 1, ..., p$ получают с помощью субъективных (или потенциально субъективных) оценок и используют вместе с x_i, для $i = 1, ..., p$, полученными на основе оценок неопределенности типа А. Таким образом, для определения оценок $\mu_{m+1}, ..., \mu_p$ использована нестатистическая информация (данные научных исследований, требования изготовителя или другая прямая или косвенная информация) при определении оценки неопределенности типа В.

Примечание — Иногда могут быть получены оценки неопределенности типа А и В одновременно.

8.1.7 В GUM рекомендовано для вычисления y по $x_1, ..., x_p$ использовать модель, связывающую измеряемую величину θ с входными величинами $\mu_1, ..., \mu_p$. Таким образом, измеряемую величину (или ее оценку) определяют в виде

$$y = f(x_1, ..., x_m, x_{m+1}, ..., x_p),$$

т. е. как оценку Y. При этом $y = f(x_1, ..., x_p)$ — измеряемое значение θ.

8.1.8 В GUM для оценки стандартной неопределенности $u(y)$ использован закон распространения неопределенности. Стандартную неопределенность $u(x_1), .., u(x_p)$, соответствующую $x = (x_1, ..., x_p)$, определяют на основе разложения функции $f(x_1, ..., x_p)$ в ряд Тейлора первого порядка

$$f(x_1, ..., x_p) \approx f(\mu_1, ..., \mu_p) + \sum_{i=1}^{p} c_i(x_i - \mu_i), \tag{6}$$

Пусть $\mu = (\mu_1, ..., \mu_p)$. Частные производные

$$c_i = \left.\frac{\partial f}{\partial \mu_i}\right|_{\mu = x}$$

называют коэффициентами чувствительности. В соответствие с GUM применение закона распространения неопределенности дает метод определения приближенной оценки стандартной неопределенности y:

$$u(y) \approx \sqrt{\sum_{i=1}^{p} c_i^2 u^2(x_i) + 2\sum_{i<j} c_i c_j u(x_i, x_j)}, \tag{7}$$

где $u(x_i, x_j)$ — ковариация X_i и X_j.

8.1.9 Для определения оценки стандартной неопределенности $u(y)$ в GUM использована формула Велча-Саттервейта для вычисления эффективного числа степеней свободы ν_{eff}:

$$\nu_{eff} = \frac{u^4(y)}{\sum_{i=1}^{p} \frac{c_i^4 u^4(x_i)}{\nu(x_i)}}. \tag{8}$$

Примечание — В [15] рассмотрен парадокс, в соответствии с которым в межлабораторных исследованиях доверительный интервал, построенный на основе аппроксимации Велча-Саттервейта, может быть меньше для оценки различий между лабораториями, чем внутри лаборатории для компонентов неопределенности.

8.1.10 Согласно GUM для построения доверительного интервала для θ используют величину

$$W(y,\theta)=\frac{y-\theta}{u(y)}. \quad (9)$$

Распределение $W(Y, q)$ приближенно можно считать t-распределением с ν_{eff} степенями свободы.

$$W(Y,\theta)\in t(\nu_{eff}). \quad (10)$$

Тогда доверительный интервал уровня 100 $(1-\alpha)$ % имеет вид:

$$y\pm u(y)t_{\nu_{eff},1-\alpha/2}. \quad (11)$$

Этот доверительный интервал может быть рекомендован в качестве доверительного интервала для θ с уровнем доверия 100 $(1-\alpha)$ %. Полуширина этого интервала $t_{\nu_{eff},1-\alpha/2}u(y)$ представляет собой расширенную неопределенность y.

8.1.11 Изложенное согласуется с обычной статистической практикой, когда общую неопределенность определяют с помощью оценки типа А, а наиболее часто используемой статистической оценкой для конкретной входной величины μ является выборочное среднее по n наблюдаемым величинам. Традиционным методом получения оценки стандартной неопределенности типа А является дисперсия $(S/\sqrt{n})$ с $(n-1)$ степенями свободы. Это основано на том, что $(n-1)S^2/\sigma^2$ подчиняется χ^2-распределению с $(n-1)$ степенями свободы. Этот метод применим также к статистикам вида $Y = G(X_1, ..., X_p)$, когда оценки X_i, $i = 1, ..., p$ подчиняются центральной предельной теореме. При этом стандартное отклонение Y может быть получено по формуле (7), в которой $u(x_i, x_j)$ необходимо заменить на $\text{cov}(X_i, X_j)$.

Метод GUM обобщает коллективный опыт многих метрологов. Однако он ограничен предположениями о:

- локальной линейности функции f (коэффициенты чувствительности не следует сильно менять и удалять);
- нормальности распределения точечных оценок $Y = f(X_1, ..., X_p)$ (для малых выборок это может не выполняться даже в приближении);
- законности формулы Велча-Саттервейта (возможно, формула плохо работает, если входные величины взаимно зависимы, входные данные не подчиняются нормальному распределению и стандартные неопределенности различны. Число степеней свободы для распределений, не связанных с χ^2-распределением, трудно интерпретировать, и в действительности его не используют в статистической теории).

8.1.12 Для анализа выражения (7) с точки зрения частотного подхода можно использовать понятия теории статистических решений и дисперсию (квадрат стандартной неопределенности) $u^2(y)$ интерпретировать как средний квадрат ошибки статистической оценки $f(x_1, x_2, ..., x_p)$. Это можно сделать при условии, что величины, неопределенности которых определены с использованием оценок типа В, а именно, $x_{m+1}, ..., x_p$ устранены интегрированием по их распределениям (см. [5]). Если функция f достаточно близка к линейной, выражение (7) обеспечивает приближение первого порядка среднего квадрата ошибки.

8.1.13 В примерах дана другая общепринятая для частотного похода процедура получения доверительных интервалов.

8.2 Бутстреп-метод

8.2.1 Бутстреп-метод представляет собой стратегию генерации повторных выборок [16] для оценки параметров распределения, таких как дисперсия, и определения доверительных интервалов для них в случае, когда форма основного распределения неизвестна. Ключевая идея бутстреп-метода состоит в том, что соотношение между интегральной функцией распределения вероятностей F для Y и выборкой из F аналогично соотношению между оценкой интегральной функции распределения $\hat{F}$, которая может быть не эмпирическим распределением, сгенерированным выборкой, и второй выборкой из $\hat{F}$. Если распределение F неизвестно, то никакие выводы относительно него не могут быть сделаны, но современные компьютеры позволяют получить оценку распределения $\hat{F}$ по большому количеству неизвестных распределений. Так, сначала используют первичную выборку для формирования $\hat{F}$ по F, а затем определяют выборочное распределение оценок параметров $\hat{F}$. Эта оценка основана на генера-

ции вторичных выборок и получении оценок для каждой вторичной выборки. Если $\hat{F}$ является хорошим приближением F, то H, выборочное распределение оценки на основе $\hat{F}$, является в общем случае хорошим приближением выборочного распределения оценки на основе F. Распределение H обычно называют бутстреп-распределением параметра.

8.2.2 Существует два типа бутстреп-процедур для непараметрического и параметрического случая. Непараметрический бутстреп-метод связан с рассмотрением эмпирического распределения $\hat{F}$, построенного на основе первичной выборки из F. В параметрическом случае бутстреп-распределение F имеет некий набор параметров, и $\hat{F}$ получают с помощью оценок параметров по данным.

Примечание — В типовых метрологических задачах наборы данных не являются достаточно большими, чтобы гарантировать правильность непараметрического бутстреп-подхода, поэтому этот подход не рассмотрен в настоящих рекомендациях.

8.2.3 Ключевое предположение, используемое при построении доверительного интервала в соответствии с GUM (предположение (10)), может не давать хорошей аппроксимации даже для простых задач. Однако бутстреп-метод позволяет получить доверительные интервалы без использования предположения (10). Одним из способов получения таких интервалов является t-бутстреп-подход. Эта процедура генерирует эмпирическое распределение для аппроксимации функции $W(Y, \theta)$ [для замены t-распределения в (10)]. Если предположение (10) корректно, t-бутстреп распределение воспроизводит t-распределение. Эмпирическое t-бутстреп распределение в этом случае используют для построения доверительного интервала по формуле (11) так же, как t-распределение.

Соотношение между бутстреп-методом и методами, предложенными в ГОСТ Р 54500.3.1—2011 (далее GUMS1), рассмотрено в 11.2.

8.2.4 Ниже приведена схема генерации бутстреп-выборки. Предположим, что x_1 и $u(x_1)$ — средние и стандартные отклонения случайной переменной X_1, которая, как предполагается, подчиняется распределению вероятностей с некоторым установленным количеством параметров. Далее для иллюстрации использовано нормальное распределение. Алгоритм бутстреп-выборки состоит из трех этапов:

а) По случайной выборке объема k из нормального распределения определяют оценки среднего и стандартного отклонения x_1 и $u(x_1)$.

б) Из $N(x_1, u^2(x_1))$ отбирают выборку объема k, т. е. $\left\{x^*_{1,1}, ..., x^*_{1,k}\right\}$.

в) По $\left\{x^*_{1,1}, ..., x^*_{1,k}\right\}$ вычисляют выборочное среднее x^*_1 и выборочную стандартную ошибку $u(x^*_1)$.

$\left\{x^*_1, u(x^*_1)\right\}$ — бутстреп-выборка для X_1. Точно так же B бутстреп-выборок может быть сгенерировано для любой переменной.

8.2.5 Так же, как в GUM значения $(x_i, u(x_i))$, для $i = 1, ..., p$ используют как входные параметры при вычислении y, $u(y)$ и $W(Y, \theta)$, а бутстреп-выборки $\left\{x^*_1, u(x^*_1)\right\}$ $i = 1,...,p$ (см. 8.2.4) используют как входные параметры для вычисления y^*, $u(y^*)$,

$$W^* = W(y^*, y) = \frac{y^* - y}{u(y^*)}. \tag{12}$$

8.2.6 Для получения бутстреп-распределения $W(Y, \theta)$ для больших значений B (скажем 100000) генерируют B бутстреп-выборок $\left\{x^*_i(b), u(x^*_i(b))\right\}$, $i = 1, .., p$ и для каждой вычисляют $W^*(b)$, $b = 1, ..., B$. Процентиль t-бутстреп распределения $W(Y, \theta)$ уровня 100 α % аппроксимируют значением $\hat{t}_\alpha$ так, что

$$\frac{\left|\left\{W^*(b) \le \hat{t}_\alpha\right\}\right|}{B} = \alpha,$$

где $|A|$ — количество элементов, удовлетворяющих условию A. Наконец, доверительный интервал t-бутстреп распределения уровня 100 $(1 - \alpha)$ % имеет вид

$$\left(y - \hat{t}_{1-\alpha/2} \cdot u(y)\, y + \hat{t}_{\alpha/2} \cdot u(y)\right). \tag{13}$$

Процентили распределения Стьюдента симметричны относительно нуля, и как следствие, значения в формуле (11) всегда должны быть симметричны относительно y. Напротив, процентили t-бутстреп распределения, используемые в (13), могут быть асимметричны относительно нуля, что приводит к несимметричному интервалу неопределенности относительно y, который может обеспечить более точное описание физической ситуации в некоторых случаях. Детали этого процесса при построении интервала неопределенности приведены в следующем алгоритме.

