

УДК 389.14:53.083

С.Ф. Левин

Московский институт экспертизы и испытаний, Москва, Россия

ФИЛОСОФИЯ И МАТЕМАТИЧЕСКАЯ СТАТИСТИКА В МЕТРОЛОГИИ

Рассмотрены философские и статистические аспекты проблемы истинного значения в метрологии. Проводится критический анализ терминологии российских, европейских и международных нормативных документов.

Ключевые слова: истинное значение, толерантность, неопределенность – распределение и параметр.

Введение

До введения ГОСТ 16263–70 [1], рекомендаций по межгосударственной стандартизации РМГ 29–99 [2] и ДСТУ 2681–94 [3] метрологию определяли в [4] как «учение о мерах, отрасль физики». В [1] в примечаниях к термину «физическая величина» отмечалось, что «термин допускается применять для свойств, изучаемых не только в физике, но и в химии или других науках, если для сравнения их количественного содержания требуется применение физических методов». И в математической физике отношение математики к физической реальности рассматривается с позиции аксиоматики Дедекинда–Кантора–Вейерштрасса теории действительных чисел [5] и евклидовой геометрии. «Польза математики» в том, что «в физике фундаментальное понятие измерения близко понятию сложения, а большинство физических законов суть утверждения о пропорциональности, что соответствует понятиям умножения и деления» [6].

В центральную проблему отечественной метрологии на рубеже XX-го и XXI-го веков термин «истинное значение физической величины» [1] превратило «философское» примечание, появившееся в РМГ 29–99 [2]:

3.6 истинное значение физической величины

Значение физической величины, которое идеальным образом характеризует в качественном и количественном отношении соответствующую физическую величину.

Примечание – Истинное значение физической величины может быть соотнесено с понятием абсолютной истины. Оно может быть получено только в результате бесконечного процесса измерений с бесконечным совершенствованием методов и средств измерений.

Это примечание к термину, определение которого было перенесено из стандарта [1] без изменений, не раскрыло смысл ключевой части определения «идеальным образом ... в качественном и количественном отношении». Для этого потребовалось обнаружение баланса структурной и параметрической составляющих погрешности неадекватности

математической модели объекта измерений. Далее в РМГ 29–99 [2] был сделан ряд противоречивых уточнений этого «философского определения»:

10.1 погрешность средства измерений

Разность между показанием средства измерений и истинным (действительным) значением измеряемой величины.

Примечания ... 2. Поскольку истинное значение физической величины неизвестно, то на практике пользуются ее действительным значением.

3.7 действительное значение физической величины

Значение физической величины, полученное экспериментальным путем и настолько близкое к истинному значению, что в поставленной измерительной задаче может быть использовано вместо него.

В ДСТУ 2681–94 [3] этому термину соответствует термин «умовно істинне значення».

Но если истинное значение неизвестно, то как можно узнать, на сколько именно близко к нему действительное значение?

Образовавшуюся терминологическую «пустоту» как по команде заполнило зарубежное «Руководство по выражению неопределенности измерения» (GUM) анонимных «международных» экспертов с главной идеей «при указании значения измеряемой величины необходимо давать ее наилучшую оценку и наилучшее оценивание неопределенности этой оценки» за счёт отказа от неизвестного истинного значения физической величины и погрешности измерения. Их следует заменить средним арифметическим «ряда результатов измерений» и «параметром, характеризующим дисперсию» [7]:

неопределенность (измерения) есть параметр, связанный с результатом измерения, который характеризует дисперсию значений, которые могли быть обоснованно приписаны измеряемой величине.

Удивительно не то, что параметр, характеризующий дисперсию, называют дисперсией или стандартным отклонением как корнем квадратным из дисперсии, а их оценками являются сред-

ний квадрат отклонения от среднего арифметического и «среднее квадратичное отклонение». Удивительно то, что GUM [7] был представлен «главными теоретиками по внедрению неопределенности в отечественные измерения» как новый этап развития метрологии. Но их попытки усилить его различного рода «дополнениями» и «пояснениями» [8–12] обернулись примерами доходящего до казусов некорректного применения статистических методов (подробнее см. [13–18]). Поэтому за GUM закрепилось полушутливое название «Руководства по выражению дисперсии измерения». Перевод [7] вызвал бесплоднейшую дискуссию. В её пылу не все обратили внимание на то, что в переводе GUM открытым текстом указаны его «особенности» с использованием непрофессиональной терминологии:

Приложение Е

Е.1.1 Данное *Руководство* представляет широко применяемый метод оценивания и выражения неопределенности в измерении. Оно дает скорее реалистическое, чем «безопасное» значение неопределенности ... [7, с. 52]

GUM рекомендует не интервальные («безопасные»), а точечные (реалистические) оценки стандартного отклонения, не соответствующие требованиям [19] по доверительной вероятности из ряда норм $P = \{0,90; 0,95; 0,99\}$, а для пределов допускаемых погрешностей принимают $P = 1$ [20].

Кроме того, этот «широко применяемый метод оценивания и выражения неопределенности в измерении» расходится с международными стандартами по статистическим методам [21–22] не только в методике расчётов, но и в определении терминов.

В GUM прямо указана предпочтительность «реалистических» статистических оценок по сравнению с «безопасными» при несовпадении определения термина «уровень доверия» с одноимённым термином математической статистики. Это исключает непосредственное применение концепции неопределенности в государственных поверочных схемах из-за нарушения требований по доверительной вероятности (подробнее см. [23–31]).

