Если же восстанавливающая функция является полиномом первой степени, то для первого алгоритма и его модификации можно принять $\Delta\varphi = \frac{3}{4}\Delta g$. При использовании второго алгоритма целесообразен выбор $\Delta\varphi = \Delta g$.

Библиографический список

I. Булатов В.А., Пшеничников В.В., Сабило В.П. Эффективность восстановления недостоверных значений цифровых измерительных сигналов //Автоматизация научных исследований : Куйбыш. авиац. ин-т. Куйбышев, 1988.

2. J.E. Medlin. The computer-evaluation of wild-point rejection for telemetry date compressors.- Proc. National Telemetering Conference, USA, 1969, pp. 162 to 170.

УДК 681.518.3:519.27

В.С.Соболев, Л.А.Мокрушин

МЕТРОЛОГИЧЕСКОЕ СОПРОВОЖДЕНИЕ РЕЗУЛЬТАТОВ СТАТИСТИЧЕСКОЙ ОБРАБОТКИ ИЗМЕРИТЕЛЬНОЙ ИНФОРМАЦИИ В АСНИ

(г. Ленинград)

В настоящее время в автоматизированных системах научных исследований (АСНИ) используются все более сложные алгоритмы обработки измерительной информации. Эти алгоритмы, реализуемые обычно в виде программ, выполняемых вычислительными компонентами АСНИ, фактически составляют часть измерительной процедуры и, следовательно, должны удовлетворять требованиям Государственной системы обеспечения единства измерений (ГСИ). В частности, такие программы (подпрограммы) обработки сигналов измерительной информации, как того требует ГОСТ 26.203-81, должны сопровождаться оценкой точности получаемых результатов в установленной форме /I/.

На пути реализации указанного требования встают по крайней мере две проблемы, характерные для АСНИ и не позволяющие в ряде случаев получить априори достоверные оценки характеристик погрешностей результатов обработки измерительной информации:

отсутствие достаточной априорной информации о вероятностно-статистической модели исследуемого объекта или процесса;

использование в АСНИ адаптивных, итеративных, статистических и других достаточно сложных алгоритмов измерений, в том числе алгоритмов, содержащих процедуры принятия решений по текущей измерительной информации до получения окончательного результата измерений(обработки) /2/.

В связи с этим встает актуальная задача разработки методологии, алгоритмического и программного обеспечения метрологического сопровождения результатов обработки измерительной информации в АСНИ,т.е. алгоритмов и программ, которые позволяли бы при неточно известной модели объекта или процесса одновременно с результатом измерений (обработки) получать оценки характеристик его погрешностей, например в виде среднеквадратического отклонения (СКО), смещения, доверительных интервалов и т.п.

Цель данной работы - наметить путь решения указанной задачи применительно к алгоритмам статистической обработки измерительной информации, хотя предлагаемый подход,естественно,распространяется и на другие виды "сложных" измерений и обработки.

В основе предлагаемого подхода лежат: метод складного ножа М.Кенуя-Дж.Тьюки-Р.Миллера /3/, бутстреп-метод Б.Эфрона /4/ и другие родственные им методы управления выборкой при обработке данных.

Наиболее просто основная идея этих методов иллюстрируется при рассмотрении задачи, где не нужен, в принципе, ни один их этих методов, - оценивание СКО выборочного среднего /4/.

Пусть данные, являющиеся результатами прямых измерений, представляют собой независимую одинаково распределенную (нор) выборку объема n из неизвестного распределения вероятностей на действительной прямой:

$$x_1, x_2, \ldots, x_n \overset{нор}{\sim} F. \qquad (1)$$

По значениям x_i мы вычисляем выборочное среднее $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$, которое используется в качестве оценки математического ожидания(МО) соответствующей случайной величины (процесса).

Интересно и важно с метрологической точки зрения, что данные (I) позволяют не только построить оценку МО, но и оценить ее погрешность в терминах СКО выборочного среднего $\bar{x}$, а именно:

$$\hat{\sigma}(\bar{x}) = \left[\frac{1}{n(n-1)} \sum_{i=1}^{n} (x_i - \bar{x})^2\right]^{\frac{1}{2}}. \qquad (2)$$

Трудность состоит в том, что формула (2) не допускает очевидных обобщений на отличные от $\bar{x}$ статистики, например на выборочную медиану или ковариационную функцию. Метод складного ножа и бутстреп – это два способа проведения такого обобщения.

Пусть

$$\tilde{x}_{(i)} = \frac{n\bar{x} - x_i}{n-1} = \frac{1}{n-1} \sum_{j \neq i} x_j \qquad (3)$$

выборочное среднее данных (I), из которых исключена i-я точка. Пусть также $\bar{x}_{(\cdot)} = \frac{1}{n} \sum_{i=1}^{n} \tilde{x}_{(i)}$ – среднее арифметическое средних (3).