а) Используя заданные распределения для X_i, $i = 1, .., p$ формируют B бутстреп-выборок

$$(x_i^*(1), u(x_i^*(1))), \quad ..., \quad (x_i^*(B), u(x_i^*(B))).$$

б) Для каждой бутстреп-выборки $\left(x_i^*(b), u(x_i^*(b))\right)$, $i = 1, .., p$ и $b = 1, ..., B$, вычисляют $y^*(b)$, $u(y^*(b))$ и $W^*(b) = (y^*(b) - y)/u(y^*(b))$ в соответствии с GUM.

в) Определяют процентиль t-бутстреп распределения $W(Y, \theta)$ уровня 100 α % с помощью значения $\hat{t}_\alpha$ такого, что $|\{W^*(b) \le \hat{t}_\alpha\}|/B = \alpha$.

г) Определяют доверительный интервал t-бутстреп распределения уровня 95 % $(y - \hat{t}_{0,975} \cdot u(y); \quad y + \hat{t}_{0,025} \cdot u(y))$.

8.2.7 Бутстреп-выборки могут также использоваться для замены $u(y)$ оценкой стандартного отклонения Y, когда приближение Тэйлора (6) считают неподходящим. Для этого, поскольку $i = 1, .., p$, и $b = 1,..., B$, генерируют только входные оценки $x_i^*(b)$. Для каждой бутстреп-выборки определяют $y^*(b) = f(x_1^*(b), ..., x_p^*(b))$. Бутстреп-оценкой стандартной неопределенности y является стандартное отклонение выборки объема B:

$$u_c(y) = \sqrt{\sum_{b=1}^{B} \left[y^*(b) - y^*(\cdot)\right]^2 / (B-1)}, \quad y^*(\cdot) = \sum_{b=1}^{B} y^*(b) / B.$$

8.2.8 Наконец, если приближение Тэйлора является неподходящим и имеется существенная асимметрия основного распределения Y, отбирают вложенные бутстреп-выборки $B_1 \times B_2$ для построения t-бутстреп интервала с использованием бутстреп стандартного отклонения. Пусть генерировано B_1 бутстреп-выборок (входных оценок и соответствующих им y^*). Для каждой бутстреп-выборки $u_c(y^*)$ вычисляют B_2 бутстреп-выборок второго уровня и определяют оценки

$$\frac{y^* - y}{u(y^*)}.$$

Набор B_1 таких отношений используют для определения оценки процентили распределения $W(Y, \theta)$, что позволяет определить t-бутстреп интервал в соответствии с (13). Алгоритм построения интервала неопределенности уровня 95 % с использованием вложенных бутстреп-выборок состоит в следующем:

а) Для $i = 1,..,p$, используя распределения X_i, генерируют B_1 бутстреп-выборок первого уровня $(x_i^*(1), ..., (x_i^*(B_1)))$.

б) Для каждой бутстреп-выборки первого уровня $x_i^*(B_1)$, $i = 1,.., p$, $b_1 = 1,..., B_1$ вычисляют $y^*(b_1) = f(x_1^*(b_1), ..., x_p^*(b_1))$ и $W^*(b_1) = (y^*(b_1) - y)/u(y^*(b_1))$, где для определения $u(y^*(b_1))$ применяют бутстреп-метод второго уровня, используя следующий алгоритм:

1. Для $i = 1, .., p$, используя распределение μ_i, генерируют B_2 бутстреп-выборок второго уровня $x_i^*(1),..., x_i^*(B_2)$.

2. Для каждой бутстреп-выборки второго уровня определяют оценку $y^*(b_2) = f(x_1^*(b_2), ..., x_p^*(b_2))$.

3. Формируют бутстреп-оценку стандартной неопределенности $y^*(b_1)$ в виде выборочного стандартного отклонения:

$$u(y^*(b_1)) = \sqrt{\sum_{b_2=1}^{B_2} \left[y^*(b_2) - y^*(\cdot)\right]^2 / (B_2 - 1)},$$

для B_2 повторений, где $y^*(\cdot) = \sum_{b_2=1}^{B_2} y^*(b_2) / B_2$.

в) Определяют оценку процентили уровня 100 α % *t*-бутстреп распределения $W(Y, \theta)$ с помощью значения такого, что $|\{W^*(b_1) \le \hat{t}_\alpha\}|/B_1 = \alpha$.

г) Определяют вложенный *t*-бутстреп доверительный интервал уровня 95 %

$$\left(y - \hat{t}_{0,975} \cdot u(y);\quad y + \hat{t}_{0,025} \cdot u(y)\right).$$

Это более общий подход. Вложенный бутстреп-метод требует достаточно сложных вычислений. Для примеров был выбран более простой бутстреп-метод.

8.3 Примеры

8.3.1 Общие положения

8.3.1.1 В качестве примера рассмотрим статистическую модель примера из раздела 7, а именно

$$Y_i = \theta + \beta + \varepsilon_i,\ i = 1, \ldots, n, \tag{14}$$

где θ — измеряемая величина, β представляет фоновый шум, ε_i — независимые ошибки, $\varepsilon_i \in N(0, \sigma^2)$. Если β — постоянная величина, γ — среднее данных, то в соответствии с (14) $\theta = f(\beta, \gamma) = \gamma - \beta$.

8.3.1.2 Если фоновый шум β подчиняется равномерному распределению на интервале $(a - d, a + d)$, интервал для θ в соответствии с GUM имеет вид:

$$\bar{Y} - a \pm 2\sqrt{\frac{\sigma^2}{n} + \frac{d^2}{3}}.$$

В [5] рассмотрены свойства таких интервалов и проведено их сравнение с интервалом

$$\bar{Y} - a \pm \left[2\sqrt{\frac{\sigma^2}{n}} + d\right], \tag{15}$$

рекомендованным Эйзенхартом в [17]. Так как условное распределение $\bar{Y}$ для данного β является нормальным, $N(\theta + \beta, \sigma^2/n)$,

$$P\left(\left|\bar{Y} - \theta - \beta\right| \le \frac{2\sigma}{\sqrt{n}}\right) \ge 0{,}95,$$

в то время как

$$P(|\alpha - \beta| \le d) = 1.$$

Из этого следует, что интервал Эйзенхарта является гарантированным, т. е.

$$P\left(\left|\bar{Y} - \alpha - \theta\right| \le \frac{2\sigma}{\sqrt{n}} + d\right) \ge 0{,}95. \tag{16}$$

8.3.1.3 Однако если $d > 12\sigma/\sqrt{n}$, интервал, рекомендованный GUM, содержит интервал (15), из чего видны различия этих подходов.

8.3.1.4 Интервал (15) может быть приспособлен к *t*-распределению отношения $\sqrt{n}(\bar{Y} - \alpha - \beta)/S$. Он также может быть использован для других распределений (треугольного, трапециевидного и т. д.). Действительно, в модели (14) статистика $\bar{Y}$ содержит всю информацию о θ, содержащуюся в данных (т. е. $\bar{Y}$ — достаточная статистика для θ) с плотностью вероятностей

$$\frac{\sqrt{n}}{2\sqrt{2\pi}\sigma d}\int_{a-d}^{a+d} e^{-0{,}5n(\bar{y}-\theta-\beta)^2/\sigma^2} d\beta.$$

Специальная форма этого распределения позволяет получить альтернативные доверительные интервалы (различной длины с центром в точке максимума правдоподобия ($\bar{Y} - \alpha$)) (см. [14]).

8.3.2 Пример а)

Используем исходные данные примера, приведенного в разделе 7 для модели (14). После вычислений получаем $\bar{y}$ = 3,537 и $u(\bar{y})$ = 0,153. Последним заменяем $\sigma/\sqrt{n}$ в неравенстве (16), а ко-

эффициент 2 следует заменить на процентиль t-распределения с эффективным числом степеней свободы 5,15. В примере а) фоновый шум β может быть исследован на основе измеренных величин, принадлежащих нормальному распределению. Получаем $\bar{b}$ = 1,228 и $u(\bar{b})$ = 0,059. Тогда оценка θ принимает значение $\bar{y} - \bar{b}$ = 2,309 с соответствующей стандартной неопределенностью $\sqrt{u^2(\bar{y}) + u^2(\bar{b})}$ = = 0,164. Доверительный интервал в соответствии с GUM имеет вид

$$2{,}309 \pm 2{,}548 \times 0{,}164 = 2{,}309 \pm 0{,}417 = (1{,}892;\ 2{,}727).$$

t-бутстреп интервал уровня 100(1 – α) % согласно (13) имеет вид:

$$(2{,}309 - 0{,}164\,\hat{t}_{1-\alpha/2}\,;\ 2{,}309 + 0{,}164\,\hat{t}_{\alpha/2}),$$

где $\hat{t}_{\beta}$ — процентиль W^* уровня 100 β % (12).

Для удобства пользователей в качестве иллюстрации далее приведены некоторые фрагменты R-программы (см. [18]) и WinBUGS (см. [19]). Для примера а) приведена R-программа генерации B = 10000 реализаций распределения W^*:

```
B = 10000
y.star = rnorm (B, mean=3,537, sd=0,153)
u.y.star = 0,153 * sqrt (rchisq (B, df=4)/4)
b.star = rnorm (B, mean=1,228, sd=0,059)
u.b.star = 0,059 * sqrt (rchisq (B, df=4)/4)
w.star = ((y.star-b.star)-2,309)/sqrt (u.y.star^2+u.b.star^2).
```

Доверительный t-бутстреп интервал уровня 95 %, на основе квантилей уровней 0,025 и 0,975 моделируемого распределения:

```
2.309 — quantile (w.star, c (0,975; 0,025)) *0,164
## 1,895754; 2,728817
```

Таким образом, доверительный t-бутстреп интервал уровня 95 % имеет вид (1,896; 2,729).