Так «перестройка» от единства измерений к traceability зашла в «статистический тупик» [32–33], хотя эквивалентом этого англоязычного термина в [2] было «единство измерений». А «философское» примечание продолжилось «выкручиванием рук при введении нового понятия» [34].

Философская проблема № 1: проблема истины

Согласно Международному словарю [35]

Истинное значение – значение величины согласно определению величины.

Примечания: 1 – В концепции погрешности при описании измерения истинное значение величины рассматривается единственным и, на практике, неизвестным. Концепция неопределенности признает то, что в действительности из-за недостаточно детального определения величины не существует единственного истинного значения величины, а есть совокупность истинных значений величины, соответствующих определению. Однако эта совокупность значений, в принципе и на практике, неизвестна. Другие подходы совершенно не требуют понятия истинного значения величины и опираются на понятие метрологической совместимости результатов измерений при оценивании их правильности. 2 – В специальном случае фундаментальной константы считается, что величина имеет единственное истинное значение. 3 – Когда внутренняя неопределенность, связанная с определением измеряемой величины, считается ничтожной по сравнению с другими компонентами неопределенности измерения, то можно считать, что измеряемая величина имеет по существу единственное истинное значение.

Но соотносение «истинного значения физической величины» с категорией абсолютной истины направлено не только против понятия погрешности измерения. Оно направлено против нормирования пределов и доверительных границ погрешностей средств измерений, против норм доверительных вероятностей, устанавливаемых государственными поверочными схемами. А в конечном итоге – на создание методического хаоса в этой сфере.

В [35] есть и «компромиссные» термины:

Опорное значение величины – значение величины, используемое в качестве основы для сравнения величин того же рода.

Примечания: 1 – Опорное значение величины, которое может быть истинным значением измеряемой величины, в этом случае оно неизвестно, или принятое значение величины, в этом случае оно известно. ...

Причины появления термина «опорное значение физической величины» в международной и отечественной метрологии оказались разными [14].

Зачем надо сравнивать величины одного рода?

Международный смысл термина «опорное значение» раскрывал бы аутентичный перевод стандарта о «прецизионности» [36], если бы он не был настоящей «неопределенностью», в том числе и в смысле «СКО». Но его единственное конкретное определение – «т.е.» между «математическим ожиданием» и «средним значением совокупности». А сам стандарт рассчитан на случай «отсутствия необходимых эталонов» и «может применяться для оценки точности выполнения измерений состава и свойств очень широкой номенклатуры материалов, включая жидкости, порошкообразные и твердые материалы – продукты материального производства или существующие в природе, при условии, что учитывают любую неоднородность материала» [36]. А вопрос о статистической неоднородности данных в расчётах того же СКО остаётся открытым.

И о каких же «научных принципах» и «экспериментальных работах» в метрологии можно говорить в случае «отсутствия необходимых эталонов», не говоря уже о том, что разница между математическим ожиданием и средним значением в математической статистике называется не «т.е.», а смещением, и является предметом оценивания.

Однако причём здесь абсолютная истина?

Напомним, что философская категория «истины» определяется как «верное, правильное отражение действительности в мысли, критерием которого, в конечном счёте, является практика. Характеристика истинности относится к мыслям, а не к самим вещам и средствам их языкового выражения» [37]. При этом истина как философская категория бывает не только абсолютной, но и относительной.

Но дело ещё и в том, что международные словари, а за ними и [2], определяют базовый для метрологии термин тождеством «физической величины» и «свойства физического объекта»:

3.1 физическая величина

Одно из свойств физического объекта (физической системы, явления и процесса), общее в качественном отношении для многих физических объектов, но в количественном отношении индивидуальное для каждого из них.

Физические величины представляют именем, числом и размерностью, чего в физической реальности нет. Это – символы для свойств физического объекта.

Физическая величина это не само свойство, а всего лишь символ его количественного проявления, что и зафиксировано теперь в [38].

Традиционная интерпретация опорного значения дана [38], детализирована [39] и соответствует отечественной метрологической практике. Опорное значение физической величины используют для расчёта характеристик погрешностей. Истинное значение является аналогом относительной истины [37] и, одновременно, результатом измерения государственным первичным эталоном как наиболее точным узаконенным средством измерений.

Ещё в XVIII веке Л. Эйлер в книге «Алгебра» отметил: «При определении или измерении величин всякого рода ... устанавливается некоторая известная величина этого же рода, именуемая мерой или единицей и зависящая исключительно от нашего произвола.

Затем определяется, в каком отношении находится данная величина к этой мере...».

Тем самым физический размер меры единицы измерения с её номинальным значением может быть и не связан, а физические размеры мер крат-

ных и дольных единиц при построении шкал отношений физических величин устанавливаются методом калибровки. «Философское» же примечание к «истинному значению» в части мер и шкал измерений оказывается ложным.

В то же время статус действительного значения физической величины имеет результат решения специальной метрологической измерительной задачи, для которого размерами наикратчайшего толерантного интервала по сравнению с нормами доверительных границ, установленными для решаемой измерительной задачи, можно пренебречь.

