УДК 519.6

ПОГРЕШНОСТЬ ОТ НЕАДЕКВАТНОСТИ МАТЕМАТИЧЕСКОЙ МОДЕЛИ ПРОБЛЕМНОЙ ОБЛАСТИ ПРИ СТАТИСТИЧЕСКИХ ИЗМЕРЕНИЯХ

Мирошниченко И.В.
Национальный технический университет Украины «Киевский политехнический институт имени Игоря Сикорского» (Украина)

Суммарная погрешность статистических измерений, зависящая от физического смысла вычисляемых оценок вероятностных характеристик, конечности объема выборки результатов измерений и взаимодействия исследуемой проблемной области с другими, в отличие от детерминированных измерений, всегда содержит погрешность неадекватности от несоответствия проблемной области приписываемой ей математической модели.

Для статических измерениях необходимо выполнение таких операций как, преобразование, которое лежат в основе алгоритмов обработки результатов и процедуры усреднения случайного объёма выборки. Сигналы и помехи, которые образуются в процессе исследований, имеют вероятностную составляющую. Наличие этой составляющей дает возможность применения статистических методов для обработки получаемого в процессе исследования результата. Соответствие процесса и его вероятностной модели сводится к проверке гипотез о типе распределении, однородности данных и т.д. Наличие отрицательного результата на каждом шаге проверки не предоставляет возможности исследования по стандартной технологии. Процесс проверки пригодности каждой конкретной математической модели и физического объекта, превращается в еще одно дополнительное исследование.

Наиболее совершенными являются многоканальные и многофункциональные системы обработки экспериментальных данных (СОЭД) с интегрированными программой управляемыми адаптивными измерительными каналами (ПУАИК). СОЭД реализуют не только физический процесс измерений, но и производят анализ результатов этих измерений, их сопоставление (или принятие решения) и представление оператору результаты обработки измерений.

Ключевые слова: статистические измерения, математическая модель, погрешность классификации, теоретические вероятностные характеристики и их оценки.

Постановка проблемы. Основу исследований в различных проблемных областях (PRAR – Problem area): изучении свойств сред (воды, атмосферы, плазмы); системах связи, навигации и управления; диагностике (технических систем, в биологии и медицине) и других областях науки и техники, составляют результаты процедуры "статистических измерений" различных параметров этих PRAR в активном и пассивном режимах [2, 6, 7, 9].

Статистические измерения необходимы даже в активном режиме при работе с детерминированными зондирующими сигналами $x(t)$, так как принятые сигналы (радио и гидролокация, лазерные системы, спутниковые системы навигации) являются смесью отраженного $x(t)$, подвергшегося воздействию среды (затухание сигнала в среде PRAR), и помех $n(t)$ от воздействия сигналов $x(t)$ на PRAR (отражения от местных предметов в радиолокации, реверберационная помеха в гидролокации, кавитация и другие эффекты).

Анализ последних исследований и публикаций. Пассивные методы вообще предполагают работу только с вероятностными математическими моделями (ММ) сигналов и помех [9].

Вероятностный характер сигналов и помех и представление их ММ в виде Θ – теоретических вероятностных характеристик (ВХ), дают основания для применения процедуры статистических измерений, под которыми будем понимать измерения (прямые, косвенные, совокупные и совместные) параметров PRAR, создание баз данных результатов этих измерений в виде массивов $\{*\}$ и последующей обработки результатов массива $\{*\}$ по заданному алгоритму [6].

Сигнал, для описания которого используется случайная функция $\xi(t,\omega)$ двух переменных: $t \in T$ и элементарного события $\omega \in \Omega$ всего множества событий с заданной на нем вероятностью, называется стохастическим. Сечение стохастического сигнала на любом t является случайной величиной (СВ), заданной на основной вероятностном пространстве (ξ, P, F), где F – алгебра событий, $\xi(t,\omega)$ – семейство СВ.

