УДК 519.814:519.226:006.86

И.Р. Шайняк

*НИЦ контроля и диагностики технических систем, Нижний Новгород, Россия*

# О ПОСЛЕДОВАТЕЛЬНОСТИ ИЗЛОЖЕНИЯ БАЙЕСОВСКОГО ПОДХОДА В РУКОВОДСТВЕ ПО ВЫРАЖЕНИЮ НЕОПРЕДЕЛЕННОСТИ ИЗМЕРЕНИЯ (GUM)

*Рассматривается возможность байесовской интерпретации процедур, установленных Руководством по выражению неопределенности измерения (GUM), на примере расчета интервала охвата с использованием формулы Уэлча-Саттертуэйта. Показано, что данная процедура допускает байесовскую интерпретацию и может быть использована при выполнении ряда условий, включая проверку близости доверительного интервала и байесовского интервала для данной измерительной задачи. Приведен пример, иллюстрирующий возможные существенные различия между данными интервалами.*

***Ключевые слова**: неопределенность измерения, GUM, байесовский подход, доверительный интервал, интервал охвата.*

## Введение

Выход в свет Руководства по выражению неопределенности измерений (GUM) [1] означал фактическое признание невозможности использования в метрологических целях аппарата математической статистики, основанного на частотной интерпретации вероятности, и замена его байесовским выводом [2], [3]. В этом, строго говоря, и заключена «революционность» нового подхода, все остальное – лишь вытекающие с той или иной степенью очевидности следствия указанного факта.

В литературе неоднократно встречается мнение, будто в самом Руководстве (GUM) идея байесовского вывода проведена непоследовательно, и, в первую очередь, это касается тех мест Руководства, где речь идет об оценивании неопределенности измерения по типу А на основе частотного подхода (см., например, [4], [5]). Данное обстоятельство, в частности, рассматривается рабочей группой JCGM/WG 1, отвечающей за поддержку Руководства (GUM) и разработку дополняющих его документов, в качестве первой причины планируемого пересмотра GUM [6] (см. также статью [7], соавторами которой являются члены JCGM/WG 1).

В связи с изложенным встает вопрос, насколько важно, чтобы в основополагающем метрологическом руководстве строго выдерживалась непротиворечивая интерпретация вероятностных моделей метрологических задач или, другими словами, способна ли иная («неправильная») интерпретации существенным образом сказаться на результате измерения.

**Цель настоящей статьи** – во-первых, исследовать, насколько правомерны «претензии» к GUM в части непоследовательности проведения байесовской интерпретации вероятности, и, во-вторых, показать, насколько сильно могут различаться между собой результаты, полученные в рамках старого, частотного (или, как его иначе называют, классического) подхода и в рамках нового, байесовского подхода GUM.

## Основная часть

Прежде всего, отметим, что само Руководство (GUM), опубликованное двадцать лет назад, хотя и установило новый общий подход к выражению результатов измерений, но при этом, по сути, рассматривало ограниченный класс метрологических задач, описываемых моделью измерения либо линейной, либо допускающей линеаризацию без существенных потерь в точности. Только после выхода в свет Дополнений [8] и [9] концепция неопределенности измерения приобрела относительно законченный вид и стала применима как к нелинейным моделям, так и к моделям с произвольным числом выходных переменных (измеряемых величин). Однако при проверке непротиворечивости GUM естественно оставаться в рамках допущений, принятых в GUM, т.е. предполагать, что модель измерения линейна.

Далее, полагаем, что все существенные источники неопределенности оцениваются по типу А на основе многократных наблюдений. Данное ограничение является весьма сильным для практики и сужает класс метрологических задач, по-видимому, только до некоторых видов косвенных измерений, но оно оправдано в целях сопоставления байесовского оценивания и метода, изложенного в GUM.

Добавим также следующие допущения: входные величины $X_i$ линейного уравнения измерения $Y=\sum_{i=1}^{N} c_i X_i$ независимы, и характеристики случайной величины $\xi_i$, ассоциированной с $X_i$, оценивают по выборке объема $n_i$ из нормального распределения независимых повторных наблюдений величины $X_i$.