Несоответствие одноимённых терминов математической статистики и GUM в переводе [7] на русский язык из стремления сделать «как лучше» выражено так витиевато, что в 1999 году без специальной подготовки ее сразу нельзя было понять:

6.2.2 Термины **доверительный интервал** (С.2.27, С.2.28) и **уровень доверия** (С.2.29) имеют в статистике специальные определения и применяются к интервалу, определенному U , только когда выполнены определенные условия, включая условие, чтобы все составляющие неопределенности, которые входят в $u_c(y)$, были бы получены из оценивания по типу А. Таким образом, в данном *Руководстве* слово «доверие» не используется для модификации слова «интервал», когда ссылаются на интервал, определяемый U , и термин «доверительный уровень» также не используется в связи с интервалом и предпочитается скорее термин «уровень доверия». Более конкретно, U рассматривается как задание интервала вокруг результата измерения, который содержит большую часть p распределения вероятностей, характеризуемого результатом и его суммарной стандартной неопределенностью, и p является *вероятностью охвата* или *уровнем доверия* для этого интервала.

6.2.3 Если это возможно, необходимо оценить и указать доверительный уровень p , связанный с интервалом, определяемым U . Надо признать, что умножение $u_c(y)$ на какую-то постоянную величину не дает никакой новой информации, а просто представляет ранее имевшуюся информацию в новом виде. Однако нужно также признать, что в большинстве случаев уровень доверия p (особенно для значений p , близких к 1) будет скорее неопределенным не только из-за ограниченного знания распределения вероятностей, характеризующих y и $u_c(y)$ (особенно в крайних областях), но также из-за неопределенности самой $u_c(y)$ (см. Примечание 2 к 2.3.5, 6.3.2 и Приложение G, особенно G.6.6).

Тем не менее, главное – очевидно:

6.3.2 В идеале хотелось бы иметь возможность выбрать конкретное значение коэффициента охвата k , которое обеспечивало бы интервал $Y=y\pm U=y\pm ku_c(y)$, соответствующий выбранному уровню доверия, такому как 95 или 99 процентов; равным образом, для заданного значения k хотелось бы иметь возможность четко указать уровень доверия, связанный с этим интервалом. Однако это нелегко осуществить на практике, поскольку это требует полного знания распределения вероятностей, характеризующего результатом измерения y и его суммарной неопределенностью $u_c(y)$. Хотя эти параметры обладают большой значимостью, сами по себе они недостаточны для того, чтобы установить интервалы, имеющие точно известные уровни доверия.

Оценивание «неопределенности измерения» не устанавливает интервалы, содержащие с заданной доверительной вероятностью P не менее чем заданную долю γ неизвестного распределения вероятностей измеряемой величины. Такие интервалы называются в математической статистике толерантными, а доверительные интервалы характеризуют только точность оценок параметров распределения.

Чтобы понять причину этой путаницы, достаточно сравнить тексты оригинала GUM

C.2.29 confidence coefficient; confidence level [ISO 3534-1, 2.59] – The value $(1-\alpha)$ of the probability associated with a confidence interval or statistical coverage interval. (see [ISO 3534-1] 2.57 [C.2.27], 2.58 [C.2.28]), and 2.61 [C.2.30])

C.2.30 statistical coverage interval [ISO 3534-1, 2.61] – An interval for which a given level of confidence that it contains at least a specified proportion of population.

NOTES ... 2 Also called «statistical tolerance interval». This term should not be used because it may cause confusion with «tolerance interval» which is defined in ISO 3534-2.

и его перевода на русский язык [7]

С.2.29 Коэффициент доверия; доверительный уровень [ISO 3534-1, 2.59] Значение $(1-\alpha)$ вероятности, связанное с доверительным интервалом или статистическим интервалом охвата (см. [ISO 3534-1] 2.57 [C.2.27], 2.58 [C.2.28] и 2.61 [C.2.30]).

С.2.30 Статистический интервал охвата [ISO 3534-1, 2.61] – интервал, для которого можно с заданным доверительным уровнем констатировать, что он включает, по крайней мере, определенную часть совокупности.

ПРИМЕЧАНИЯ ... 2. Его называют также «статистически допустимый интервал». Такой термин не следует использовать, так как это может вызвать путаницу с «допустимым интервалом», определенным в ISO 3534-2.

В художественном отношении перевод выполнен, может быть, и хорошо. А в профессиональном?

Наверное, это уже стало «нормой»:

ГОСТ Р 50779.10–2000 (ISO 3534-1) [40]	
2.59 доверительная вероятность; уровень доверия	<i>en</i> confidence coefficient;
Величина $(1-\alpha)$ – вероятность, связанная с доверительным интервалом или со статистическим накрывающим интервалом.	<i>confidence level</i>
2.61 толерантный интервал	<i>en</i> statistical coverage interval,
Интервал, для которого можно утверждать с данным уровнем доверия, что он содержит, по крайней мере, заданную долю определенной совокупности.	<i>fr</i> intervalle statistique de dispersion
ГОСТ Р 50779.11–2000 (ISO 3534-2) [41]	
1.4.5 поле [область] допуска	<i>en</i> tolerance interval,
Множество значений показателя между предельными значениями, включая последние	<i>tolerance zone; fr</i> intervalle de tolérance

Видимо, трудности восприятия концепции неопределенности дополняются качеством перевода и научного редактирования. В результате для метрологии был «потерян» толерантный интервал, фигурирующий в государственных поверочных схемах в качестве доверительных границ погрешности с указанием нормы доверительной вероятности [19].

Как же так получилось, что долгие годы переводчики, научные редакторы и «главные теоретики по внедрению неопределенности в отечественные измерения» не могли перевести «tolerance interval» как «толерантный интервал»?

Слово «композиция» упоминалось ещё в ГОСТ 8.207–76 [42], но без формулы для композиции распределений случайной и неисклѳенной систематической составляющих погрешности результата «измерения» как среднего арифметического данных в ряду измерений. При этом отмечалось, что погрешность, возникающая из-за пренебрежения одной из составляющих погрешности при их отношении в пределах от 0,8 до 8, не превышает 15%.