Если: t – время, то $\xi(t,\omega)$ называют случайным процессом $\xi(t)$;

-P – счетное множество целых чисел, то $\xi(t,\omega)$ называют случайной последовательностью $\xi(t_i)$, счетномерным вектором или процессом с дискретным временем;

-T – вещественный интервал, то говорят о процессе с непрерывным временем [6].

Теоретически достаточно полное описание $\xi(t)$ производится с помощью их Θ: функций распределения вероятностей, моментных функций, характеристических функций, кумулянтов и т.д. [5].

Поскольку статистические измерения в качестве объекта исследований предполагают реальный физический процесс, то наиболее естественной формой представления этого процесса является использование реализаций $X_i(t)$ – последовательных зависимостей мгновенных значений $x_i(t_j)$ от времени, или последовательностей $X(t_i)$ – совокупность мгновенных значений $x_i(t_j)$ в один и тот же момент t или совокупность значений СВ, соответствующих временному сечению t_j.

Для реализаций $X(t)$ обязательно предполагается существование распределения вероятностей в пространстве аргумента t, причем $X(t)$ рассматриваются как выборки из некоторых бесконечных генеральных совокупностей СВ. При этом предполагается, что вместо одной, возможно очень сложной ММ, может быть предложена последовательность менее сложных (более приближенных), но тоже вероятностных моделей [6].

С статистическими измерениями связана значительная часть приложений, при этом возможна их интерпретация для других $\xi(t,\omega)$, например, случайных полей $\xi(l,\omega)$ (градиентов напряженностей, концентраций веществ и растворов, полей шероховатостей и волнистости), где в качестве аргумента может выступать пространственная координата l.

Адекватность ММ, базирующейся на массивах мгновенных значений (реализациях или последовательностях) может иметь место только при бесконечно большом объеме данных, что выражается в понятиях бесконечно большого ансамбля реализаций и реализаций бесконечной длительности.

Адекватная ММ случайного процесса $\xi(t)$ может быть представлена в виде массива $\{*\}$ данных о мгновенных значениях $x_i(t_j)$:

– *по реализациям* $X_i(t) = \{x_i(t)\}$,

причем $i=\overline{1,\infty}, t\in[0,\infty] \vee i=\overline{1,\infty}, t\in[t_1t_2] \vee i\in I, t\in[0,\infty]$;

– *по совокупности последовательностей* $X(t_i) = \{x(t_i)\}$,

если $i=\overline{1,\infty}, j=\overline{1,\infty} \vee i=\overline{1,\infty}, j\in\{j\} \vee i=\{i\}, j=\overline{1,\infty}$.

Так как вся рассматриваемая ограниченная совокупность массива $\{*\}$ представляется реальными объектами, а теория статистических измерений оперирует действительными функциями, чаще всего одномерными, то результатом статистических измерений является результат вычисления значения Θ по ограниченной совокупности выборочных данных объема d, называемый ее статистической характеристикой (СХ) или оценкой Θ^*.

Оценка Θ^* является количественным значением Θ –

многомерной теоретической вероятностной характеристики ММ исследуемой PRAR, полученным по алгоритму:

$\langle g[x(t)]\rangle)$ $\Theta^*[x(t)] = L_1[x_i(t)_1^n]$ или $\Theta^*[x(t)] = L_2[x(t_i)n_1^n]$,

где: L_1 и L_2 – операторы преобразования массива $\{*\}$ в массив $[*]$ для n выборочных значений реализаций или последовательностей.

Формулирование целей статьи. Рассмотреть погрешность от неадекватности математической модели проблемной области при статистических измерениях.

