УДК 621.317

И.П. Захаров

Харьковский национальный университет внутренних дел

ВЛИЯНИЕ КОРРЕЛЯЦИИ НА ДОСТОВЕРНОСТЬ ОЦЕНИВАНИЯ НЕОПРЕДЕЛЕННОСТИ ИЗМЕРЕНИЙ

Оценена неопределенность нахождения наблюдаемого коэффициента корреляции по экспериментальным данным. Получена оценка погрешности, обусловленной полным и частичным пренебрежением корреляции при оценивании суммарной стандартной неопределенности. Проведено сравнение достоверной оценки расширенной неопределенности с оценкой, полученной по рекомендациям GUM.

коэффициент корреляции, оценка погрешности, неопределенность измерений

Введение

В настоящее время к измерениям, помимо традиционных физических дисциплин, прибегают в самых разных общественных науках, прежде всего в психологии, социологии, экономике [1]. В этих областях рамки классической концепции метрологии, в частности теории погрешности, оказываются узкими для решения задач обработки информации в условиях структурной и параметрической неопределенности.

В последнее время широкое распространение во всех областях измерений получила концепция неопределенности [2]. Разработанное семью ведущими международными организациями «Руководство по выражению неопределенности в измерениях» (GUM) [3], лежащее в основе этой концепции, является неформальным международным стандартом оценивания качества измерений. В нем даются рекомендации по нахождению суммарной стандартной неопределенности u_c с учетом наблюдаемой и логической корреляции между входными величинами. Этот момент выгодно отличает [3] от отечественных документов по оцениванию погрешности измерений [4] в которых производится учет только наблюдаемой корреляции, который ограничивается применением метода приведения, что не всегда возможно, поскольку требует проведения одинакового числа наблюдений для всех входных величин.

Следует заметить, что недостоверность оценок неопределенности входных величин, как и коэффициента корреляции $r_{i,k}$, приводят к недостоверности оценок суммарной стандартной неопределенности u_c. Именно поэтому в некоторых работах [5] учет корреляции при оценивании точности измерений считается нецелесообразным и ей либо пренебрегают, либо производят ее приближенную оценку.

Постановка задачи

Цель данного исследования состоит в сравнении влияния неопределенности оценки коэффициента корреляции и пренебрежения (или приближенного оценивания) корреляции на достоверность оценок неопределенности измерений.

Для достижения этой цели необходимо решить следующие задачи:

1) определить неопределенность оценки наблюдаемого коэффициента корреляции $u\left(r_{1,2}\right)$;

2) оценить влияние $u\left(r_{1,2}\right)$ на $u\left(u_c\right)$;

3) произвести оценку влияния пренебрежения или приближенного учета наблюдаемой или логической корреляции на $u\left(u_c\right)$;

4) сравнить результаты, полученные в результате решения задач 2 и 3;

5) получить достоверную оценку расширенной неопределенности U и сравнить ее с оценкой, получаемой при использовании рекомендаций [3].

1. Оценивание неопределенности нахождения наблюдаемого коэффициента корреляции по экспериментальным данным

Наблюдаемый коэффициент корреляции оценивается по экспериментальным данным по известной формуле [2]

$$r_{1,2}=\frac{\sum_{j=1}^{n}\left(x_{1j}-\bar{x}_1\right)\left(x_{2j}-\bar{x}_2\right)}{\sqrt{\sum_{j=1}^{n}\left(x_{1j}-\bar{x}_1\right)^2\sum_{j=1}^{n}\left(x_{2j}-\bar{x}_2\right)^2}}, \qquad (1)$$

где x_{1j}, x_{2j} - одновременно наблюдаемые значения первой и второй входных величин; $\bar{x}_1$, $\bar{x}_2$ - их средние арифметические значения.

Методом Монте-Карло [6] был произведен расчет зависимости $u\left(r_{1,2}\right)$ от значений $r_{1,2}$ в диапазоне −1…1 для числа степеней свободы ν от 1 до 30. Результаты расчета показаны на рис. 1.