В 1984 году РД 50–453 [43] зафиксировал, что при расчѳтах характеристик погрешностей по математическому ожиданию и K -кратному стандартному отклонению для усечѳнных распределений погрешность определения среднего значения коэффициента K может колебаться в пределах от 7 до 65 %.

Тайну происхождения погрешностей в расчѳтах погрешностей приоткрыла в 1987 году метрологическая справочная книга [44]: «Для нахождения погрешности результата измерения надо построить композицию распределения случайных и неисклѳенных систематических погрешностей. Но построение композиции затруднено, поэтому используют эмпирическую формулу расчѳта доверительной погрешности результата.

$$\Delta(P) = t_{\Sigma} \cdot S_{\Sigma}, \tag{6.17}$$

где

$$t_{\Sigma} = [\Theta(P) + \varepsilon(P)] / [S(\bar{A}) + \Theta / \sqrt{3}] -$$

коэффициент, соответствующий q -му уровню значимости композиции распределений случайных и неисключённых систематических погрешностей;

$$S_{\Sigma} = \sqrt{S^2(\bar{A}) + \Theta^2 / 3} -$$

СКО композиции случайных и неисключённых систематических погрешностей.

Вычисление доверительной погрешности результата измерения по формуле (6.17) даёт пренебрежимо малую погрешность, не превышающую 12%. Однако это вычисление довольно громоздко».

В 1998 году в отзыве Российского центра испытаний и сертификации «РОСТЕСТ–Москва» на проект РМГ 29–99 была отмечена некорректность определений терминов и обозначений в формулах, связанных с рассеянием и усреднением данных в ряду измерений. Но эти замечания были отклонены.

В 1999 году в переводе GUM [7] было заявлено, что «расширенная неопределённость» и «интервал охвата» – это не *tolerance interval*, принятый в математической статистике. А математической основой GUM стала центральная предельная теорема теории вероятностей и т.н. «нормальная теория».

В 2000 году в метрологии появилось количественное определение погрешности неадекватности математической модели объекта измерений [39] и «погрешности формул» для расчёта погрешностей.

Требования к формулам в метрологии устанавливались и ранее. Согласно МИ 2091–90 [45] «погрешность, обусловленная несоответствием модели объекту измерений, не должна превышать 10 % от предела допускаемой погрешности измерений». Но и композиция составляющих погрешности также должна учитывать погрешности неадекватности [39]. Но это уже было за рамками традиционной логики статистического вывода. Основной её дефект лежал на поверхности, а именно: критерии согласия, реагируя на расхождение эмпирического и теоретического распределений, в дальнейшем это расхождение в конечном положительном результате проверки гипотезы никак не учитывали.

Этот дефект позволили преодолеть теорема о наибольшем отклонении статистической функции распределения от функции распределения вероятностей (расстоянии Колмогорова), формула Уолша-Уилкса-Роббинса о вероятности попадания между крайними членами вариационного ряда заданной доли распределения и теорема о связи вероятности согласия распределений с расстоянием Колмогорова. Так выяснилось, что в государственных поверочных схемах для погрешностей указаны границы не доверительных, а толерантных интервалов.

В 2001 году анализ [46] примера калибровки термометра [7], иллюстрирующий главную идею GUM, показал, что в этом примере GUM и традиционная теория погрешностей при усреднении методом наименьших квадратов дают решения в виде линейных моделей и поправки, совпадающие до 4-го знака после запятой. Учёт же погрешности неадекватности согласно Р 50.2.004–2000 [39] дал параболическую модель в 1,6 раза более точную.

Однако расчёт значений тока согласно закону Ома при измерениях напряжения и сопротивления в случае распределений Гаусса даёт распределение Коши, а у него дисперсия и, следовательно, «неопределённость измерения», не существует.

В этом же 2001 году в книге [47] было описано предложение профессора Л.К. Исаева – использовать GUM только в тех видах измерений, в которых отсутствуют государственные поверочные схемы.

В 2002 году эксперимент по проверке методом калибровки граммового набора гирь класса точности E_2 и расчёт протоколу с учётом погрешности неадекватности [39] в композиции распределений составляющих показал, что набора требовал перевода в более низкий класс точности F_1 [48].

В 2003 году в Р 50.2.028–2003 [49] был дан пример расчёта методом наименьших квадратов 95%-й «расширенной неопределённости» градуировочной характеристики хроматографа, в полосу которой вошло 12 из 35 градуировочных точек [50].

В 2004 году анализ [51] примера определения плотности твёрдого тела из МИ 2083–90 [52] показал, что необоснованное использование гипотезы о «нормальности распределения» и формулы Тейлора привело к несостоятельным оценкам и существенному завышению точности результатов.

В 2005 году появился первый нормативный документ по расчёту характеристик погрешностей на основе композиции распределений МИ 2916–2005 [38] и стандарты по статистическим методам ГОСТ Р ИСО 16269–6–2005 [53] и ГОСТ Р ИСО 16269–8–2005 [54] с таблицами для границ толерантных и предикционных интервалов, а позже – статья ведущих метрологов ВНИИФТРИ [55] о том, что «формулы, рекомендуемые в МИ 1552–86 и МИ 2083–90 для вычисления доверительной границы погрешности результата измерения, являются неверными».