Основная часть. При представлении ММ случайного процесса $\xi(t)$ ансамблем реализаций его теоретические Θ выражаются через пределы выборочных средних, а так как при этом используются две независимые переменные – время t и номер реализации N, то получаются 3 предела выборочных средних:

$$\Theta[x(t)] = \lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N} g[x_i(t)] = \lim_{T\to\infty}\frac{1}{T}\int_0^T g[x_i(t\ )]dt = \lim_{T\to\infty}\frac{1}{T}\int_0^T g[x_i(t\ )]dt\ . \qquad (1)$$

Аналогично, при представлении $\xi(t)$ набором последовательностей:

$$\Theta[x(t)] = \lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N} g[x_i(t_j)] = \lim_{N\to\infty}\frac{1}{N}\sum_{j=1}^{N} g[x_i(t_j)] = \lim_{N_1\to\infty, N_2\to\infty}\frac{1}{N_1 N_2}\sum_{i=1}^{N}\sum_{j=1}^{N} g[x_i(t_j)]. \qquad (2)$$

Выражения (1) и (2) адекватны существу процесса “статистических измерений” и позволяют перейти от теоретических значений $\Theta[x(t)]$, полученных на основе анализа ММ $\xi(t)$ и принимаемых в статистических измерениях за образцовые (“истинные”) значения, к их оценкам $\Theta^*[x(t)]$ по результатам вычислений по алгоритму $\langle g[x(t)]\rangle$ параметров PRAR (реальных физических объектов, сигналов), описание которых входило в ММ. Поэтому значения $\Theta^*[x(t)]$ рассматриваются только как возможные оценки соответствующих вероятностей, вероятностных параметров или числовых характеристик $\xi(t)$ [6].

Принципиальные ограничения достижимой точности вычислений Θ^* по результатам измерений определяются общими и специфическими группами факторов. Общие ограничения обусловлены уровнем развития теории [1, 9] и физическим смыслом Θ^*, а также объективным взаимодействием исследуемой PRAR с другими. К специфическим ограничениям можно отнести конечность объема d выборки мгновенных значений $\xi(t)$, полученных в процессе статистических измерений.

Объединив (1) и (2) в общее выражение получим:

$$\Theta^*[x(t)] = \lim Sd[g[x(t)]], \qquad (3)$$

где: $d = N \vee T \vee NT$ – параметр, характеризующий объем данных $\{*\}$,

$Sd = \frac{1}{N}\sum_{i=1}^{N} \vee \frac{1}{T}\int_{0}^{T} \vee \frac{1}{NT}\sum_{i=1}^{N}$ – оператор усреднения и, исключив практически нереализуемые ММ, можно считать, что в процессе статистических измерений получается оценка:

$$\Theta^*\left[x(t)\right] = Sd\left[g\left[x(t)\right]\right] \neq \Theta\left[x(t)\right] \text{ теоретической ВХ.} \qquad (4)$$

При этом из (4) следует, что $\Theta^*\left[x(t)\right]$ всегда содержит $\Delta_H\left[\Theta\left[x(t)\right]\right]$ – погрешность от неадекватности реальной PRAR и приписываемой ей ММ, называемая также погрешностью классификации.

Положения общей теории измерений [4] об объективной невозможности экспериментального установления “истинного” значения физической величины полностью приложимо к задачам статистических измерений, которые можно рассматривать как расширение классических детерминированных измерений по виду алгоритма измерений $\left\langle g\left[x(t)\right]\right\rangle$ и по соотношению вычисляемых по результатам измерений $\Theta^*\left[x(t)\right]$ и теоретических значений Θ, принимаемых за “истинные” (далее используется термин “образцовые” вместо “истинные”).

Познавательная функция является специфической именно для такой технической процедуры, которой является измерение и отражается в формулировке задачи определения измеряемой величины. В современной интерпретации познавательная функция формулируется как обратная задача и является некорректной.

Имеется достаточно развитая теория решения обратных задач в условиях наличия априорной информации [1,6], одним из основных приемов которой можно считать использование методов и алгоритмов, устойчивых к наличию некоторой априорной неопределенности, при увеличении которой решение не теряет устойчивости, но становится менее точным. Примерами могут быть адаптивные алгоритмы оценивания, гарантирующие обеспечение с заданной вероятностью устойчивость фильтрации в условиях параметрической неопределенности задания модели PRAR.