8.3.3 Пример б)

При отсутствии статистических данных о фоновом шуме предполагается, что β подчиняется равномерному распределению в интервале (1,126; 1,329). Тогда приближенный доверительный интервал в соответствии с GUM имеет вид:

$$3{,}537 - 1{,}228 \pm 2{,}533\sqrt{\frac{0{,}342^2}{5} + \frac{0{,}102^2}{3}} = 2{,}310 \pm 0{,}415 = (1{,}895;\ 2{,}724).$$

Доверительный интервал Эйзенхарта несколько шире, а именно,

$$3{,}537 - 1{,}228\ \left[2{,}776\frac{0{,}342}{\sqrt{5}} + 0{,}102\right] = 2{,}310 \pm 0{,}526 = (1{,}783;\ 2{,}836).$$

На основе данных примера а) для θ может быть построен доверительный t-бутстреп интервал. Для данного примера оценки и соответствующая стандартная неопределенность для y, β и θ представлены в цифровой форме так же, как в примере а), за исключением того, что β определен на основании опыта или оценок экспертов и оценка неопределенности B представляет собой оценку типа В. Поэтому генерация W^* выполнена не так, как в примере а), а только для отбора бутстреп-выборки b^* и определения оценки соответствующей неопределенности. Бутстреп-выборка объема b^* теперь генерирована из известного равномерного распределения (1,126; 1,329) со стандартной неопределенностью 0,059. Программа генерации B = 10000 реализаций W^* приведена ниже.

```
B = 10000
y.star = rnorm (B, mean=3,537, sd=0,153)
u.y.star = 0,153 * sqrt (rchisq (B, df=4)/4)
b.star = runif (B, min=1,126, max=1,329)
u.b.star = 0,059
w.star = ((y.star-b.star)-2,309)/sqrt (u.y.star^2+u.b.star^2).
```

Доверительный t-бутстреп интервал уровня 95 %, основанный на квантилях уровней 0,025 и 0,975

```
2,309 — quantile (w.star, c (0,975; 0,025)) *0,164
## 1,918643; 2,699749.
```

Таким образом, доверительный t-бутстреп интервал уровня 95 % имеет вид (1,919; 2,700).

8.3.4 Пример в)

Так как $\bar{y}$ = 1,196 и s_y = 0,047, оба интервала имеют отрицательные нижние границы. Если среднее θ положительно, то нижнюю границу можно заменить нулем, что приводит к рекомендуемому GUM интервалу (0; 0,124) и к интервалу Эйзенхарта (0; 0,202).

R-программа генерации B = 10000 реализаций W^* и получения бутстреп-интервала такая же, как для примера б) с $\bar{y}$ = 1,196 и $u(\bar{y})$ = 0,047.

```
B = 10000
y.star = rnorm (B, mean=1,196, sd=0,047)
u.y.star = 0,047 * sqrt (rchisq (B, df=4)/4)
b.star = runif (B, min=1,126, max=1,329)
u.b.star = 0,059
w.star = (y.star-b.star) +0,032)/sqrt (u.y.star^2+u.b.star^2).
Доверительный t-бустреп интервал уровня 95 %
-0,032 -quantile (w.star, c (0,975;0,025)) *0,075
##-0,1762648; 0,1128422
```

Таким образом, доверительный t-бустреп интервал уровня 95 % имеет вид (-0,176; 0,113).

Если известно, что θ положительная величина, то усеченный доверительный t-бустреп интервал уровня 95 % для θ имеет вид (0; 0,113).

9 Байесовский подход

9.1 Основной метод

9.1.1 В метрологии измеряемая величина и входные переменные модели (1) являются постоянными физическими величинами. При байесовском подходе соответствующие параметры μ_i и θ рассматривают как случайные величины в том смысле, что их распределения вероятностей суммируют знания об этих величинах.

9.1.2 Байесовский подход позволяет получить распределения вероятностей без физических данных, например, используя требования изготовителей или экспертные оценки. На практике в типичных метрологических задачах результаты измерений (данные) физических величин могут быть использованы для оценки параметров входных величин. В таких случаях, используя теорему Байеса, можно получить плотность распределения следующим образом. Пусть $p(\mu_i)$ — плотность распределения μ_i, известная заранее до получения физических данных. Эта функция называется априорной плотностью вероятностей μ_i. Пусть Y — случайная величина, для которой существует ее реализация y (данные). Плотность $p(y|\mu_i)$ для Y называют статистической моделью. Так как μ_i — случайная величина, знак «|» обозначает, что плотность вероятностей Y является условной по μ_i. Для конкретной реализации y величины Y плотность $p(y|\mu_i)$, как функция μ_i, называется функцией правдоподобия. В соответствии с теоремой Байеса

$$p(\mu_i|y) = \frac{p(y|\mu_i)p(\mu_i)}{\int p(y|\mu_i)p(\mu_i)d\mu_i}. \qquad (17)$$

Плотность распределения μ_i обобщает знания о μ_i с учетом полученных значений y_i.

9.1.3 Если заранее о μ_i ничего не известно, то используют так называемое неинформативное априорное распределение [20]. При наличии априорной информации ее представляют в виде информативного распределения вероятностей. Это один из механизмов байесовского подхода для включения информации, используемой для получения оценки неопределенности типа В. Форму функции правдоподобия обычно выбирают на основе знаний о процессе, поставляющем данные.

9.1.4 Вид функции правдоподобия и априорная плотность определяют форму апостериорной плотности. Важно тщательно выбрать функцию правдоподобия и априорную плотность и провести анализ чувствительности результатов относительно возможных изменений этих распределений. Для априорных распределений это может означать сопоставление результатов использования нескольких различных плотностей. Проверка соответствия функции правдоподобия (статистическая модель, которая описывает данные об измерениях) представляет собой форму валидации модели [21], которую проводят не только для байесовской, но и для других моделей (частотной и фидуциальной).

9.1.5 Определение оценки неопределенности измерений в контексте байесовской статистики может быть интерпретировано (по отношению к апостериорному распределению вероятностей для измеряемой величины θ) следующим образом: стандартная неопределенность — это стандартное отклонение случайной величины, характеризуемой этим распределением вероятностей. Для определения стандартного отклонения необходимо сначала найти объединенное распределение вероятностей μ_i и затем применить формулу замены переменных, предложенную в [14] для получения распределения θ.

Простой способ получения моментов этого распределения состоит в следующем. Математическое ожидание и дисперсия функции $h(\theta)$ имеют вид:

$$E(h(\theta)) = \int \ldots \int h(f(\mu_1, \ldots, \mu_p)) p(\mu_1, \ldots, \mu_p) d\mu_1 \ldots d\mu_p,$$

$$Var(\theta) = E(\theta^2) - [E(\theta)]^2.$$

Для интегрирования часто используют методы Монте-Карло [20].

9.1.6 Совместное распределение вероятностей для независимых случайных величин представляет собой произведение частных распределений. Однако μ_i не всегда являются независимыми. Для такого случая возьмем, например, функцию распределения вероятностей Y, зависящую только от двух переменных μ_1 и μ_2. Здесь $p(y|\mu_1, \mu_2)$ — статистическая модель $p(\mu_1, \mu_2) \neq p(\mu_1)\, p(\mu_2)$. Для этого примера апостериорная плотность (μ_1, μ_2) имеет вид:

$$p(\mu_1, \mu_2 | y) = \frac{p(y | \mu_1, \mu_2) p(\mu_1, \mu_2)}{\int p(y | \mu_1, \mu_2) p(\mu_1, \mu_2) d\mu_1 d\mu_2}.$$

9.1.7 Распространенная ситуация, которая приводит к такой зависимости, появляется тогда, когда статистическая модель является функцией θ и μ_i. Оба примера, рассмотренных в данном разделе, попадают в эту категорию, иллюстрируя тезис о том, что при байесовском подходе, когда результаты измерений доступны, процесс определения связанных распределений вероятностей требует соответствующего определения статистической модели. Выполнение этого правила автоматически приводит к функциям правдоподобия, необходимым для применения теоремы Байеса и к соответствующей апостериорной плотности. Таким образом, процесс можно описать следующим образом:

а) определение результатов измерений, относящихся к исследуемым физическим величинам;

б) установление статистической модели (также называемой моделью наблюдений), связывающей данные с параметрами, которыми могут быть μ_i и измеряемая величина θ;

в) установление априорных распределений для всех используемых параметров;

г) применение теоремы Байеса для получения апостериорных распределений параметров;

д) вычисление апостериорного среднего и апостериорного стандартного отклонения измеряемой величины;

е) выполнение анализа чувствительности результатов относительно возможных изменений апостериорных распределений.

9.1.8 Чтобы упростить числовые вычисления, там, где это возможно, можно использовать последовательное приближение Тэйлора и предположение о нормальности распределений. В частности, разложение в ряд Тэйлора $f(\mu_1, \ldots, \mu_p)$ в точке математических ожиданий μ_i вместе с предположением о нормальности могут быть использованы для утверждения о том, что $f(\mu_1, \ldots, \mu_p)$ имеет приближенное распределение $N(f(E(\mu_1), \ldots, E(\mu_p)), \omega^2)$, где

$$\omega^2 = \sum_{i=1}^{p} c_i^2 Var(\mu_i) + 2 \sum_{i<j} c_i c_j Cov(\mu_i, \mu_j).$$

$Cov(\mu_i, \mu_j)$ — ковариация μ_i и μ_j, c_i — частная производная θ по переменной μ_i, вычисленная в точке математического ожидания μ_i.

Примечание — Формулы (6) и (7) использованы в 8.1.8. В данном случае разложение использовано для определения дисперсии оценки θ, а не самого θ.