В 2006 году ГОСТ Р ИСО 10576–1–2006 [56] указал: «4.1.4 Устанавливаемые предельные значения не должны включать в себя (в явном или неявном виде) неопределённость измерений». Но она появилась в браковочном условии методики поверки термометров ГОСТ Р 8.624–2006 [57].

Ещё к проекту этого стандарта возникли вопросы по рассеянию в ряду измерений и примеру расчёта неопределённости измерения. Так, в [58] было указано о его несоответствии стандарту по стати-

стическим методам ГОСТ Р 50779.21–2004 [59]. И было отмечено, что «внедряемые в нормативные документы по поверке средств измерений расчеты согласно «Руководству» завышают оценки точности результатов измерений по ряду причин»:

суммарная стандартная неопределенность не является состоятельной оценкой параметра рассеяния в композиции негауссовых распределений;

уровень доверия стандартных неопределенностей типа А не согласован с доверительной вероятностью, установленной поверочной схемой;

не учитывается неопределенность эквивалентной функции распределения вероятностей искомой при поверке величины и её неадекватность;

расширенная неопределенность не учитывает то, что все реальные распределения усечены.

Игнорирование этих обстоятельств ставит под сомнение и корректность выбора эталонов, и состоятельность «оценённой неопределённости».

Принятое в документах ГСИ по поверке браковочное условие по максимуму абсолютного значения случайной составляющей погрешности согласно критериям математической статистики предпочтительнее расширенной неопределенности. С точки зрения метода максимума правдоподобия такая логика принятия решения эквивалентна выбору не распределения Гаусса, когда расчёт погрешностей ведётся по среднему арифметическому и СКО, а равномерного распределения [20, 42, 43, 60].

И только после статьи [58] последовала «эмоциональная» реакция «главных теоретиков по внедрению неопределенности в отечественные измерения». В «сухом остатке» она сводилась к следующему: «Следует признать, что толерантные интервалы, несправедливо не получившие широкого распространения при обработке данных в метрологии, могли бы эффективно использоваться, на наш взгляд, в тех ситуациях, когда надо охарактеризовать разброс некоторой совокупности данных, например, результатов измерений продукции при ее выборочном контроле и т.д. Однако нельзя согласиться с их применением при обработке данных калибровки в примере, рассматриваемом в [58], поскольку в этом случае определяется МХ конкретного СИ и результатом является ее оценка с соответствующей неопределенностью».

Иначе, присовокупить к МХ СИ неопределённость можно, а толерантный интервал – нельзя [61].

Тем не менее, история толерантного интервала получила неожиданное продолжение.

В 2008 году выяснилось, что в расчетах погрешности действительной массы гири класса F_1 из примера в оригинале и подлиннике МИ 1747–87 [62] и проекта стандарта по поверке мер массы при общих исходных данных во всех трёх случаях получены разные результаты [63].

Появление в ГОСТ 28311-89 [64] «предела допускаемого относительного СКО при доверительной вероятности $\gamma = 0,95$ » прошло почти незаметно. Но когда ГОСТ Р 8.648–2008 [65] ввёл «предел допускаемых относительных доверительных погрешностей при доверительной вероятности 0,95», а за ним ГОСТ Р 8.663-2009 [66] – «предел допускаемых значений доверительных границ относительной погрешности», РОССТАНДАРТ инициировал для метрологов, специалистов и преподавателей, программу повышения квалификации «Статистические методы решения измерительных задач» на базе МИ 2916–2005. И 9 сентября 2010 года программа была согласована и утверждена.

Решение РОССТАНДАРТА вызвано критическим состоянием метрологического обеспечения в 1990-е годы, большими потерями опытных специалистов в это период и, как следствие, снижением уровня метрологической и математической подготовки разработчиков нормативных документов.

С 1 января 2011 года в России введено в действие Изменение № 2 в РМГ 29–99. Были учтены замечания 1998 года – исправлены некорректные названия и определения терминов «Средняя квадратическая погрешность результатов единичных измерений в ряду измерений» и «Средняя квадратическая погрешность результата измерений среднего арифметического», а также ошибки в формулах. Так в отечественной метрологии было восстановлено СКО ряда измерений и оценка стандартного отклонения параметра его положения.

Изменение № 2 стало и признанием того, что коэффициент Стьюдента использовался в ряде нормативных документах не по назначению. Он входит в формулы для доверительных границ оценки параметра положения совокупности данных в ряду измерений и характеризует неисключённую систематическую составляющую. А в оценках доверительных границ параметра рассеяния применяют коэффициенты на основе другого распределения – χ^2 .

С 1 января 2012 года введён в действие ГОСТ 8.381–2009 [67], в котором из РМГ 43–2001 [8] взят, но уже без исходных данных, расчёт по эмпирическим формулам точности вторичного эталона единицы длины, штрихового метра. В примере доверительная граница погрешности штрихового метра при доверительной вероятности $P = 0,95$ и расширенная неопределенность измерений при уровне доверия 0,95 по данным поверки на государственном первичном эталоне оценены соответственно $\Delta_{\Sigma}(0,95) = 0,0713$ мкм $\approx 0,07$ мкм и $U(0,95) = 0,068$ мкм $\approx 0,07$ мкм. Тогда как в государственной поверочной схеме средств измерений длины установлено $P = 0,99$. А расчёт на основе композиции показал, что оценка границы доверительной погрешности занижена не менее чем на 79 %!