Погрешность классификации $\Delta_H\left[\Theta\left[x(t)\right]\right]$, определяемая несоответствием реальной PRAR приписываемой ему ММ, в большинстве технических приложений, например, при последовательном, т.е. невоспроизводимом во времени анализе серии измерений, зависит от многих частично или полностью неконтролируемых факторов. При этом представления о ММ в виде $\xi(t)$ не предполагают объективного суждения о виде закона распределения вероятностей этой ММ, а подразумевают только его неповторимость или непредсказуемость [7].

Отсюда следует, что представления о сходимости $\Theta^*\left[x(t)\right]$ к

$\Theta[x(t)]$ для некоторых $x(t)$ и $n(t)$ могут не подтверждаться. Поэтому при статистических измерениях особое внимание должно быть уделено *несмещенности, состоятельности* и *эффективности* оценок Θ^*, их *сходимости* и другим вопросам, традиционным для математической статистики [1, 7].

Отделение процедуры принятия решений от задачи вычисления Θ^* по результатам измерений, в общем случае не всегда обязательное, дает следующие практические преимущества:

– становится возможным аппаратурный статистический анализ при отсутствии многих априорных сведений о $\xi(t)$, кроме общепринятых при измерениях детерминированных сигналов (например, динамический диапазон амплитуд, частотные границы и т.д.), что отвечает практике и перспективам использования СОЭД с интегрированными ПУАИК для вычислений Θ^*;

– не требуется оправдания применения (или строгих доказательств) таких Θ^*, которые связаны со смыслом прикладных задач, например, характеристики выбросов над порогом и другие, но не имеют аналогов в теории вероятностей и могут быть даже смещенными, несостоятельными и неэффективными [2];

– метрологическая аттестация и поверка средств статистических измерений может основываться на общепринятых правилах [4].

Процедура получения вероятностных выводов (процедура принятия решений) по результатам статистических измерений почти всегда включает элемент субъективизма (и даже неосмотрительности), что часто приводит к необратимой утрате исходных данных и, как следствие, невозможность сделать последующие достоверные выводы. Применение принципа комплексирования статистических измерений может служить одним из путей уменьшения априорной неопределенности ММ о PRAR [3].

Выводы. Для многих проблемных областей процедура проверки соответствия процесса приписываемой ему математической модели обычно стоит вне измерений и сводится, чаще всего, к проверке статистических гипотез о типе распределения, однородности выборки и т.п. Ввиду того, что на каждом этапе последовательного выдвижения и проверки гипотез может быть получен отрицательный результат, процедура проверки не может быть представлена в виде стандартной методики и именно поэтому вопрос о пригодности вероятностной математической модели в каждом конкретном случае превращается в самостоятельный этап научного исследования [5, 6, 9].

Литература

1. Вентцель Е.С. Теория вероятностей. М.: Наука, 1969. 576 с.
2. Виленкин С.Я. Статистическая обработка результатов исследования случайных функций. М.: Энергия, 1979. 320 с.

3. Детлінг В.С., Мірошниченко І.В., Павленко В.І., Тихоход В.О. Вибір параметрів адаптивних систем обробки експериментальних даних. *Адаптивні системи автоматичного управління: міжвідомчий науково-технічний збірник*. Київ: Національний технічний університет України "Київський політехнічний інститут". 2012. Вип. 20(40). С.41-51.
4. Дороговець М., Мотало В., Стадник Б. Основи метрології та вимірювальної техніки. У двох томах. Том 1. Основи метрології, Львів, вид. "Львівська політехніка, 2005. 650 с.
5. Малахов А.Н. Кумулянтный анализ негауссовых процессов и их преобразований. М.: Советское радио, 1978. 376 с.
6. Цветков Э.И. Основы теории статистических измерений. 2-е изд., перераб. и дополн. Л.: Энергоатомиздат, 1986. 286 с.
7. Bendat Y. Mathematical analyses of average response values for non-stationary data. IEEE Trans. Jun. 1964. p. 47-52.
8. Detling V.S., Kartunov C., Miroshnichenko J.V. Information-logical model error of random statistical characteristics. International Scientific Conference, Gabrovo, 23-24 Nov. 2007. P. 322-327.
9. Rise S.O. Filtered thermal noise fluctuation of energy as a function of interval length. JASA, 1943. Vol.14, №4. P. 216-227.