9.2 Пример

9.2.1 Общие положения

Проиллюстрируем процесс на данных примера раздела 7. Измеряемой величиной в этом примере является θ. Модель измерений в соответствии с 8.3.1.1 имеет вид:

$$\theta = \gamma - \beta. \tag{18}$$

9.2.2 Пример а)

9.2.2.1 Имеется два набора данных: Набор данных I — пять результатов измерений y_i (сигнал плюс фоновый шум), полученных независимо друг от друга. Набор данных II — пять результатов измерений значений фонового шума b_i, также независимых. Каждую величину в наборе данных I рассматривают как реализацию случайной величины Y_i, подчиняющуюся нормальному распределению со средним $\gamma = \theta + \beta$ и стандартным отклонением σ_Y. Те же предположения сделаны относительно случайной величины B_i в наборе данных II, со средним β и стандартным отклонением σ_B. Таким образом, статистические модели для Y_i и пяти результатов измерений независимых величин имеют вид:

$$Y_i \mid \theta, \beta, \sigma_Y^2 \in N(\theta + \beta, \sigma_Y^2),$$

$$p(y_1, \ldots, y_5 \mid \theta, \beta, \sigma_Y) = \left(\frac{1}{\sigma_Y\sqrt{2\pi}}\right)^5 \exp\left\{-\frac{\sum_{i=1}^{5}(y_i - \theta - \beta)^2}{2\sigma_Y^2}\right\}.$$

9.2.2.2 Статистическая модель для B_i

$$B_i \mid \beta, \sigma_B^2 \in N(\beta, \sigma_B^2),$$

$$p(b_1, \ldots, b_5 \mid \beta, \sigma_B) = \left(\frac{1}{\sigma_B\sqrt{2\pi}}\right)^5 \exp\left\{-\frac{\sum_{i=1}^{5}(b_i - \beta)^2}{2\sigma_B^2}\right\}.$$

9.2.2.3 Так как два набора наблюдений взаимно независимы, статистическая модель для Y и B имеет вид:

$$p(y, b \mid \theta, \beta, \sigma_Y, \sigma_B) = p(b_1, \ldots, b_5 \mid \beta, \sigma_B)\, p(y_1, \ldots, y_5 \mid \theta, \beta, \sigma_Y).$$

9.2.2.4 Таким образом, существуют четыре параметра θ, β, σ_Y и σ_B, для которых необходимо задать априорные распределения. В данном примере нет никакой информации об этих параметрах, кроме того, что они являются неотрицательными, поэтому предположим, что эти параметры независимы. Желательно, чтобы влияние вида априорных распределений на результаты анализа было минимальным. Такой результат может быть получен при использовании так называемой опорной априорной плотности [20]. Для параметров, связанных со средними, например θ и β, такая плотность может быть аппроксимирована для θ равномерным распределением (0, *c*), для β ~ равномерным распределением (0, c) с большими значениями *c*. Для параметров масштаба σ_Y и σ_B опорные априорные плотности вида:

$$p(\sigma_Y) = 1/\sigma_Y,\ p(\sigma_B) = 1/\sigma_B$$

являются неподходящими, так как их трудно объединить в единую плотность. Так как этот аспект может вызвать трудности при вычислениях, для больших значений использованы более подходящие плотности: для σ_Y — равномерное распределение (0, *c*) или Гамма-распределение (*c*, *c*).

Гамма (ϕ_1, ϕ_2) представляет собой гамма-распределение с параметрами ϕ_1 и ϕ_2, т. е. для случайной переменной X эта плотность вероятностей имеет вид:

$$p(x \mid \phi_1, \phi_2) = \frac{\phi_2^{\phi_1}}{\Gamma(\phi_1)} x^{\phi_1 - 1} e^{-x\phi_2}.$$

9.2.2.5 Применение теоремы Байеса приводит к объединенной апостериорной плотности для θ, β, σ_Y и σ_B следующего вида:

$$p(\theta, \beta, \sigma_Y, \sigma_B \mid y, b) = \frac{p(y, b \mid \theta, \beta, \sigma_Y, \sigma_B) p(\theta) p(\beta) p(\sigma_Y) p(\sigma_B)}{\int p(y, b \mid \theta, \beta, \sigma_Y, \sigma_B) p(\theta) p(\beta) p(\sigma_Y) p(\sigma_B) d\theta d\beta d\sigma_Y d\sigma_B}.$$

Апостериорная плотность измеряемой величины θ выражается через интеграл и имеет вид:

$$p(\theta \mid y, b) = \int p(\theta, \beta, \sigma_Y, \sigma_B \mid y, b) d\beta d\sigma_Y d\sigma_B.$$

Это апостериорное распределение содержит всю информацию о θ после выполнения измерений. Оценка математического ожидания этого распределения является оценкой физической величины, а стандартное отклонение этого распределения — оценкой стандартной неопределенности этой оценки. Это прямой путь получения интервала охвата для измеряемой величины на основе такого распределения. Такой интервал охвата представляет собой интервал возможных значений θ с заданной вероятностью. В статистике Байеса этот интервал называют байесовским доверительным интервалом. Во многих случаях при применении теоремы Байеса для вычисления интеграла используют численные методы. Одним из возможных решений является получение апостериорного распределения методами Монте-Карло по схеме марковской цепи (см. [22]) с использованием программного обеспечения WinBUGS (см. [19]). Программа для этого примера, с равномерными априорными распределениями и $c = 100$ приведена ниже

```
Example 1a {
theta~dunif (0,100)
beta~dunif (0,100)
gamma<- theta+beta
sigma.Y~dunif (0,1)
sigma.B~dunif (0,1)
tau.Y <- 1 / (sigma.Y*sigma.Y)
tau.B <- 1 / (sigma.B*sigma.B)
for (i in 1:n) {
y[i]~dnorm (gamma, tau.Y)
b[i]~dnorm (beta, tau.B)}
}
```

С данными из 7.2 для $n = 5$ программа дает апостериорное среднее θ, равное 2,309, и апостериорное стандартное отклонение 0,247. Байесовский доверительный интервал для θ уровня 95 % имеет вид (1,805; 2,815). Анализ чувствительности относительно изменений формы четырех предшествующих распределений может быть выполнен с помощью изменения значений c (см. 9.2.2.4) и подстановки строк

```
tau.Y~dgamma (1,0E-5; 1,0E-5)
tau.B~dgamma (1,0E-5; 1,0E-5)
```

в четыре строки

```
sigma.Y~dunif (0,1)
sigma.B~dunif (0,1)
tau.Y <- 1/(sigma.Y*sigma.Y)
tau.B <- 1/(sigma.B*sigma.B)
```

и сравнения полученных значений апостериорного среднего и стандартного отклонения. По результатам очевидна их устойчивость к таким изменениям.

9.2.3 Пример б)

9.2.3.1 Информация о β представлена в форме распределения вероятностей, полученного на основе оценки неопределенности типа В. В этом случае модель наблюдений для набора данных (I) в примере а) (9.2.2), имеет вид:

$$Y_i \mid \theta, \beta, \sigma_Y^2 \in N(\theta + \beta, \sigma_Y^2).$$

9.2.3.2 Имеется три параметра, которым необходимо назначить априорные распределения. Для параметра β в качестве априорной плотности выбрано равномерное распределение на интервале

(1,126; 1,329), для θ — равномерное распределение (0, *c*) и для σ_Y — равномерное распределение (0, *c*) с большими значениями *c*.

9.2.3.3 Программа WinBUGS для данного примера:

```
Example 1b {
theta~dunif (0,100)
beta~dunif (1,126;1,329)
sigma.Y~dunif (0,1)
gamma< -theta+beta
tau.Y < -1/(sigma.Y*sigma.Y) для (i in 1:n) {
y[i]~dnorm (gamma, tau.Y)
}
```

Эта программа дает апостериорное среднее θ, равное 2,309 и апостериорное стандартное отклонение θ, равное 0,232. Доверительный байесовский интервал уровня 95 % для θ имеет вид (1,832; 2,788). Анализ чувствительности также удовлетворительный.

9.2.4 Пример в)

9.2.4.1 Единственным отличием данного примера от примера б) является различие фактических измеряемых величин (которые теперь близки к фоновому шуму), поэтому могут быть использованы та же модель и та же программа WinBUGS, что и в примере б). Апостериорное среднее θ теперь составляет 0,069, апостериорное стандартное отклонение θ равно 0,067, а байесовский доверительный интервал уровня 95 % для θ имеет вид (0,000; 0,188). Эти результаты устойчивы к изменению значений *c* в априорном равномерном распределении. Замена формы априорной плотности для σ_Y равномерного распределения на Гамма-распределение приводит к значениям апостериорного среднего и апостериорного стандартного отклонения 0,058 и 0,052, соответственно, и байесовскому доверительному интервалу уровня 95 % (0,000; 0,150). Это большее изменение, чем в предыдущих примерах. Видно, что из-за близости данных к значению фонового шума они дают меньше информации об измеряемой величине. Значение σ_Y (управляемое до некоторой степени с помощью априорного распределения, так как есть только пять значений, на которых основана эта оценка) влияет на информативность данных. В этом случае консервативное (гарантированное) решение состоит в использовании равномерного распределения, обеспечивающего более широкий байесовский доверительный интервал. Очевидно, что использование большего количества данных снижает влияние априорной плотности σ_Y на результаты. Дополнительные сведения о байесовских интервалах приведены в [23]. Там показано, что для моделей, рассмотренных в примере, байесовским доверительным интервалам, построенным на основе априорного равномерного распределения, соответствует вероятность охвата, близкая к 95 %, в то время как интервалам, основанным на априорном гамма-распределении, соответствует меньшая вероятность.

9.2.5 Выводы

В примере а) показана ситуация, когда для оценки неопределенности измерений использованы две независимые измеряемые величины. В примере б) показано, как информация о фоновом шуме, используемая для определения оценки неопределенности типа В, может быть включена в байесовскую модель. В примере в) показано, что ограничения (например, положительность) значений измеряемой величины могут быть включены в байесовскую модель. В этом примере также показано, как выбор неинформативного априорного распределения влияет на результаты.

10 Фидуциальный подход

10.1 Основной метод

10.1.1 Для функции измерений (1) оценка неопределенности измеряемой величины θ может быть основана на фидуциальном распределении θ. Приведенные ниже примеры иллюстрируют способ получения фидуциальных распределений для исследуемых параметров.