В работе [4] обращается внимание на различие следующих двух процедур.

*Процедура 1 (GUM).* Способ оценивания неопределенности по GUM, основанный на трансформировании неопределенностей, предполагает получение стандартной неопределенности u(y) для оценки математического ожидания y случайной величины η, ассоциированной с измеряемой величиной Y, в виде $u^2(y)=\sum_{i=1}^{N}c_i^2u^2(x_i)$, где в качестве стандартной неопределенности $u(x_i)$ для оценки $x_i$, полученной в виде выборочного среднего по $n_i$ наблюдениям входной величины $X_i$, берут ее стандартное отклонение $s_i = S_i/\sqrt{n_i}$, $S_i$ – выборочное стандартное отклонение. Для получения расширенной неопределенности $U(y)$ берут коэффициент охвата $k_p$, равный $(1+p)/2$ -квантили t-распределения с числом степеней свободы $\nu_{eff}$, определяемых из формулы Уэлча-Саттертуэйта (p – вероятность охвата) [1, раздел G.4].

*Процедура 2 (байесовская).* В рамках байесовского подхода распределения $\xi_i$ должны рассматриваться как апостериорные распределения, полученные на основе неинформативного априорного распределения $\xi_i$ (фактически, двух неинформативных распределений – для математического ожидания μ и дисперсии нормального распределения $\sigma_i^2$), построенной по $n_i$ наблюдениям функции правдоподобия и применения теоремы Байеса. Каждое из таких апостериорных распределений представляет собой t-распределение с $(n_i-1)$ степенями свободы, математическим ожиданием $x_i$ и стандартным отклонением $s_i\times\sqrt{(n_i-1)/(n_i-3)}$. Далее, если оставаться в рамках способа оценивания неопределенности по GUM [1, раздел 8], стандартная неопределенность u(y) должна быть получена из уравнения $u^2(y)=\sum_{i=1}^{N}c_i^2s_i^2(n_i-1)/(n_i-3)$, но выбор значения коэффициента охвата уже связан с некоторыми проблемами. Строго говоря, распределение η должно быть полностью определено распределениями $\xi_i$, поскольку эти случайные величины связаны тем же линейным уравнением $\eta=\sum_{i=1}^{N}c_i\xi_i$. Однако распределение суммы случайных величин, имеющих t-распределение каждая, но с разным числом степеней свободы, не имеет аналитического выражения. Распределение η могло бы быть получено числено с применением метода Монте-Карло (см. [8]), однако работа [4] была опубликована до появления [8], поэтому в ней предлагается распределение величины η считать нормальным, а значение коэффициента охвата брать равным двум за исключением только некоторых особых случаев.

Здесь уместно обратимся к примечанию к пункту 4.1.4 в [1]. В нем рассматривается ситуация, когда в дополнение к вышеизложенным допущениям предполагается, что число наблюдений каждой выходной величины одинаково, и $n_i = n$ для всех $i=1,...,N$. В этом случае от наблюдений входных величин можно перейти к наблюдениям выходной величины, представив каждое такое наблюдение в виде $Y_j=\sum_{i=1}^{N}c_iX_{i,j}$, $j=1,...,n$. Оценка математического ожидания y и ее стандартное отклонение u(y), рассчитанные по выборке $\{Y_j\}$, будут теми же, что рассчитанные по стандартной процедуре оценивания неопределенности по GUM. Но что более важно, в рассматриваемом случае применения *Процедуры 1* и *Процедуры 2* дают, в принципе, одинаковые результаты. В самом деле, апостериорным распределением η, построенным по выборке $\{Y_j\}$, будет t-распределение с $(n-1)$ степенями свободы, и это распределение будет совпадать с распределением случайной величины $\sum_{i=1}^{N}c_i\xi_i$, где каждая $\xi_i$ также имеет t-распределение с $(n-1)$ степенями свободы. В *Процедуре 1* рассчитанное по формуле Уэлча-Саттертуэйта эффективное число степеней свободы $\nu_{eff}$ будет равно $(n-1)$, а в *Процедуре 2* необходимо отказаться от предположения о нормальности распределения η и заменить его истинным t-распределением.