ПОХИБКА ВІД НЕАДЕКВАТНОСТІ МАТЕМАТИЧНОЇ МОДЕЛІ ПРОБЛЕМНОЇ ОБЛАСТІ ПРИ СТАТИСТИЧНИХ ВИБІРКАХ

Мірошниченко І.В.

Загальна похибка статистичних замірів, залежить від фізичного сенсу обчислюваних оцінок імовірнісних характеристик, кінцівки об'єму вибірки результатів вибірки і взаємодії досліджуваної проблемної області з іншими, на відміну від детермінованих вибірок, завжди містить погрішність неадекватності від невідповідності проблемної області приписуваної їй математичної моделі. Для статичних вимірах потрібне виконання таких операцій як, перетворення, яке лежать в основі алгоритмів обробки результатів і процедури усереднювання випадкового об'єму вибірки. Сигнали і перешкоди, які утворюються в процесі досліджень, мають імовірнісну складову. Наявність цієї складової дає можливість застосування статистичних методів для обробки отримуваного в процесі дослідження результату. Відповідність процесу і його імовірнісної моделі зводиться до перевірки гіпотез про тип розподілу, однорідності даних і так далі. Наявність негативного результату на кожному кроці перевірки не надає можливості дослідження за стандартною технологією. Процес перевірки придатності кожної конкретної математичної моделі і фізичного об'єкту, перетворюється на ще одно додаткове дослідження. Найбільш

досконалими є багатоканальні і багатофункціональні системи обробки експериментальних даних (СОЕД) з інтегрованими програмою керованими адаптивними вимірювальними каналами (ПУАВК). СОЕД реалізують не лише фізичний процес вимірів, але і роблять аналіз результатів цих вимірів, їх зіставлення (чи ухвалення рішення) і представлення операторові результати обробки вимірів.

Ключові слова: статистичні виміри, математична модель, погрішність класифікації, теоретичні імовірнісні характеристики і їх оцінки.

ERROR FROM INADEQUACY OF MATHEMATICAL MODEL TO PROBLEM AREA AT STATISTICAL MEASURING

Miroshnichenko I.

The total error of the statistical measuring, depending on physical sense of the calculated estimations of probabilistic descriptions, extremity of sample of results of measuring and cooperation of the investigated problem area size with other, unlike the determined measuring, always contains the error of inadequacy from disparity of problem area of the mathematical model added to her. For static measuring implementation of such operations is needed as, transformation that are the basis of algorithms of treatment of results and procedure of усреднения of casual sample size. Signals and hindrances, that appear in the process of researches, have a probabilistic constituent. The presence of this constituent gives an opportunity of application of statistical methods for treatment of the result got in the process of research. Accordance of process and his probabilistic model is taken to verification of hypotheses about a type distribution, homogeneity of data etc. The presence of negative result on every step of verification does not give research possibility on standard technology. Process of verification of fitness of every certain mathematical model and physical object, grows into another additional research. Most perfect are the multichannel and multifunction systems of processing of experimental data (СОЭД) with integrated by the program by the guided adaptive measuring channels (ПУАИК). СОЭД will realize not only the physical process of measuring but also the results of treatment of measuring produce the analysis of results of these measuring, their comparison (or decision-making) and presentation to the operator.

Keywords: statistical measuring, mathematical model, error of classification, theoretical probabilistic descriptions and their estimations.