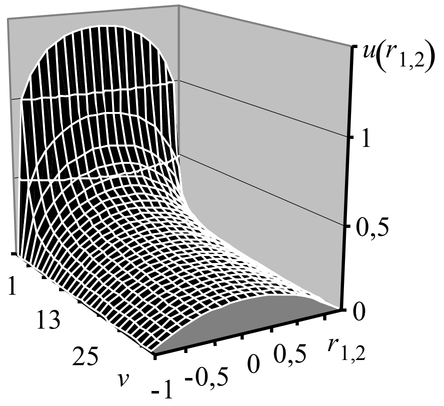

Рис. 1. Зависимость неопределенности оценки коэффициента корреляции $u(r_{1,2})$ от значения $r_{1,2}$ для числа степеней свободы ν

Анализ рис. 1 показывает, что неопределенность оценки коэффициента корреляции достигает максимума при $r_{1,2}=0$ и увеличивается с уменьшением числа степеней свободы (количества наблюдений). При $r_{1,2}=0$ зависимость $u(r_{1,2})=f(\nu)$ хорошо аппроксимируется выражением

$$u(r_{1,2}) = 1/\sqrt{\nu-0,5}\,. \tag{2}$$

При числе степеней свободы больше 10 зависимость $u(r_{1,2})$ аппроксимируется выражением

$$u(r_{1,2}) = \left(1-r_{1,2}^2\right)/\sqrt{\nu-0,5} \tag{3}$$

и ее максимальное значение не превышает 0,3.

Следует отметить, что относительная неопределенность оценки стандартного отклонения аппроксимируется выражением [7, 8]

$$\tilde{u}(u) = u(u)/u = 1/\sqrt{2(\nu\text{-}0,5)}\ , \tag{4}$$

в которое, в отличие от (2) входит постоянный множитель $1/\sqrt{2}$.

Таким образом, неопределенность нахождения коэффициента корреляции $u(r_{1,2})$ должна вносить вклад в неопределенность оценки u_c двух коррелированных величин, соизмеримый с суммарным вкладом неопределенностей оценок входных величин. Эта закономерность будет показана ниже.

2. Оценка влияние неопределенности оценивания коэффициента корреляции на достоверность нахождения суммарной стандартной неопределенности

В соответствии с законом распространения неопределенности [1] выражение для суммарной стандартной неопределенности двух коррелированных входных величин имеет вид

$$u_c = \sqrt{u_1^2+u_2^2+2r_{1,2}u_1u_2}\ , \tag{5}$$

где u_1, u_2 – вклады неопределенности первой и второй входных величин; $r_{1,2}$ – коэффициент корреляции.

Нетрудно заметить, что недостоверность оценок неопределенности входных величин, как и коэффициента корреляции, приводят к недостоверности оценок суммарной стандартной неопределенности u_c.

Так, для выражения (5), неопределенность оценки u_c будет равна

$$\begin{aligned} u(u_c) &= \\ &= \sqrt{\left[\frac{\partial u_c}{\partial u_1}u(u_1)\right]^2+\left[\frac{\partial u_c}{\partial u_2}u(u_2)\right]^2+\left[\frac{\partial u_c}{\partial r_{1,2}}u(r_{1,2})\right]^2} = \\ &= \frac{1}{u_c}\left\{\left[\left(u_1+r_{1,2}u_2\right)u(u_1)\right]^2+\left[\left(u_2+r_{1,2}u_1\right)u(u_2)\right]^2+\right. \\ &\quad \left.+\left[u_1u_2u(r_{1,2})\right]^2\right\}^{0,5}. \end{aligned} \tag{6}$$

С учетом выражений (3), (4) это выражение можно переписать для $\nu \ge 10$ в виде

$$u(u_c) = \sqrt{\frac{\left(u_1^2+u_2^2\right)^2+4r_{1,2}u_1u_2}{2\left(u_1^2+u_2^2+2r_{1,2}u_1u_2\right)(\nu-0,5)}} \tag{7}$$

При игнорировании корреляции это выражение примет вид

$$u(u_{нк}) = \sqrt{\frac{u_1^2+u_2^2}{2(\nu-0,5)}} \tag{8}$$

При этом относительная неопределенность $\tilde{u}(u_{нк})$ будет определяться выражением (4). Ее значения для $\nu=10$ и $\nu=30$ будут равны 23 % и 13 % соответственно.