Необходимо, однако, отметить следующее. *Процедура 1* основана на работе [10], где, вообще говоря, рассматривается построение доверительного, а не байесовского интервала для Y. В случае, когда выборки берут из нормальной генеральной совокупности, доверительный интервал, основанный на статистике $(y-Y)/u(y)$, и байесовский интервал численно совпадают, и на «частотное происхождение» *Процедуры 1* можно не обращать внимания.

Естественным является желание распространить указанное решение на случай, когда $n_i$ не равны между собой. В этом случае выборка $\{Y_j\}$ отсутствует, и применить байесовский вывод непосредственно для распределения, ассоциированного с выходной величиной, невозможно. В [1, пункт G.1.4] указан принцип построения точного аналитического решения для Y в виде свертки распределений для входных величин $X_i$. Изложенный далее в том же разделе метод с использованием формулы Уэлча-Саттертуэйта (метод У-С) рассматривается как приближенное решение, реализующее следующую логическую цепочку: если все $n_i$ равны между собой, то доверительный и байесовский интервалы для Y совпадают → если не все $n_i$ равны между собой, то доверительный и байесовский интервалы можно считать близким друг другу, и построенный доверительный интервал рассматривать как приближение байесовского.

В работе [10] показано, что при несовпадающих $n_i$ статистика $(y-Y)/u(y)$, вообще говоря, зависящая от $\sigma_i^2$, имеет распределение, близкое к независящему от $\sigma_i^2$ t-распределению с $\nu_{eff}$ степенями свободы. Таким образом, *Процедура 1* может также рассматривается как байесовская, если принять, что в ней использованы два приближения:

– берется доверительный интервал в качестве интервала охвата (байесовского);

– сам доверительный интервал получен в некотором приближении.

Точность последнего приближения оценена в [10]. Данных же о том, к каким последствиям ведет замена интервала охвата доверительным интервалом, в настоящее время собрано недостаточно (см., например, [4]).

Если учесть, что *Процедура 2* тоже дает только некоторое приближенное решение (приближение связано с заменой истинного апостериорного распределения для Y нормальным), то можно сделать вывод, что эти две процедуры мало отличаются с точки зрения «идеологической чистоты». Их обе в равной степени можно считать байесовскими, а необходимость использования и в той, и в другой некоторых приближений обусловлена отсутствием на момент разработки и опубликования GUM общедоступной вычислительных средств, позволяющих рассчитывать точные апостериорные распределения для выходных величин численными методами, например, с использованием метода Монте-Карло [8].

Вместе с тем следует признать, что в стремлении придать GUM возможно более универсальный характер метод У-С без должных оснований был распространен на гораздо более широкий класс задач, чем те, для которых была получена формула Уэлча-Саттертуэйта. Причем, если для того, чтобы приписать выходной величине нормальное распределение, рекомендуется проверить выполнение условий центральной предельной теоремы, то для применимости метода У-С требуется только, чтобы число входных величин было не менее двух (хотя в [1, пункт G.5.4] и упоминается о возможных ограничениях метода). Данная проблема была решена Дополнением [8], дающим возможность получать истинные (в пределах точности численного метода) апостериорные распределения для выходной величины, после чего метод У-С в значительной степени утратил актуальность. Здесь, как и выше, под истинным распределением понимается то, что определено моделью измерения при условии известных распределений для входных величин.

## Пример

Если вернуться к использованной в *Процедуре 1* замене интервала охвата доверительным интервалом, то следует отметить необычность данной ситуации, поскольку одним из недостатков классического частотного подхода является именно сложность или невозможность построения доверительного интервала для многих метрологических задач и, напротив, простота построения байесовского интервала, требующего, правда, соответствующих вычислительных ресурсов [3].