С 1 июля 2012 года введён в действие ГОСТ OIML R 111-1–2009 [68] с примером расчёта массы эталонной гири класса точности F_1 из МИ 1747–87, упоминавшимся выше с ошибками в формулах и не удачной подгонкой под правильный результат. Расчёт на основе композиции показал, что согласно данным поверки гиря должна быть забракована.

И, наконец, с 1 октября 2012 года введён в действие ГОСТ Р 54500.3–2011 [69], дублирующий GUM и его редакцию ISO/IEC GUIDE 98-3:2008(E). В нём исправлены ошибочные определения GUM терминов «толерантный интервал» и «доверительная вероятность». «Рассеяние» (dispersion) заменено «разбросом», а совокупности значений величин наконец-то разделены на генеральные и выборочные.

C.2.29 доверительная вероятность <i>en confidence coefficient, confidence level</i> Значение $(1-\alpha)$ вероятности, связанной с доверительным или толерантным интервалом [см. ИСО 3534-1:1993, словарные статьи 2.57 (C.2.27), 2.58 (C.2.28) и 2.61 (C.2.30)].	
C.2.30 толерантный интервал <i>en statistical coverage interval</i> Интервал, для которого можно утверждать с определенной доверительной вероятностью, что он содержит долю генеральной совокупности, не меньшую заданной.	

Так как эти определения в оригинале GUM выглядели иначе, возникает вопрос: «Кто виноват?»

Заинтересованные в ответе на этот вопрос специалисты ВНИИМС (этот институт в 2012 году выступил ведущей организацией по расчётам характеристик погрешности средств измерений по данным поверки или калибровки на основе МИ 2916–2005) любезно предоставили автору оригинальный текст ISO/IEC GUIDE 98-3:2008(E):

C.2.29 confidence coefficient, confidence level The value $(1-\alpha)$ of the probability associated with a confidence interval or a statistical coverage interval [See ISO 3534-1:1993, definitions 2.57 (C.2.27), 2.58 (C.2.28) and 2.61 (C.2.30).] NOTE $(1-\alpha)$ is often expressed as a percentage. [ISO 3534-1:1993, definition 2.59]
C.2.30 statistical coverage interval An interval for which it can be stated with a given level of confidence that it contains at least a specified proportion of the population NOTE 2 Also called “statistical tolerance interval”. This term should not be used because it may cause confusion with “tolerance interval” which is defined in ISO 3534-2:1993. [ISO 3534-1:1993, definition 2.61]

Нетрудно заметить, что принципиальных изменений по сравнению с определениями GUM нет.

Другими словами, зарубежные разработчики GUM и его редакции 2008 года тоже оказались не в курсе точки зрения математической статистики на толерантные интервалы. А в безграмотность авторов тех книг по статистике, которые перечислены в библиографии, верится с трудом.

В итоге, исправление в ГОСТ Р 54500.3–2011 [69] ошибочных переводов и определений терминов «доверительная вероятность» и «толерантный интервал» для одних сделало ещё более очевидными вероятностно-статистические дефекты GUM, а для других ситуация с применением концепции неопределённости в государственных поверочных схемах ещё больше запуталась.

Заключение

Важнейшей проблемой фундаментальной метрологии вдруг оказалось оценивание точности результатов при решении измерительных задач.

В чём же эта проблема выразилась?

Во-первых, в игнорировании смысла доверительной вероятности, нормируемой государственными поверочными схемами. Этой нормой надо руководствоваться при поверке и калибровке!

Во-вторых, в путанице между доверительными и «несправедливо не получившими широкого распространения» толерантными интервалами.

В-третьих, в использовании «простых», но приближенных формул вместо строгих соотношений теории вероятностей по суммированию неопределённых и случайных величин, требующих «вычисления композиции» – распределения вероятностей суммы. И только после введения контурных оценок выяснилось, что «громоздкость» вычислений композиций оказалась сильно преувеличенной.

В-четвёртых, в некорректной интерпретации истинного и действительное значений физической величины на основе понятия абсолютной истины.

В-пятых, в затянувшемся на десятилетие признании ошибочности определений РМГ 29–99 терминов, связанных с оцениванием параметров распределения вероятностей погрешности как характеристик её систематической, случайной и неисключённой систематической составляющих. А когда математические конструкции были названы своими именами, выявилась неадекватность расчётов доверительных границ на основе коэффициентов Стьюдента и «средней квадратической погрешности».

Эти обстоятельства – следствие некорректной формулировки измерительной задачи метода многократных измерений в ГОСТ 8.207–76 как принципа усреднения: «За результат измерения принимают среднее арифметическое результатов наблюдений...». Этот принцип у метрологов получил название принципа «средней температуры по больнице».

С 1 января 2013 года взамен ГОСТ 8.207–76 [42] введен в действие ГОСТ Р 8.736–2011 [70], в котором определение результата «измерения» перефразировано. Теперь за оценку измеряемой величины \bar{x} принимают среднее арифметическое значение исправленных результатов измерений, но для его неисключённых систематических погрешностей используют всё те же «эмпирические» формулы.

И, как в известной английской песенке, заканчивающейся словами «оттого, что в кузнице не было гвоздя», цепочка из «философского» примечания, неполноты формулы Тейлора для погрешности и «закона распространения неопределённости», пробелов в знаниях статистики и боязни громоздкости композиции довела до «чрезмерной торопливости буквально навязывания силой нового понятия».