10.1.2 Пусть $Y \in N(\theta, 1)$, где θ — измеряемая величина, процесс измерений имеет известную дисперсию, равную 1, и *Y* — наблюдаемые значения случайной величины. Соотношение между результатами и случайной экспериментальной погрешностью процесса имеет вид:

$$Y = \theta + E, \tag{19}$$

где E — случайная величина, характеризующая ошибку измерений, подчиняющаяся нормальному распределению *N*(0,1). Каждому результату измерений соответствует случайная ошибка измерений. Пред-

положим, что выполнено единственное измерение, и результат измерений составил 10. Обозначим соответствующую ошибку измерений e. Тогда:

$$10 = \theta + e.$$

Следовательно, $\theta = 10 - e$. Если бы значение e было известно, то результат измерений был бы известен точно. Поскольку распределение случайной величины e известно, можно определить совокупность наиболее вероятных значений θ. Например, насколько вероятно, что $\theta = 2$. Для этого, необходимо, чтобы $e = 8$. Значение $e = 8$ маловероятно для $N(0, 1)$. Таким образом, значение $\theta = 2$ маловероятно. Какова вероятность того, что θ принимает значения из интервала от 10 до 12. Для этого e должно находиться в интервале от 0 до 2, и вероятность этого события равна $\Phi(2) - \Phi(0)$, где $\Phi(z)$ — значение интегральной функции нормированного нормального распределения в точке z. Таким образом, по вероятностям, соответствующим ошибке E, могут быть определены вероятности для θ. Знание о θ, основанное на результате измерений 10, может быть описано распределением случайной величины $\tilde{\theta} = 10 - E$, распределение которой имеет вид $N(10,1)$. Это фидуциальное распределение θ (т. е. распределение $\tilde{\theta}$). Случайную переменную $\tilde{\theta}$ также называют фидуциальной величиной для θ. Эту величину называют в литературе обобщенной центральной величиной (см. [24], [25]) или фидуциальной обобщенной центральной величиной (см. [26], [27]).

10.1.3 Предположим, что в рассмотренном выше примере выполнено два измерения. Пусть Y_1 и Y_2 — случайные величины, обозначающие возможные значения результатов измерений:

$$\begin{aligned} Y_1 &= \theta + E_1 \\ Y_2 &= \theta + E_2. \end{aligned} \qquad (20)$$

Предположим, что фактические результаты измерений равны 10 и 8. Тогда, следующие уравнения связывают результаты измерений, измеряемую величину и реализованные значения e_1 и e_2

$$\begin{aligned} 10 &= \theta + e_1, \\ 8 &= \theta + e_2. \end{aligned}$$

Вероятные значения θ связаны с вероятными значениями (e_1, e_2). В отличие от предыдущего примера предположим известно, что $e_1 - e_2 = 2$. Теперь набор возможных значений (e_1, e_2) ограничен этим требованием. Известно, что (e_1, e_2) подчиняется двумерному нормированному нормальному распределению, но ограничены линейным соотношением $e_1 - e_2 = 2$.

Следовательно фидуциальная величина $\tilde{\theta}$ имеет распределение, которое равно условному распределению $(10 - E_1)$, при условии, что $(E_1 - E_2) = 2$. Это то же самое распределение, что и условное распределение $(8 - E_2)$, при условии, что $(E_1 - E_2) = 2$. Простые вычисления показывают, что распределение $\tilde{\theta}$ подчиняется $N(\bar{y}, 1/2)$, где $\bar{y} = (y_1 + y_2)/2 = (10 + 8)/2 = 9$.

10.1.4 Для n независимых результатов измерений из $N(\theta, \sigma^2)$,

$$\begin{aligned} Y_1 &= \theta + \sigma E_1, \\ Y_2 &= \theta + \sigma E_2, \\ &\cdots\cdots, \\ Y_n &= \theta + \sigma E_n, \end{aligned} \qquad (21)$$

где $E_1,..., E_n$ — независимые случайные величины из $N(0,1)$. Объединенное фидуциальное распределение для (θ, σ) может быть получено следующим образом. Используя первые два (или любые два) из выше приведенных n уравнений необходимо выразить θ и σ как функции y_1, y_2, E_1, и E_2, обозначая их $\tilde{\theta}$ и $\tilde{\sigma}$. Объединенное фидуциальное распределение (θ, σ) является объединенным распределением $(\tilde{\theta}, \tilde{\sigma})$ при условии E_i, наложенном остальными $(n - 2)$ уравнениями. В частности фидуциальное распределение θ

$$\tilde{\theta} = \bar{y} - \frac{s}{\sqrt{n}} T_{n-1} \qquad (22)$$

представляет собой смещенное и масштабированное t-распределение с $(n - 1)$ степенями свободы. Здесь $\bar{y}$ и s — значения выборочных среднего $\bar{X}$ и стандартного отклонения S по n результатам измерений, а T_{n-1} — случайная величина, подчиняющаяся t-распределению с $(n - 1)$ степенями свободы.

10.1.5 Далее приведен альтернативный и более простой метод формирования фидуциального распределения, чем тот, что описан в предыдущей части раздела формулой (22).

10.1.6 Приведенные выше аргументы можно обобщить. Фидуциальные распределения могут быть получены для параметров модели в более общих задачах. Отправной точкой этого процесса является структурное уравнение [28]. Структурное уравнение имеет вид $Y = G(\beta, E)$. Для единственного измерения уравнение (19) является структурным уравнением. Для n результатов измерений уравнения (21) представляют собой n структурных уравнений. Структурные уравнения связывают результаты измерений Y с параметрами модели β и ошибкой измерений E, распределение которой полностью известно. Например, для единственного результата измерений распределение E известно полностью. Для любых постоянных значений β распределение E и структурные уравнения $G(\cdot)$ определяют распределение Y. После выполнения измерений Y роли данных и параметров могут поменяться местами. В частности, если значение Y установлено, распределение E и структурные уравнения $G(\cdot)$ используют для определения распределения β. Это то, что составляет фидуциальный вывод.

10.2 Пример

10.2.1 Пример а)

10.2.1.1 Рассмотрим пример а) из раздела 7, в котором необходимо определить оценку величины θ на основе результатов измерений в соответствии с моделью

$$Y_i = \theta + \beta + \varepsilon_i,\ i = 1,..,\ n, \tag{23}$$

где ε_i — независимые ошибки измерений, $\varepsilon_i \in N(0, \sigma_Y^2)$. β — параметр фонового шума в соответствии с моделью

$$B_i = \beta + \delta_i,\ i = 1,..,\ n, \tag{24}$$

где δ_i — независимые ошибки измерений, $\delta_i \in N(0,\ \sigma_B^2)$. Предполагается, что ε_i и δ_i независимы. Из (23) и (24) следует, что случайная величина $(\overline{Y} - \overline{B})$ подчиняется нормальному распределению со средним θ и дисперсией $\left(\sigma_Y^2/n + \sigma^2/n_b\right)$, где $\overline{Y}$ и $\overline{B}$ — средние Y_i и B_i, соответственно. Для них справедливо соотношение

$$\overline{Y} - \overline{B} = \theta + \sqrt{\frac{\sigma_Y^2}{n} + \frac{\sigma_B^2}{n_b}}Z, \tag{25}$$

где Z — нормированная нормальная случайная величина. Уравнение (25) является структурным уравнением для $(\overline{Y} - \overline{B})$.

$$W_y = \frac{(n-1)S_y^2}{\sigma_Y^2} \sim \chi^2(n-1),$$

$$W_b = \frac{(n_b-1)S_b^2}{\sigma_B^2} \sim \chi^2(n_b-1),$$

где $\chi^2(\nu)$ — случайная величина, подчиняющаяся χ^2-распределению с ν степенями свободы, S_y^2 и S_b^2 — выборочные дисперсии для Y_i и B_i, соответственно.

Структурное уравнение для S_y^2

$$S_y^2 = \frac{\sigma_Y^2 W_y}{n-1}. \tag{26}$$

Структурное уравнение для S_b^2

$$S_b^2 = \frac{\sigma_B^2 W_b}{n_b-1}. \tag{27}$$

Решая приведенные выше три структурных уравнения для θ, σ_Y и σ_B, можно получить фидуциальную величину $\tilde{\theta}$ в виде

$$\tilde{\theta} = \overline{y} - \overline{b} - \sqrt{\frac{(n-1)s_y^2}{nW_y} + \frac{(n_b-1)s_b^2}{n_b W_b}}Z. \tag{28}$$

10.2.1.2 Фидуциальный интервал для θ, соответствующий вероятности (1 – α), имеет вид $\left(\tilde{\theta}_{\alpha/2}, \tilde{\theta}_{1-\alpha/2}\right)$, где $\tilde{\theta}_{\alpha}$ — квантиль распределения $\tilde{\theta}$ уровня α. В простых ситуациях эти квантили могут быть определены аналитически. Однако более удобно использовать их приближения с помощью метода Монте-Карло. Этот метод обеспечивает генерацию большого количества реализаций распределения $\tilde{\theta}$ и позволяет получить эмпирические квантили уровней α/2 и (1 – α/2), которые используют в качестве оценок $\tilde{\theta}_{\alpha/2}$ и $\tilde{\theta}_{1-\alpha/2}$. Единственная реализация $\tilde{\theta}$ может быть получена следующим образом.

а) Генерируют реализацию нормированной нормальной случайной величины *Z*.

б) Генерируют реализацию независимых случайных величин W_y и W_b из χ^2-распределений с $(n-1)$ и (n_b-1) степенями свободы, соответственно.

в) Вычисляют $\tilde{\theta}$ в соответствии с (28).

Для данного примера $n = n_b = 5$; $\bar{y} = 3{,}537$; $s_y = 0{,}342$, $\bar{b} = 1{,}228$ и $s_b = 0{,}131$. R-программа для генерации 500000 реализаций приведена ниже.

```
nrun = 500000
Z = rnorm (nrun)
W1 = rchisq (nrun, 4)
Wb = rchisq (nrun, 4)
theta = 3,537 – 1,228 – sqrt (4*0,342^2 / (5*W1) + 4*0,131^2 / (5*Wb)) *Z
Среднее моделируемого распределения:
mean (theta)
## 2,308893
```

Фидуциальный интервал уровня 95 % на основе квантилей уровней 0,025 и 0,975 моделируемого распределения:

```
quantile (theta, c (0,025; 0,975))
## 2,5 %; 97,5 %
## 1,857814; 2,760931
```

Таким образом, фидуциальный интервал уровня 95 % имеет вид (1,858; 2,761).

10.2.2 Пример б)

10.2.2.1 При отсутствии статистических данных о фоновом шуме предполагается, что известно распределение β и что β и ε_i независимы. Кроме того, предполагается, что распределение β полностью известно, т. е. не имеет неизвестных параметров.