Выше было показано, что в ряде случаев доверительный интервал и интервал охвата (байесовский) близки друг к другу, однако существуют задачи, в которых различие между ними может быть сколь угодно велико. В качестве примера рассмотрим простейшую модель измерения вида

$$Y = X, \tag{1}$$

которая часто применяется, когда доминирует один источник неопределенности, например, связанный со средством измерений. В качестве метрологической характеристики средства измерений часто указывают пределы погрешности $\pm\Delta$, и в этом случае инструментальную неопределенность принято характеризовать равномерным распределением в указанных пределах.

Повторные наблюдения величины X могут иметь место, когда одну и ту же величину Y измеряют разными средствами измерений. Примером может служить измерение потока жидкости или газа в трубе несколькими расходомерами. Тогда повторными наблюдениями будут являться показания этих расходомеров. В случае модели измерения вида (1) повторные наблюдения входной величины можно интерпретировать как повторные наблюдения выходной величины. Это соответствует, например, ситуации, когда измерения одной и той же величины проводятся разными лабораториями, и распределения, характеризующие результаты измерений, совпадают между собой с точностью до параметра положения.

Пусть число повторных наблюдений $\left\{X_j\right\}$ (или $\left\{Y_j\right\}$) равно двум, $j = 1, 2$. Поучаемая в классическом подходе оценка максимального правдоподобия y будет средним арифметическим наблюдений с соответствующим симметричным треугольным распределением на интервале $y \pm \Delta$. При доверительной вероятности p минимальный доверительный интервал будет иметь вид $y \pm \Delta\left(1 - \sqrt{1-p}\right)$. Получаемая в байесовском подходе GUM апостериорная вероятность для случайной величины η, ассоциированной с Y, будет иметь вид равномерного распределения на интервале $y \pm \left(\Delta - \left|X_1 - X_2\right|/2\right)$. Соответственно, интервал охвата при вероятности охвата p будет $y \pm p\left(\Delta - \left|X_1 - X_2\right|/2\right)$.

Как видим, в данном примере доверительный интервал и интервал охвата могут различаться очень сильно. Длина доверительного интервала постоянна и не зависит от наблюдений. В свою очередь, длина интервал охвата зависит от наблюдений и может варьироваться в пределах от 0 (с исчезающе малой вероятностью) до $p\Delta$ (с наибольшей вероятностью), т.е. быть как больше, так и много меньше длины доверительного интервала. Здесь в полной мере проявляется «философское» различие между этими двумя понятиями. Доверительный интервал характеризует точность измерения «в среднем», поэтому его длина может быть рассчитана до проведения наблюдений, в то время как байесовский интервал характеризует субъективное представление исследователя об измеряемой величине, складывающееся после проведения конкретного измерения, и поэтому он в максимальной степени чувствителен к результатам наблюдений.

Если длина байесовского интервала оказывается близкой к нулю, то имеет смысл повторно рассмотреть предположения, положенные в основу модели (1), и, возможно, учесть дополнительные источники неопределенности.

## Выводы

Внимательный анализ Руководства [1] позволяет сделать вывод о том, что предъявляемые ему упреки в непоследовательности изложения байесовской идеологии при оценивании неопределенности измерения небезосновательны. В тексте Руководства, действительно, неоднократно встречаются ссылки на частотный характер расчета стандартной неопределенности, оцениваемой по типу А (см. [1], пункты 3.3.5, 4.1.6, G.4.2). С этой точки зрения гораздо более последовательными в изложении концепции неопределенности на основе байесовского подхода являются Дополнения [8] и [9]. Однако во всех трех документах результат измерения рассматривается как субъективное суждение исследователя, и процедуры, изложенные в [1], включая метод У-С, допускают байесовскую интерпретацию.

Метод У-С по-прежнему может использоваться там, где необходимо и есть возможность получить приближенную оценку интервала охвата без привлечения численных методов расчета. Однако при этом требуется критично рассмотреть вопрос применимости этого метода, которому в [1] уделено недостаточно внимания. Существенная нелинейность модели измерения, негауссовость распределений, приписанных значимым входным величинам, коррелированность входных величин – наличие этих факторов может поставить применимость метода У-С под сомнение.