Очевидно, что в данном случае $\bar{x}_{(\cdot)} = \bar{x}$, но удобно сохранить введенные выше обозначения. Оценка СКО методом складного ножа есть

$$\hat{\sigma}_{Jack}(\bar{x}) = \left[\frac{n-1}{n} \sum_{i=1}^{n} (\tilde{x}_{(i)} - \bar{x}_{(\cdot)})^2\right]^{\frac{1}{2}}. \qquad (4)$$

Нетрудно проверить, что эта величина в точности совпадает с оценкой СКО по формуле (2). Преимущество формулы (4) в том, что она легко обобщается на любую статистику $\hat{\theta} = \hat{\theta}(x_1, x_2, \ldots, x_n)$, являющуюся результатом произвольного алгоритма статистической обработки данных (I). Отличие состоит лишь в подстановке $\hat{\theta}_{(i)} = \hat{\theta}(x_1, x_2, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n)$ вместо $\tilde{x}_{(i)}$ и $\hat{\theta}_{(\cdot)} = \frac{1}{n}\sum_{i=1}^{n} \hat{\theta}(i)$ вместо $\bar{x}_{(\cdot)}$.

С помощью бутстрепа оценка СКО обобщается несколько иначе. Пусть $\hat{F}$ – эмпирическое распределение данных (I), приписывающее массу $1/n$ каждому x_i, и пусть имеется гипотетическая нор-выборка из распределения $\hat{F}$:

$$X_1^*, X_2^*, \ldots, X_n^* \overset{\text{нор}}{\sim} \hat{F}. \qquad (5)$$

Другими словами, X_i^* – случайные величины, извлеченные с возвращением из наблюдаемых данных (I), причем так, что вероятность извлечь x_i при каждом выборе равна $1/n$ для любого i = I, 2,..., n.

Выборка (5) называется бутстреп-выборкой, а ее выборочное среднее

$$\bar{X}^* = \frac{1}{n}\sum_{i=1}^{n} X_i^* \tag{6}$$

бутстреп-реализацией выборочного среднего. Оценки характеристик этой величины и предлагаются в качестве оценок характеристик выборочного среднего в бутстреп-методе. В данном случае их можно вычислить аналитически. В частности, дисперсия $\bar{X}^*$ при описанной схеме выбора – это просто дисперсия величины (6) при распределении (5):

$$D_{\hat{F}}(\bar{x}^*) = \frac{1}{n^2}\sum_{i=1}^{n} D_{\hat{F}}(X_i^*) = \frac{1}{n^2}\sum_{i=1}^{n}(x_i - \bar{x})^2. \tag{7}$$

Бутстреп-оценка СКО статистики $\bar{x}$ при этом

$$\hat{\sigma}_{BOOT}(\bar{x}) = \left[D_{\hat{F}}(\bar{X}^*)\right]^{\frac{1}{2}} = \frac{1}{n}\left[\sum_{i=1}^{n}(x_i - \bar{x})^2\right]^{\frac{1}{2}}. \tag{8}$$

Она допускает простое обобщение для произвольной статистики $\hat{\theta} = \hat{\theta}(x_1, x_2, \ldots, x_n)$:

$$\hat{\sigma}_{BOOT}(\hat{\theta}) = \left[D_{\hat{F}}\hat{\theta}(X_1^*, X_2^*, \ldots, X_n^*)\right]^{\frac{1}{2}}. \tag{9}$$

Возможные варианты использования формулы (9) в общем случае связаны с двумя обстоятельствами: какую оценку $\hat{F}$ мы возьмем и какой способ оценки дисперсии по выбранной мере $\hat{F}$ мы примем. В рассмотренном примере нам удалось вычислить ее аналитически, однако, как только алгоритм вычисления оценки $\hat{\theta}$ становится более сложным, приходится прибегать к другому аппарату. В качестве такого аппарата может использоваться метод статистического моделирования Монте-Карло, позволяющий получить независимые реализации бутстреп-выборок $(X_{1j}^*, X_{2j}^*, \ldots, X_{nj}^*)$; $j = 1, 2, \ldots, B$; требуемое число раз B.

Для каждой такой реализации вычисляется бутстреп-реализация соответствующей статистики

$$\hat{\theta}_j(x_{1j}^*, x_{2j}^*, \ldots, x_{nj}^*); \ j=1,2,\ldots,B; \tag{10}$$

по исходному алгоритму статистической обработки. B -выборка бутстреп-реализаций (10) используется для получения выборочной бутстреп-оценки СКО данной статистики, например по формуле

$$\hat{\sigma}_{BOOT}(\hat{\theta}) = \Big[\frac{1}{B}\sum_{j=1}^{B}(\hat{\theta}_j - \hat{\theta}_{(\cdot)})^2\Big]^{\frac{1}{2}}, \tag{11}$$

где $\hat{\theta}_{(\cdot)} = \frac{1}{B}\sum_{j=1}^{B}\hat{\theta}_j$.