10.2.2.2 Структурным уравнением для $\bar{Y}$ является

$$\bar{Y} = \theta + \beta + \frac{\sigma}{\sqrt{n}} Z. \qquad (29)$$

Используя эту формулу вместе со структурным уравнением для S_y^2 из (26) можно получить фидуциальную величину $\tilde{\theta}$ в виде

$$\tilde{\theta} = \bar{y} - \beta - \frac{s_y}{\sqrt{n}} \frac{Z}{\sqrt{W_y/(n-1)}}.$$

Так как $Z / \sqrt{W_y/(n-1)} = T_{n-1}$ — случайная величина, подчиняющаяся *t*-распределению с $(n-1)$ степенями свободы, то

$$\tilde{\theta} = \bar{y} - \beta - \frac{s_y}{\sqrt{n}} T_{n-1}. \qquad (30)$$

Единственная реализация $\tilde{\theta}$ может быть получена следующим образом.

а) Генерируют реализацию T_{n-1} из распределения Стьюдента с $(n-1)$ степенями свободы.

б) Генерируют β в соответствии с его распределением независимо от T_{n-1}.

в) Вычисляют $\tilde{\theta}$ в соответствии с (30).

Для данного примера предполагается, что β подчиняется равномерному распределению на интервале (1,126; 1,329). Ниже приведена программа генерации 500000 реализаций $\tilde{\theta}$.

```
beta = runif (nrun, 1,126; 1,329)
theta = 3.537 – beta – 0,342/sqrt (5) *rt (nrun,4)
```

Среднее моделируемого распределения:

mean (theta)

2,309454

Фидуциальный интервал уровня 95 % на основе квантилей моделируемого распределения уровней 0,025 и 0,975:

quantile(theta, c(0,025; 0,975))

2,5 %; 97,5 %

1,871685; 2,745590

Таким образом, фидуциальный интервал уровня 95 % имеет вид (1,872; 2,746).

10.2.2.3 Полученный фидуциальный интервал соответствует интервалу неопределенности, полученному в соответствии с GUMS1.

10.2.3 Пример в)

10.2.3.1 Рассмотрим данные примера б). Пусть $\bar{y}$ = 1,196 и s_y = 0,106.

Для 500000 реализаций $\tilde{\theta}$ программа имеет вид:

theta = 1,196 – beta – 0,106/sqrt (5) *rt (nrun,4)

Среднее реализаций:

mean(theta)

##-0,03158058

Среднее находится вне области значений θ. Количество реализаций за пределами области значений θ:

length ((1:nrun) [theta<0])

319168

Ограничения на параметры приводят к усеченному фидуциальному распределению на ограниченном пространстве параметров. Таким образом, используют max($\tilde{\theta}$, 0) для получения реализаций фидуциального распределения θ. Фидуциальный интервал уровня 95 %

quantil(pmax (theta, 0), c (0,025; 0,975))

2,5 %; 97,5 %

0,0000000; 0,1361553

Таким образом, фидуциальный интервал уровня 95 % имеет вид (0,000; 0,136).

10.2.3.2 Способ, приведенный в 10.2.1.1, 10.2.2.2, может быть обобщен на произвольные статистические модели. Описание построения фидуциальных величин приведено в [29]. Более простой способ, применимый для большей части задач, где существуют достаточные статистики, приведен в [30] и рассмотрен в [24] и [25]. Таким образом, общий подход включает следующие этапы.

а) Каждую достаточную статистику представляют в виде функции одного или более параметров и случайных величин, распределения которых полностью известны (не включают неизвестные параметры). Таким образом, получают структурное уравнение для каждой достаточной статистики.

б) В каждом структурном уравнении каждый параметр представляют в виде функции достаточных статистик и случайных величин, распределения которых полностью известны.

в) Получают фидуциальную величину для каждого параметра, заменяя достаточные статистики соответствующими им наблюдаемыми значениями.

11 Обсуждение результатов

11.1 Сопоставление оценок неопределенности, полученных на основе трех статистических подходов

11.1.1 В таблице 3 приведены результаты примера. Решения для примера а) и примера б) на основе частотного бутстреп-метода, байесовского и фидуциального подходов очень похожи. Бутстреп-решения и решения в соответствии с GUM дают более короткие интервалы для обоих примеров а) и б). Более существенные различия присутствуют в решении примера в). В этом случае байесовское решение на основе равномерной априорной плотности распределения дают интервалы, которые заметно больше, чем интервалы, построенные другими методами. Только консервативный (гарантированный) интервал Эйзенхарта больше байесовского интервала.

Таблица 3 — Интервалы неопределенности для трех статистических подходов (см. пример)

Пример	GUM	Интервал Эйзенхарта	Бутстреп-подход	Байесовский подход	Фидуциальный подход
Пример а)	(1,89; 2,73)	(1,89; 2,73)	(1,83; 2,66)	(1,81; 2,82)	(1,86; 2,76)
Пример б)	(1,90; 2,72)	(1,78; 2,84)	(1,86; 2,64)	(1,83; 2,79)	(1,87; 2,75)
Пример в)	(0,00; 0,12)	(0,00; 0,20)	(0,00; 0,11)	(0,00; 0,19)	(0,00; 0,14)

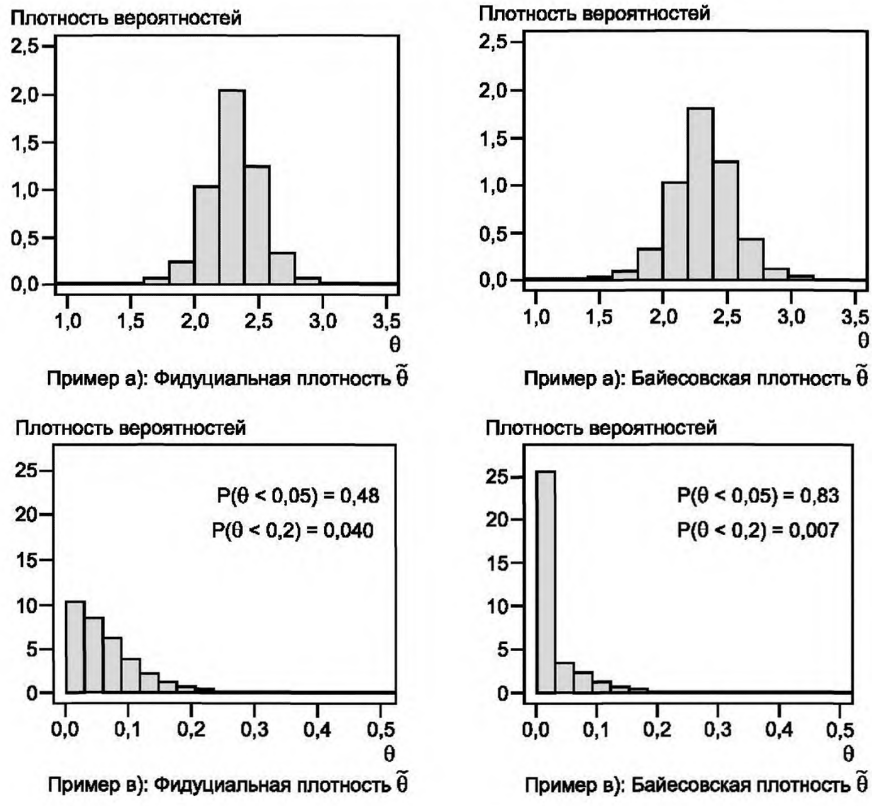


Рисунок 1 — Сопоставление аппроксимаций байесовской и фидуциальной плотности для примеров а) и в)

11.1.2 Поскольку байесовский и фидуциальный подходы дают распределения вероятностей для измеряемой величины θ, то в дополнение к сравнению интервалов расширенной неопределенности в таблице 3 результаты примеров а) и в) также представлены на рисунке 1. Результаты примера б) не показаны, поскольку они визуально неразличимы с результатами примера а). Из гистограммы на рисунке 1 видно, что в случае, когда сигнал заметно превышает фоновый шум, байесовское апостериорное распределение вероятностей и фидуциальное распределение для θ аналогичны. Однако, если сигнал близок к фоновому шуму, то эти два распределения имеют различные характеристики в связи с различными методами включения физических ограничений, присущих каждой задаче.

11.1.3 В частотном подходе предполагается, что измеряемая величина θ и входные величины $\mu_1, \ldots, \mu_p$ в модели измерений (1) являются неизвестными постоянными величинами. Этот подход является достаточно разумным, если измеряемая величина представляет собой физическую константу, для которой предыдущие исследования не обеспечивают информативное априорное распределение или структурное уравнение. Этот подход одобрен статистиками, которые не считают, что все параметры могут быть смоделированы в виде случайных величин (хотя это, как правило, относится к неопределенности, полученной методами типа В), описаны распределениями вероятностей и интегральной функцией распределения. Такой подход похож на байесовский подход, в котором все параметры характеризуют распределениями вероятностей, но требуется меньшее количество предположений о распределениях.

11.1.4 Бутстреп-метод является известным статистическим методом, который может заменить сложные и часто приближенные доверительные интервалы компьютерным моделированием. Существуют различные бутстреп-схемы построения доверительных интервалов в различных условиях. Параметрический *t*-бутстреп интервал, приведенный в настоящих рекомендациях, является усовершенствованным интервалом Стьюдента (*t*-интервала), приведенного в GUM. Преимущество бутстреп-метода состоит в его простоте и возможности построения доверительных интервалов (см. примеры).

11.1.5 В примерах показано также, что байесовская оценка неопределенности, использующая статистическую модель, концептуально проста, и может быть применена в сложных задачах метрологии без изменения основного метода. Систематические влияния, которые не могут быть оценены с помощью результатов измерений (при отсутствии функций наблюдений, математические ожидания которых равны систематическому влиянию), и для которых получают оценку неопределенности типа В, могут легко быть включены в байесовскую модель. Вычисление апостериорных распределений может быть выполнено с использованием метода Монте-Карло по схеме Марковской цепи с применением существующего программного обеспечения.