Относительная неопределенность u_c, обусловленная только неопределенностью оценки коэффициента корреляции, определяется выражением:

$$\tilde{u}_к(u_c) = \frac{u_к(u_c)}{u_c} = \frac{u(r_{1,2})u_1u_2}{u_1^2+u_2^2+2r_{1,2}u_1u_2} \tag{9}$$

При $u_1 = u_2$ это выражение достигает максимума и принимает вид

$$\tilde{u}_к(u_c) = \frac{u(r_{1,2})}{2(1+r_{1,2})}. \tag{10}$$

Эта зависимость, построенная с учетом результатов, полученных в п.3 настоящей статьи, приведена на рис. 2.

Для $\nu \ge 10$ эту зависимость можно аппроксимировать выражением

$$\tilde{u}_к(u_c) = \frac{(1-r_{1,2})}{2\sqrt{\nu-0,5}}. \tag{11}$$

Анализ рис. 2 показывает, что погрешность оценивания суммарной неопределенности уменьшается при уменьшении $r_{1,2}$ и числа степеней свободы. Для снижения влияния этой составляющей погрешности при отрицательном коэффициенте корреляции следует его оценку производить при значительном количестве наблюдений (больше 30).

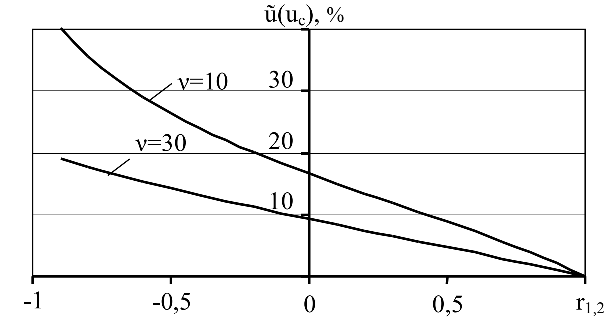

Рис. 2. Относительная неопределенность u_c в зависимости от $r_{1,2}$ для разного числа степеней свободы ν

Однако, как будет показано ниже, полное игнорирование корреляции при оценивании суммарной стандартной неопределенности приводит к более существенным погрешностям, чем погрешности, обусловленные неточностью определения коэффициента корреляции.

3. Оценка влияния пренебрежения корреляцией или приближенного учета коэффициента корреляции

При полном игнорировании корреляции погрешность оценивания суммарной стандартной неопределенности $\delta_{нк}$, очевидно, определяется как

$$\delta_{нк} = \frac{u_{нк}}{u_c} - 1 = \sqrt{\frac{1 + \left(u_2/u_1\right)^2}{1 + \left(u_2/u_1\right)^2 + 2r_{1,2}\, u_2/u_1}} - 1, \qquad 2)$$

где $u_{нк}$ – суммарная неопределенность, рассчитанная без учета корреляции входных величин.

Зависимости этой погрешности от отношения неопределенности каждой из суммируемых входных величин u_2/u_1 и коэффициента корреляции $r_{1,2}$ показана на рис. 3 пунктирными линиями. Из рис. 3 видно, что значения погрешности по модулю нарастают с увеличением u_2/u_1 и $r_{1,2}$ и достигают -30 % для положительных $r_{1,2}$ и более 100 % для отрицательных.

При $u_1 = u_2$ выражение (8) достигает максимума и принимает вид

$$\delta_к = \frac{1}{\sqrt{1 + r_{1,2}}} - 1. \qquad (13)$$

Для уменьшения этой погрешности в литературе [4] предлагается способ приближенного учета коэффициента корреляции, заключающийся в разделении погрешностей на сильно коррелированные $r_{1,2} \le -0{,}7$ и $r_{1,2} \ge 0{,}7$, для которых принимают $r_{1,2} = \mp 1{,}0$, и слабо коррелированные $-0{,}7 \le r \le 0{,}7$, корреляцию для которых не учитывают ($r_{1,2} = 0$).