Но, главное, во многих метрологических задачах наименьший доверительный интервал (где такой интервал вообще существует) и наименьший интервал охвата (т.е. байесовский интервал, существующий всегда) могут сильно отличаться друг от друга. Это является практически значимым доводом в пользу точки зрения о несовместимости двух подходов – классического, основанного на понятии погрешности, и подхода GUM.

## Список литературы


1. *ISO/IEC Guide 98-3:2008 Uncertainty of measurement – Part 3: Guide to the expression of uncertainty in measurement (GUM:1995).*
2. *Weise, K. A Bayesian theory of measurement uncertainty / K. Weise, W. Wöger // Meas. Sci. Technol. – 1992. – Vol. 3. – P. 1-11.*
3. *Шайняк И.Р. Об изменении некоторых метрологических представлений в связи с опубликованием руководств в области неопределенности измерений / И.Р. Шайняк // Законодательная и прикладная метрология. – 2011. – № 6. – С. 14-21.*
4. *Kacker, R. N. On use of Bayesian statistics to make the Guide to the Expression of Uncertainty in Measurement consistent / R.N. Kacker, A.T. Jones // Metrologia. – 2003. – Vol. 40. – P. 235-248.*
5. *Kacker, R.N. Bayesian alternative to the ISO-GUM's use of the Welch–Satterthwaite formula / R.N. Kacker // Metrologia. – 2006. – Vol. 43. – P. 1-11.*
6. *GUM (survey report) [Электронный ресурс]. – Режим доступа к документу: http://www.bipm.org/wg/JCGM/JCGM-WG1/Allowed/sub-committee_5/WG1-SC5-N12-4b_UM_survey_report.pdf.*
7. *Bich W. Revision of the 'Guide to the Expression of Uncertainty in Measurement' / W. Bich et al. // Metrologia. – 2012. – V. 49. – P. 702-705.*
8. *ISO/IEC Guide 98-3:2008/Supplement 1:2008 Propagation of distributions using a Monte Carlo method.*
9. *ISO/IEC Guide 98-3:2008/Supplement 2:2011 Extension to any number of output quantities.*
10. *Welch, B.A. The generalization of 'Student's' problem when several different population variances are involved / B.A. Welch, // Biometrika. – 1947. – V. 34. – P. 28-35*


*Поступила в редколлегию 16.02.2013*

**Рецензент**: д-р техн. наук А.Г. Чуновкина, Всероссийский НИИ метрологии им. Д.И. Менделеева, Санкт-Петербург.

### ПРО ПОСЛІДОВНІСТЬ ВИКЛАДЕННЯ БАЙЄСОВСЬКОГО ПІДХОДУ В НАСТАНОВІ З ВИРАЖЕННЯ НЕВИЗНАЧЕННОСТІ У ВИМІРЮВАННЯХ (GUM)

І.Р. Шайняк

*Розглядається можливість баєсовської інтерпретації процедур, які встановлено Настановою з подання невизначеності вимірювань. (GUM), на прикладі розрахунку інтервалу охоплення з застосуванням формули Уелча-Саттертуейта. Показано, що дана процедура допускає баєсівську інтерпретацію та може бути використана при виконанні ряду умов, включаючи перевірку близькості довірчого інтервалу та баєсовського інтервалу для даної вимірювальної задачі. Наведено приклад, який ілюструє ймовірні суттєві різниці між даними інтервалами.*

***Ключові слова**: невизначеність вимірювань, GUM, баєсівський підхід, довірчий інтервал, інтервал охвату.*

### ON THE CONSISTENCY OF BAYESIAN APPROACH IN THE GUIDE TO THE EXPRESSION OF UNCERTAINTY IN MEASUREMENT (GUM)

I.R. Szajniak

*The possibility of the Bayesian interpretation of procedures established by the Guide to the Expression of Uncertainty in Measurement (GUM) on the example of the coverage interval calculation using the Welch-Satterthwaite formula is considered. It is shown that this procedure allows the Bayesian interpretation and can be used under certain conditions, including the verification of the confidence interval proximity and the Bayesian interval proximity for the given measurement task. An example of possible significant difference between these intervals is given.*

***Keywords**: the measurement uncertainty, GUM, the Bayesian approach, the confidence interval, the coverage interval.*