11.1.6 У байесовского метода существуют некоторые недостатки. Самый серьезный недостаток состоит в том, что априорные распределения должны быть определены для всех параметров модели измерений, включая измеряемые величины. Даже при том, что в метрологии информация для построения априорных распределений часто представлена в виде оценок неопределенности типа В, не редко бывает так, что для одного или двух параметров назначают неинформативные априорные распределения из-за недостатка предварительной информации. Как было показано в примере в) такие распределения могут влиять на результаты. Поэтому желательно выполнять анализ чувствительности для определения степени такого влияния. Существенные влияния, являющиеся результатом априорных, неинформативных характеристик, требуют проведения исследований измерительной системы. Наличие таких влияний означает, что в полученных данных недостаточно информации об измеряемой величине и, таким образом, априорное распределение имеет значительное влияние на результат. В некоторых случаях эта проблема может быть решена путем увеличения количества повторных измерений или изменением способа сбора данных. Существуют ситуации, когда в используемой математической модели слишком много параметров, для которых отсутствует реальная априорная информация, и следовательно, модель должна быть упрощена.

11.1.7 Когда существенная априорная информация об измеряемой величине действительно существует, она может быть введена и эффективно обновляться через теорему Байеса. Далее, чувствительность к форме априорного распределения не только для измеряемой величины, но и для стандартного отклонения является признаком наличия проблем с измерительной системой. В этом случае они должны быть изучены и устранены.

11.1.8 Фидуциальный подход является основой, позволяющей связать распределение с исследуемым параметром. Результаты исследования ГОСТ ИСО/МЭК 17025 показали, что фидуциальный вывод является реальным статистическим методом. Приведенные примеры показали, что фидуциальный подход помогает легко включать информацию о неопределенности в модель измерений и определять оценки измеряемой величины и соответствующей стандартной неопределенности.

Нет никакой необходимости в распространении неопределенности на основе последовательных разложений Тэйлора или методе Велча-Саттервейта при использовании фидуциального подхода.

11.1.9 В фидуциальном подходе вследствие выбора формы структурного уравнения существует проблема не единственности. Однако следует отметить, что в большинстве случаев физический процесс, в результате которого получены данные, известен. В этом случае выбранные структурные уравнения должны отражать этот процесс, таким образом, устраняя проблему не единственности. В метрологии при измерении некоторой величины и использовании некоторых известных процессов, можно утверждать, что случайные погрешности влияют на результат измерений некоторым указанным способом. Полученные результаты измерений отражают модель измерений, которая включает результаты измерений и ошибки измерений в форме влияющих величин. Эта модель может быть выбрана в качестве структурного уравнения.

11.2 Соотношение между методами, предложенными в GUMS1 и тремя статистическими подходами

11.2.1 GUMS1 требует генерации случайных чисел из распределения вероятностей для выходной величины *Y* в соответствии с моделью измерений, которая описывает знание этой величины, осно-

ванное на знании входных величин в соответствии с их функциями плотности вероятностей. В GUM подчеркнуто, что плотность распределения вероятностей не следует понимать как частотную функцию плотности (гистограмму). В качестве измеряемой величины в GUM определена входная величина *Y*. Таким образом, результатами анализа в соответствии с GUMS1, такими как среднее и стандартное отклонение, полученными на основе метода Монте-Карло, являются оценки параметров плотности вероятностей измеряемой величины. Следовательно, возможно сопоставление методов GUMS1, фидуциального метода и традиционного байесовского метода. Интервалы неопределенности по GUMS1 могут быть исследованы на частоту охвата, но их не следует интерпретировать как обычные доверительные интервалы.

11.2.2 В соответствии с 9.1.1 и 9.1.2 традиционные байесовские методы основаны на статистической модели, которая учитывает предварительные сведения об измеряемой величине. Это утверждение не верно для GUMS1, поскольку этот метод основан на модели измерений, где измеряемая величина является выходной величиной, и, таким образом, ее распределение вероятностей полностью определяют плотности вероятностей входных величин. Таким образом, любое прямое сопоставление выводов традиционных байесовских методов и методов GUMS1 ограничено случаем отсутствия предварительной информации об измеряемой величине.

11.2.3 В [34] выполнено такое сопоставление для частной, но широко распространенной задачи измерений. В [34] измеряемая величина μ является функцией α и β, т. е. модель измерений имеет вид $\mu = f(\alpha, \beta)$. Оценка параметра *a* может быть определена по данным, которые представляют собой реализации случайной величины *X*, подчиняющейся нормальному распределению. Для определения оценки параметра β нет никаких данных, но известно его фидуциальное распределение. Анализ, приведенный в GUMS1 (см. п. 6.4.9.2), устанавливает масштабированное смещенное *t*-распределение для α, и затем определяет распределение β через распределение α и функцию *f*. В [34] показано, что этот анализ эквивалентен вычислению байесовской плотности вероятностей для функции $f(\alpha, \beta)$, при условии, что эти два параметра независимы, функция правдоподобия для *X* является нормальной со средним α, для α задано равномерное априорное распределение, а плотность для β задана распределением, построенным на основе предположений. Следует отметить, что априорная плотность μ в этом случае не используется.

11.2.4 Предположим, что существует функция *g*, такая, что $\alpha = g(\mu, \beta)$. Традиционный байесовский анализ использует нормальную функцию правдоподобия для *X* со средним $g(\mu, \beta)$ и априорными распределениями μ и β. В отсутствие дополнительной информации об измеряемой величине для β может быть использовано равномерное распределение, но могут быть использованы и другие распределения. В основе выбора априорного распределения лежит доверенность в распределении β. Обычно μ и β являются независимыми случайными величинами. Следует отметить, что в данной модели, априорная плотность α не использована.

11.2.5 В GUMS1 и традиционном байесовском анализе использована различная параметризация одной и той же статистической модели. Модель, использованная в GUMS1, не предполагает знания средних плотностей вероятностей μ, а использует неинформативное априорное распределение α. Как указано в [34] это различные предположения. Традиционный байесовский анализ использует для измеряемой величины μ неинформативное априорное распределение. В [34] показано, что два эти исследования приводят к идентичным распределениям вероятностей для измеряемой величины, когда в байесовском анализе использовано равномерное априорное распределение для μ и функция *f* линейна. Для нелинейных функций распределения вероятностей μ, полученные этими двумя способами параметризации, не совпадают. Важно отметить, что если неинформативное априорное распределение преобразовано в априорное распределение μ, то соответствующий байесовский анализ приводит к тем же самым результатам что и GUMS1.

11.2.6 Как было отмечено GUMS1 позволяет на основе модели измерений получить плотность вероятностей измеряемой величины путем распространения плотностей вероятностей входных величин. Полученная плотность описывает знания об измеряемой величине, содержащиеся в наблюдаемых данных и предположениях, сделанных при назначении объединенной плотности вероятностей входных данных. Во многих стандартных моделях с результатами измерений, подчиняющимися одномерному нормальному распределению, интервалы неопределенности, полученные в соответствии с GUMS1 и фидуциальным методом, очень похожи и даже идентичны. В соответствии с моделью измерений примера а)

$$\theta = \gamma - \beta,$$

с $Y_i \in N(\gamma, \sigma_Y^2)$, $i = 1,...,5$ и $B_j \in N(\beta, \sigma_B^2)$, $j = 1,...,5$. На основании GUMS1, в качестве плотности распределения γ и β назначено смещенное масштабированное t-распределение. Плотность распределения γ совпадает с распределением случайной величины

$$\overline{y} - \frac{s_y}{\sqrt{5}} T_4^{(1)},$$

где $T_4^{(1)}$ — случайная величина, подчиняющаяся распределению Стьюдента с четырьмя степенями свободы, плотность распределения β имеет то же распределение, что и случайная величина

$$\overline{b} - \frac{s_b}{\sqrt{5}} T_4^{(2)},$$

где $T_4^{(2)}$ — случайная величина, подчиняющаяся распределению Стьюдента с четырьмя степенями свободы, независимая от $T_4^{(1)}$. Следовательно, плотность распределения измеряемой величины θ может быть получена из распределения величины

$$\overline{y} - \overline{b} - \frac{s_y}{\sqrt{5}} T_4^{(1)} + \frac{s_b}{\sqrt{5}} T_4^{(2)}.$$

Ниже приведена R-программа генерации 500000 реализаций из вышеупомянутого распределения.

```
nrun = 500000
T1 = rt (nrun, 4)
T2 = rt (nrun, 4)
theta = 3,537 – 1,228 – 0,342/sqrt (5) *T1 + 0,131/sqrt (5) *T2
```

Интервал неопределенности уровня 95 % на основе квантилей аппроксимации плотности распределения уровней 0,025 и 0,975:

```
quantile (theta, (0,025; 0,975))
## 2,5 %; 97,5 %
```

Полученный интервал (1,853703; 2,763999) идентичен фидуциальному интервалу для этого примера. Точно так же подход GUMS1 и фидуциальный подходы дают один и тот же интервал неопределенности для задач примеров б) и в).

11.2.7 Существует много других ситуаций, когда приведенный в GUMS1 и фидуциальный методы приводят к различным результатам. Особый случай описан в задаче [35]. В случае [35] измеряемой величиной является комплексная величина

$$\Gamma = \Gamma_1 + i\Gamma_2.$$

Таким образом

$$|\Gamma| = \sqrt{\Gamma_1^2 + \Gamma_2^2}.$$

Если $X_1 \in N(\Gamma_1, \sigma^2)$ и $X_2 \in N(\Gamma_2, \sigma^2)$ с известным σ, метод ГОСТ Р 54500.3.1 назначает $N(x_1, \sigma^2)$ в качестве плотности распределения Γ_1 и $N(x_2, \sigma^2)$ — для Γ_2. Следовательно, для |Γ| в соответствии с GUMS1 плотность распределения имеет вид:

$$\sqrt{(x_1 - \sigma Z_1)^2 + (x_2 - \sigma Z_2)^2}, \qquad (31)$$

где Z_1 и Z_2 — независимые случайные величины, подчиняющиеся нормированному нормальному распределению. В [35] показано, что интервалы, построенные методом GUMS1 для |Γ|, имеют вероятности охвата менее заданных, если |Γ| мало по сравнению с σ. Это происходит потому, что случайная величина в выражении (31) положительна и, следовательно, нижняя граница интервала неопределенности также положительна и возможна ситуация, если |Γ| близко к 0, когда интервал не накрывает |Γ|.

11.2.8 Фидуциальное решение этой задачи может быть получено на основе того, что $\frac{X_1^2 + X_2^2}{\sigma^2}$ подчиняется нецентральному χ^2-распределению с двумя степенями свободы и параметром нецентральности $\lambda = |\Gamma|^2/\sigma^2$. Это свойство может быть использовано для формирования структурного уравнения, которое связывает наблюдаемую статистику $\left(x_1^2 + x_2^2\right)/\sigma^2$ с параметром λ, включающим исследуемый параметр |Γ|. На основе этого структурного уравнения можно построить фидуциальный интервал для |Γ|. В [36] показано, что этот фидуциальный интервал обеспечивает охват во всех ситуациях.