Список литературы

- ГОСТ 16263–70. ГСИ. Метрология. Термины и определения.
- РМГ 29–99. ГСИ. Метрология. Основные термины и определения.
- ДСТУ 2681–94. Метрологія. Терміни та визначення.
- Словарь иностранных слов. Изд. 6-е, перераб. и доп. М.: Советская энциклопедия, 1964.
- Математическая энциклопедия. Т. 1. А–Г. М.: Советская энциклопедия, 1977.
- Джеффрис Г., Свирлс Б. Методы математической физики. Вып. 1. М.: Мир, 1969.
- Руководство по выражению неопределённости измерения. Пер. с англ. Научный редактор Слаев В.А. СПб: ВНИИМ им. Д.И. Менделеева, 1999.
- РМГ 43–2001. ГСИ. Применение «Руководства по выражению неопределённости измерений».
- РМГ 91–2009. ГСИ. Совместное использование понятий «погрешность измерений» и «неопределённость измерений». Общие принципы.
- ОКРМ 104:2009. Введение к «Руководству по выражению неопределённости измерения» и сопутствующим документам – Оценка данных измерений.
- МИ 3281–2010. ГСИ. Оценка результатов измерений. Пояснения к «Руководству по выражению неопределённости измерений».
- Приложение 1 к «Руководству по выражению неопределённости измерения»: Трансформирование распределений с использованием метода Монте–Карло. СПб: Профессионал, 2010.
- Левин С.Ф. Проблемы несоответствий в метрологии // Системы обработки информации. 2007. № 6(64). С. 56–62.
- Левин С.Ф. Неопределённость как параметр распределения погрешности // Системы обработки информации. 2010. № 4(85). С. 13–19.
- Левин С.Ф. «Новые» «Пояснения по оценке результатов измерений» к «Руководству по выражению неопределённости измерения» // Системы обработки информации. 2011. № 1(91). С. 15–19.
- Левин С.Ф. «Руководство по выражению неопределённости измерения» и измерительные задачи обеспечения единства измерений. Системы обработки информации. 2011. № 6(96). С. 9–14.
- Левин С.Ф. Статистические методы теории измерительных задач и классическая неопределённость // Системы обработки информации. 2011. № 6 (96). С. 26–36.
- Левин С.Ф. Оценивание точности решений измерительных задач // Электротехнические и компьютерные системы. 2012. № 06 (82). С. 27–33.
- ГОСТ 8.061–80 ГСИ. Поверочные схемы. Содержание и построение.
- МИ 1317–2004. ГСИ. Результаты и характеристики погрешности измерений. Формы представления. Способы использования при испытаниях образцов продукции и контроле их параметров.
- ГОСТ Р 50779.21–2004 (ISO 2854:1976) Статистические методы. Правила определения и методы расчета статистических характеристик по выборочным данным. Часть 1. Нормальное распределение.
- ГОСТ Р ИСО 16269–6–2005 Статистические методы. Статистическое представление данных. Определение статистических толерантных интервалов.
- Чуйко В.Г. О влиянии новых терминов на работу практикующего метролога // Измерительная техника. 2004. № 1. С. 20–23.
- Кокс М., Харрис П. Основные положения Приложения 1 к Руководству по выражению неопределённости в измерении // Измерительная техника. 2005. № 4. С. 17–24.
- Левин С.Ф. Неопределённость в узком и широком смысле результатов поверки средств измерений // Измерительная техника. 2007. № 9. С. 15–19.
- Левин С.Ф. Проблема доверительной вероятности // Измерительная техника. 2008. № 9. С. 33–39.
- Левин С.Ф. Нерешенные проблемы неопределённости // Главный метролог. 2009. № 4. С. 13–24.
- Левин С.Ф. Нерешенные проблемы «Руководства по выражению неопределённости измерения» // Метрология. 2009. № 6. С. 3–21.
- Левин С.Ф. Неопределённость как параметр распределения вероятностей: Прикладная нормативно-математическая точка зрения // Главный метролог. 2010. № 5. С. 10–20.
- Рабинович С. Г. О необходимости создания новых рекомендаций по оцениванию погрешностей и неопределённости измерений // Системы обработки информации. 2010. № 4(85). С. 23–26.
- Левин С.Ф. Нужны ли «Пояснения по оценке результатов измерений» к «Руководству по выражению неопределённости измерения»? // Советник метролога. 2011. № 1. С. 49–56.
- Левин С.Ф. О нетрадиционных для метрологии положениях концепции по выражению неопределённости измерения // Метрология. 2011. № 9. С. 4–12.
- Чуйко В. Г. Поверочная схема как инструмент контроля прослеживаемости измерений // Математическая, статистическая и компьютерная поддержка качества измерений: Материалы международного НТС. СПб: КОOMET, ВНИИМ им. Д. И. Менделеева, 2006. С. 84–87.
- Тищенко В.А., Токатлы В.И., Лукьянов В.И. О переводе и заимствовании терминологии из международных метрологических документов // Измерительная техника. 2003. № 10. С. 12–16.
- Международный словарь по метрологии: Основные и общие понятия и соответствующие термины. СПб: Профессионал, 2009. 82 с.
- ГОСТ Р ИСО 5725–2002 Точность (правильность и прецизионность) методов и результатов измерений.
- Философский словарь. 5-е изд./ Под ред. И.Т. Фролова. М.: Политиздат, 1987. 590 с.
- МИ 2916–2005. ГСИ. Идентификация распределений вероятностей при решении измерительных задач.

39. Р 50.2.004–2000. ГСИ. Определение характеристик математических моделей зависимостей между физическими величинами при решении измерительных задач. Основные положения.

40. ГОСТ Р 50779.10–2000 (ИСО 3534.1–93) Статистические методы. Вероятность и основы статистики.