Для случая приближенного учета наблюдаемой или логической корреляции можно записать выражение для погрешности оценки суммарной неопределенности в виде

$$\delta_{нк} = \begin{cases} \sqrt{\dfrac{1+\left(u_2/u_1\right)^2 + 2u_2/u_1}{1+\left(u_2/u_1\right)^2 + 2r_{1,2}\, u_2/u_1}} - 1, & r_{1,2} \ge 0{,}7; \\ \sqrt{\dfrac{1+\left(u_2/u_1\right)^2}{1+\left(u_2/u_1\right)^2 + 2r_{1,2}\, u_2/u_1}} - 1, & -0{,}7 < r_{1,2} < 0{,}7; \\ \sqrt{\dfrac{1+\left(u_2/u_1\right)^2 - 2u_2/u_1}{1+\left(u_2/u_1\right)^2 + 2r_{1,2}\, u_2/u_1}} - 1, & r_{1,2} \le -0{,}7. \end{cases} \qquad (14)$$

Эти зависимости указаны на рис. 3 сплошными линиями. Из рис.3 видно, что при положительных $r_{1,2}$ модуль погрешности уменьшается незначительно (до -25 %), а при отрицательных $r_{1,2}$ – увеличивается свыше 100 %.

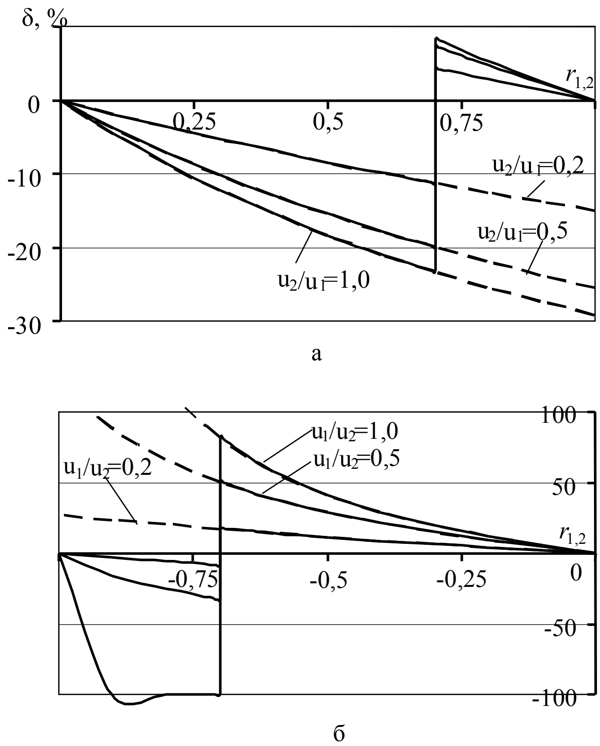

Рис. 3. Зависимости погрешности, обусловленной пренебрежением корреляцией для положительных (а) и отрицательных (б) значений коэффициента корреляции

4. Оценивание расширенной неопределенности при суммировании нескольких коррелированных составляющих

Чаще всего с попарной корреляцией можно встретиться при проведении косвенных многократных измерений. При этом суммарная стандартная неопределенность находится по формуле (5), а оценивание расширенной неопределенности в [3] четко не регламентировано.

Если для этого воспользоваться коэффициентом охвата k, определяемым как коэффициент Стьюдента с эффективным числом степеней свободы ν_{eff}, вычисляемым по формуле Велча-Саттерсвейта, то будем иметь следующие результаты. Для двух входных коррелированных величин с равным числом степеней свободы $\nu_1 = \nu_2 = \nu$ получим эффективное число степеней свободы равное

$$\nu_{eff} = \nu\, u_c^4(y) \Big/ \left[u_1^4(y) + u_2^4(y) \right]. \qquad (15)$$

Если суммарную стандартную неопределенность в этом выражении оценивать по формуле (5), то при изменении $r_{1,2}$ от -1 до 1 эффективное число степеней свободы ν_{eff} будет изменяться от 0 до 8. При $\nu_{eff} = 0$ коэффициент Стьюдента будет обращаться в бесконечность, а при $\nu_{eff} = 8\nu$ будет равен наименьшему возможному значению (для уровня доверия 0,95 k = 2) уже для $\nu = 2$.

Для нахождения достоверных значений коэффициента охвата для этого случая методом Монте-Карло осуществлялось моделирование композиции законов распределения Стьюдента для коррелированных с заданным коэффициентом корреляции $r_{1,2}$ нормально распределенных величин с известными стандартными отклонениями u_1, u_2 [9]. Результатом моделирования является зависимость коэффициента охвата получаемой композиции от числа степеней свободы ν, представленная на рис. 4.