12 Заключение

12.1 В настоящих рекомендациях рассмотрено три подхода к построению интервалов неопределенности при наличии четкой вероятностной интерпретации.

Анализ, приведенный в [9], показывает, что интервалы неопределенности, полученные при использовании разных подходов, часто дают близкие результаты, однако интерпретации их различны.

12.2 При частотном подходе справедливо утверждение о том, что при многократном выполнении конкретной процедуры построения интервала неопределенности, вероятность соответствует заданной в идентичных условиях. Таким образом, вероятностное утверждение не относится непосредственно к измеряемой величине, а лишь к соотношению процедуры, которой интервал был построен, и измеряемой величины. После того, как получены результаты измерений и интервал неопределенности вычислен частотным методом, в результатах больше нет ничего случайного. Несмотря на то, что неизвестно накроет ли конкретный интервал измеряемую величину, построенные таким образом интервалы будут накрывать измеряемую величину с установленной вероятностью. В отличие от традиционного доверительного интервала, основанного только на статистических данных, частотный интервал неопределенности, как правило, строят так, чтобы заданный доверительный уровень в среднем был достигнут после объединения всех распределений вероятностей, полученных с использованием оценок неопределенности типа В.

12.3 При построении байесовского и фидуциального интервалов неопределенности использованы распределения вероятностей, основанные на знании измеряемой величины. Методы построения этих двух типов интервалов различны, но результаты подобны. Байесовские результаты получены путем комбинирования распределений вероятностей для каждого параметра с вероятностной моделью, описывающей вариацию данных на основе теоремы Байеса. Получающиеся апостериорные распределения для каждого параметра отражают вероятность реализации конкретного значения параметра с учетом априорной информации и данных. Фидуциальные результаты получены путем обращения вероятностной модели при заданных параметрах для получения распределения параметров при заданных данных.

12.4 Если бы числовые результаты всегда были аналогичны, то каждая из интерпретаций была бы применима (по крайней мере, приближенно) к каждому интервалу неопределенности. Однако, как показано в примерах настоящих рекомендаций, результаты в некоторых случаях могут отличаться друг от друга, даже при том, что каждый имеет вероятностную природу и одинаковый уровень доверия (чаще всего 95 %). Могут быть и другие различия. Например, если один из доминирующих источников неопределенности соответствует случайной величине, имеющей асимметричное распределение, то интервалы неопределенности, полученные с использованием байесовского или доверительного интервалов, отражают эту асимметрию, в то время как доверительный интервал, построенный на основе подхода GUM, дает симметричный интервал неопределенности (это означает, что он является более длинным). Результаты частотного метода, основанного на других статистических принципах, могут в некоторых случаях соответствовать байесовским или фидуциальным результатам, но в общем случае они не совпадают, потому что каждый подход основан на своем наборе математических предположений и критериев.

12.5 Существование разных подходов для оценки неопределенности может вызывать затруднения. Однако следует воспринимать это как возможность дальнейших улучшений, которые позволят создать понятные и удобные в применении методы, эффективно использующие ресурсы, применимые ко многим типам измерений.

Библиография

[1] ISO/IEC Guide 98-3:2008, Uncertainty of measurement — Part 3: Guide to the expression of uncertainty in measurement (GUM:1995)

[2] ISO/IEC 17025:2005, General requirements for the competence of testing and calibration laboratories, 2005

[3] ISO/IEC Guide 98-3:2008/Suppl 1:2008, Uncertainty of measurement — Part 3: Guide to the expression of uncertainty in measurement (GUM:1995) — Supplement 1: Propagation of distributions using a Monte Carlo method

[4] ISO/IEC Guide 99:2007, International vocabulary of metrology — Basic and general concepts and associated terms (VIM)

[5] GLESER, L.J. Assessing uncertainty in measurement. Statistical Science, 13:277—290, 1998

[6] KACKER, R. and JONES, A. On use of Bayesian statistics to make the guide to the expression of uncertainty in measurement consistent. Metrologia, 40:235—248, 2003

[7] ELSTER, C. W., WÖGER, W. and COX, M.G. Draft GUM Supplement 1 and Bayesian analysis. Metrologia, 44:L31—L32, 2007

[8] WILLINK, R. A procedure for the evaluation of measurement uncertainty based on moments. Metrologia, 42:329–343, 2005

[9] LIRA, I. and WÖGER, W. Comparison between the conventional and Bayesian approaches to evaluate measurement data. Metrologia, 43: S249—S259, 2006

[10] GUTHRIE, W.F., LIU, H.K., RUKHIN, A.L., TOMAN, B., WANG, C.M. and ZHANG, N.F. Three statistical paradigms for the assessment and interpretation of measurement uncertainty. Data Modeling for Metrology and Testing in Measurement Sciences, edit. Pavese, F. and Forbes, A. B. Birkhauser, Boston, 2008

[11] KIRKUP, L. and FRENKEL, B. An Introduction to Uncertainty in Measurement Using the GUM. Cambridge University Press, Cambridge UK, 2006

[12] BAYES, T. An essay toward solving a problem in the doctrine of chances. Philosophical Transactions of the Royal Society of London, 53:370—418, 1764. (facsimile available at) gost.e-ecolog.ru

[13] FISHER, R.A. Inverse probability. Proc. Comb. Philos. Soc., 26:528—535, 1930

[14] CASELLA, G. and BERGER, R. Statistical Inference. Duxbury, MA, 2 edition, 2002

[15] GUTHRIE, W.F. Should (T1-T2) have larger uncertainty than T1? Proceedings of the 8th International Conference on Temperature: Its Measurements and Control, 2:887—892, 2002. gost.e-ecolog.ru guthrie-2002-01.pdf)

[16] EFRON, B. and TIBSHIRANI, R.J. An Introduction to the Bootstrap. Monographs of Statistics and Applied Probability, volume 57. Chapman and Hall, 1993

[17] EISENHART, C. Expression of the uncertainties of final measurements results. NBS special publication, NIST, Gaithersburg, MD, 1983

[18] R Development Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, 2003. ISBN 3-900051-00-3, gost.e-ecolog.ru

[19] LUNN, D.J, THOMAS, A., BEST, N. and SPIEGELHALTER, D. WinBUGS — a Bayesian modeling framework: concepts, structure, and extensibility. Statistics and Computing, 10:325-337, 2000

[20] BERNARDO, J.M. and SMITH, A.F.M. Bayesian Theory. John Wiley and Sons Ltd, 1994

[21] HASTIE, T., TIBSHIRANI, R. and FRIEDMAN, J. The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Springer-Verlag, New York, 2001

[22] GELMAN, A., CARLIN, J.B., STERN, H.S., and RUBIN, D.B. Bayesian Data Analysis. Chapman and Hall, 1995

[23] BROWNE, W.J. and DRAPER, D. A comparison of Bayesian and likelihood-based methods for fitting multi-level models. Bayesian Analysis, 1:473—514, 2006

[24] WANG, C.M. and IYER, H.K. Propagation of uncertainties in measurements using generalized inference. Metrologia, 42:145—153, 2005

[25] WANG, C.M. and IYER, H.K. A generalized confidence interval for a measurand in the presence of Type-A and Type-B uncertainties. Measurement, 39:856—863, 2006

[26] HANNIG, J., IYER, H. K., and PATTERSON, P. L. Fiducial generalized confidence intervals. Journal of the American Statistical Association, 101:254—269, 2006

[27] WANG, C.M. and IYER, H.K. Uncertainty analysis for vector measurands using fiducial inference. Metrologia, 43:486—494, 2006

[28] FRASER, D.A.S. The Structure of Inference. New York: Krieger, 1968

[29] HANNIG, J. On fiducial inference — the good, the bad and the ugly. Technical Report 2006/3, Department of Statistics, Colorado State University, Fort Collins, CO, 2006. URL gost.e-ecolog.ru

[30] IYER, H.K. and PATTERSON, P.L. A recipe for constructing generalized pivot quantities and generalized confidence intervals. Technical Report 2002/10, Department of Statistics, Colorado State University, Fort Collins, CO, 2002. URL gost.e-ecolog.ru

[31] RUKHIN, A.L. and SEDRANSK, N. Statistics in metrology: international key comparisons and interlaboratory studies. Journal of Data Science, 7:393—412, 2007

[32] EFRON, B. Six questions raised by the bootstrap. In: Exploring the Limits of Bootstrap (R. LePage and L. Billard, editors) pages 99—126. Wiley, NY, 1992

[33] LINDLEY, D. and SMITH, A.F.M. Bayes estimates for the linear model, JRSS B., 34:1—41, 1972

[34] ELSTER, C. and TOMAN, B. Bayesian uncertainty analysis under prior ignorance of the measurand versus analysis using the Supplement 1 to the Guide: a comparison, Metrologia, 46:261-266, 2009

[35] HALL, B.D. Evaluating methods of calculating measurement uncertainty, Metrologia, 45:L5-L8, 2008

[36] WANG, C.M. and IYER, H.K. Fiducial intervals for the magnitude of a complex-valued quantity, Metrologia, 46:1 81—86, 2009

УДК 389.14:006.354 ОКС 17.020 Т80

Ключевые слова: измерения, измеряемая величина, результат измерения, неопределенность, стандартная неопределенность, байесовский анализ, достаточная статистика, модель наблюдений, структурное уравнение, частотный подход, бутстреп-метод, фидуциальный подход, оценка неопределенности типа А, оценка неопределенности типа В

Редактор *Л.Б. Базякина*
Технический редактор *В.Н. Прусакова*
Корректор *М.М. Малахова*
Компьютерная верстка *Л.А. Круговой*

Сдано в набор 01.04.2015. Подписано в печать 22.05.2015. Формат 60×84⅛. Гарнитура Ариал.
Усл. печ. л. 3,72. Уч.-изд. л. 3,20. Тираж 70 экз. Зак. 1995.

Издано и отпечатано во ФГУП «СТАНДАРТИНФОРМ», 123995 Москва, Гранатный пер., 4.

gost.e-ecolog.ru gost.e-ecolog.ru

gost.e-ecolog.ru