41. ГОСТ Р 50779.11–2000 (ИСО 3534.2–93) Статистические методы. Статистическое управление качеством. Термины и определения.

42. ГОСТ 8.207-76. ГСИ. Прямые измерения с многократными наблюдениями. Методы обработки результатов наблюдений. Основные положения.

43. РД 50–453–84 МУ. Характеристики погрешности средств измерений в реальных условиях эксплуатации. Методы расчета.

44. Селиванов М.Н., Фридман А.Э., Кудряшова Ж.Ф. Качество измерений. Л.: Лениздат, 1987.

45. МИ 2091–90. ГСИ. Измерения физических величин. Общие требования.

46. Левин С.Ф. Метрология. Математическая статистика. Легенды и мифы XX-го века: Легенда о неопределенности // Партнёры и конкуренты. 2001. № 1. С. 13–25.

47. Кузнецов В.А., Ялунина Г.В. Общая метрология. М.: ИПК Издательство стандартов, 2001.

48. Левин С.Ф. Математическая теория измерительных задач. Часть 7: Метод прямого измерения // Контрольно-измерительные приборы и системы. 2002, № 6, с. 30–34; 2003, № 1, с. 35–36.

49. Р 50.2.028–2003. ГСИ. Алгоритмы построение градуировочных характеристик средств измерений состава веществ и материалов и оценивание их погрешностей (неопределенностей). Оценивание погрешности (неопределенности) линейных градуировочных характеристик при использовании метода наименьших квадратов.

50. Левин С.Ф. Обеспечение единства измерений при градуировке измерительных преобразователей // Измерительная техника. 2006. № 7. С. 8–14.

51. Левин С.Ф. Схема приведения в методе косвенного измерения // Измерительная техника. 2004. № 3. С. 5–9.

52. МИ 2083–90. ГСИ. Измерения косвенные. Определение результатов измерений и оценивание их погрешностей.

53. ГОСТ Р ИСО 16269-6–2005 Статистические методы. Статистическое представление данных. Определение статистических толерантных интервалов.

54. ГОСТ Р ИСО 16269-8–2005 Статистические методы. Статистическое представление данных. Определение предикционных интервалов.

55. Тищенко В.А., Токатлы В.И., Лукьянов В.И. Комментарии к метрологическим документам, регла-

ментирующим обработку результатов измерений // Законодательная и прикладная метрология. 2006. № 4. С. 7–12.

56. ГОСТ Р ИСО 10576-1–2006. Статистические методы. Руководство по оценке соответствия установленным требованиям. Часть 1. Общие принципы.

57. ГОСТ Р 8.624–2006. ГСИ. Термометры сопротивления из платины, меди и никеля. Методика поверки.

58. Левин С.Ф. Неопределенность в узком и широком смысле результатов поверки средств измерений // Измерительная техника. 2007. № 9. С. 15–19.

59. ГОСТ Р 50779.21-2004. Статистические методы. Правила определения и методы расчета статистических характеристик по выборочным данным. Часть 1. Нормальное распределение.

60. Левин С.Ф. Обеспечение единства измерений при поверке средств измерений // Измерительная техника. 2005. № 8. С. 14–18.

61. Левин С.Ф. Чего на самом деле должны опасаться ведущие специалисты по внедрению неопределенности в отечественные измерения // Измерительная техника. 2008. № 12. С. 61–64.

62. МИ 1747–87. МУ ГСИ. Меры массы образцовые и общего назначения. Методика поверки.

63. Левин С.Ф. Проблема доверительной вероятности // Измерительная техника. 2008. № 9. С. 33–39.

64. ГОСТ 28311-89. Дозаторы медицинские лабораторные. Общие технические требования и методы испытаний.

65. ГОСТ Р 8.648–2008. ГСИ. Государственная поверочная схема для средств измерений переменного электрического напряжения до 1000 В в диапазоне частот от $1 \cdot 10^{-2}$ до $2 \cdot 10^9$ Гц.

66. ГОСТ Р 8.663-2009. ГСИ. Государственная поверочная схема для средств измерений силы.

67. ГОСТ 8.381–2009. ГСИ. Эталоны. Способы выражения точности.

68. ГОСТ ОИМЛ R 111-1–2009. ГСИ. Гири классов E_1 , E_2 , F_1 , F_2 , M_1 , M_{1-2} , M_2 , M_{2-3} , M_3 . Часть 1. Метрологические и технические требования.

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

70. ГОСТ Р 8.736–2011. ГСИ. Измерения прямые многократные. Методы обработки результатов измерений. Основные положения.

Поступила в редколлегию 25.02.2013

Рецензент: д-р техн. наук, проф. И.В. Руженцев, Харьковский национальный университет радиоэлектроники, Харьков.

ФІЛОСОФІЯ ТА МАТЕМАТИЧНА СТАТИСТИКА В МЕТРОЛОГІЇ

С.Ф. Левін

Розглянуто філософські та статистичні аспекти проблеми істинного значення в метрології. Проведено критичний аналіз термінології російських, європейських та міжнародних нормативних документів.

Ключові слова: істинне значення, толерантність, невизначеність – розподіл та параметр.

PHILOSOPHY AND MATHEMATICAL STATISTICS IN METROLOGY

S.F. Levin

Philosophical and statistical aspects of a problem of true value in metrology are considered. The critical analysis of terminology of Russian, European and international standard documents is carried out.

Keywords: true value, tolerance, uncertainty – distribution and parameter.