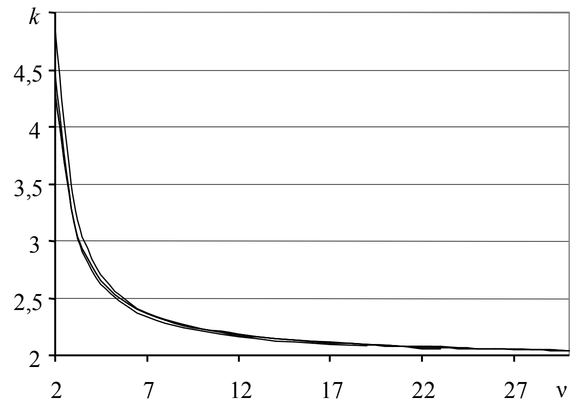

Рис. 4. Зависимость коэффициента охвата композиции законов распределения Стьюдента коррелированных входных величин от числа степеней свободы ν при разных соотношениях их неопределенностей и коэффициентов корреляций

Полученная зависимость результатов моделирования при изменении u_2/u_1 от 0 до 1 и $-1 \le r_{1,2} \le 1$ показывает слабую зависимость коэффициента охвата от коэффициента корреляции и хорошую аппроксимацию его коэффициентом Стьюдента для числа степеней свободы ν.

Эта закономерность следует из выражения для коэффициента охвата, предложенного в работе [10] при подстановке в него равных степеней свободы ν для любых соотношений u_2/u_1 и значений $r_{1,2}$:

$$k = \frac{\sqrt{k_{0,95}^2(\nu_1)u_1^2 + k_{0,95}^2(\nu_2)u_2^2 + k_{0,95}^2 2r_{1,2}u_1u_2}}{\sqrt{u_1^2 + u_2^2 + 2r_{1,2}u_1u_2}}. \qquad (16)$$

Исследование погрешности аппроксимации полученной зависимостью показывает, что она находится в диапазоне от $-1{,}5\ldots+5$ % для $\nu \ge 3$.

Выводы

Таким образом, на основании результатов, полученных в статье можно сделать вывод, что пренебрежение или приближенный учет коэффициента корреляции приводит к существенной недостоверности оценивания суммарной неопределенности. При оценивании расширенной неопределенности рекомендации (GUM) дают недостоверную оценку при наличии корреляции, а наилучшим аппроксимирующим выражением для коэффициента охвата композиции как некоррелированных, так и коррелированных законов распределения Стьюдента, является выражение (16).

Список литературы

1. *Карел Берка. Измерения. Понятия, теории, проблемы. – М.: Прогресс, 1987. – 320 с.*

2. *Захаров И.П. Теория неопределенности в измерениях. – Х.: Консум, 2002. – 256 с.*

3. *Guide to the Expression of Uncertainty in Measurement. – ISO, Geneva, First Edition. – 1995. – 101 p.*

4. *МИ 2083-90 ГСИ. Измерения косвенные. Определение результатов измерений и оценивание их погрешностей. – М.: Изд-во стандартов, 1990. – 9 с.*

5. *Новицкий П.В., Зограф И.А. Оценка погрешностей результатов измерений. – Л.: Энергоатомиздат. Ленингр. отд-ние, 1991. – 304 с.*

6. *Соболь И.М. Метод Монте-Карло. 4-е изд. доп. и перераб. – М.: Наука, 1985. – 78 с.*

7. *Рабинович С.Г. Погрешности измерений. – Л.: Энергия, 1978. – 261 с.*

8. *Грановский В.А., Сирая Т.Н. Методы обработки экспериментальных данных при измерениях. – Л.: Энергоатомиздат, Ленинградское отд-ние, 1990. – 288 с.*

9. *Захаров И.П. Композиция законов распределения Стьюдента // Системи обробки інформації. – Х.: ХУ ПС, 2005. – Вип. 8 (38). – С. 28-35.*

10. *Захаров И.П. Учет корреляции при оценивании неопределенности результатов многократных измерений // Системи обробки інформації. – Х.: ХУ ПС, 2005. – Вип. 9 (39). – С. 43-45..*

Поступила в редколлегию 16.01.2007

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