Заметим, что указанный механизм и получаемая с его помощью бутстреп-выборка статистик (10) могут использоваться не только для получения оценки СКО (11), но и других характеристик погрешности искомой статистики $\hat{\theta}(x_1, x_2, \ldots, x_n)$: смещение, полный средний квадрат погрешности, доверительные интервалы и т.д.

В некоторых случаях, однако, этот механизм оказывается не нужным, если удается получить аналитическое выражение для $\hat{\sigma}_{BOOT}(\hat{\theta})$ или других характеристик погрешности через элементы исходной выборки (1), как это было в рассмотренном выше примере (8). Это так называемый аналитический бутстреп, вообще говоря, давно применяемый в практике статистических измерений и заключающийся в замене неизвестных параметров распределения, входящих в выражение характеристик погрешностей оценивания, их выборочными оценками, получаемыми по той же выборке. В этом смысле описанный выше вариант бутстрепа, основанный на методе статистического моделирования, является просто изящным вычислительным приемом, позволяющим распространить идею аналитического бутстрепа на те сложные алгоритмы статистической обработки, где получить аналитику не удается.

Основное допущение бутстрепа, как следует из рассмотренного, заключается в том, что данные, являющиеся исходным материалом для статистической обработки, распределены так, как они распределены, а именно в соответствии со своей эмпирической функцией распределения $\hat{F}$. Такое допущение при отсутствии дополнительной априорной информации о вероятностно-статистической модели данных представляется вполне обоснованным.

С другой стороны, ничто не противоречит применению бутстрепа и когда имеется априорная информация о виде закона распределения исходных данных (1). Например, если известно, что выборка (1) извлечена из нормального распределения, то реализации бутстреп-выборки (5) могут моделироваться как выборки из нормального распределения с параметрами, оцененными по выборке (I). Это так называемый параметрический бутстреп /4/.

Важно отметить, что при применении бутстрепа погрешность оценки произвольной вероятностной характеристики (ВХ)

$$\Delta\hat{\theta}(X) = \hat{\theta}(X) - \theta(F) \tag{I2}$$

ассоциируется с бутстреп-погрешностью

$$\Delta\hat{\theta}(X^*) = \hat{\theta}(X^*) - \theta(\hat{F}), \tag{I3}$$

характеристики которой в действительности и вычисляются с помощью бутстреп-метода.

Здесь X – исходная выборка данных (I);

X^* – бутстреп-выборка (5);

$\hat{\theta}(X)$ – оценка ВХ по исходному алгоритму;

$\theta(F)$ – истинное значение оцениваемой ВХ при истинной модели F распределения;

$\hat{\theta}(X^*)$ – бутстреп-реализация оценки ВХ;

$\hat{\theta}(\hat{F})$ – значение оцениваемой ВХ, если бы истинное распределение F точно совпадало с эмпирическим $\hat{F}$.

Корректность оценки характеристик погрешности (I2) с помощью оценки характеристик погрешности (I3) является фундаментальным вопросом корректности бутстреп-метода, положительный ответ на который дает практика статистической обработки реальных экспериментальных данных /4/.

В реальной ситуации отсчеты исходной выборки (I) содержат погрешности, обусловленные неидеальным преобразованием измерительной информации в измерительных каналах АСНИ, а сами оценки ВХ и бутстреп-оценки их погрешностей подвержены определенному влиянию ошибок округления в вычислительных компонентах АСНИ. Кроме того, дополнительные флуктуации может вносить статистическое моделирование по методу Монте-Карло. Таким образом, оставляя за пределами данной работы формализацию и исследование структуры полной погрешности, оцениваемой с помощью бутстрепа, отметим, что она представляет собой сложную композицию составляющих, обусловленных следующими причинами:

неадекватность алгоритма оценивания, или неадекватность формулы, по которой вычисляется статистика;

конечность объема исходной выборки (I) и выборки бутстреп-реализаций соответствующей статистики (I0);

погрешности измерительных каналов АСНИ;

промежуточные округления при вычислениях в процессоре;

неидеальность программных генераторов случайных чисел, используемых при статистическом моделировании.

Практическая реализация рассмотренных методов метрологического сопровождения результатов статистической обработки измерительной информации предполагает включение в состав программного обеспечения АСНИ специализированных метрологических программ, включая программные генераторы случайных чисел, позволяющих вместе с результатом статистической обработки получать оценку характеристик его погрешности по одному из вариантов бутстреп-метода.

Отметим в заключение, что бутстреп допускает естественное обобщение на более сложные статистические структуры данных, чем структура (I), например временные ряды, многофакторные конфигурации, многомерные данные и т.д. /4/. Границы применимости бутстрепа и других методов управления выборкой не всегда достаточно ясны в силу новизны этих методов. Однако, последние вооружают нас новым взглядом на статистические методы, в рамки которого хорошо укладываются многие старые подходы, а для практических задач измерений и автоматизации эксперимента могут стать методологической основой конструирования метрологического обеспечения процедур обработки данных любой сложности, практически не требуя априорной информации о вероятностно-статистической модели данных.

Библиографический список

I. ГОСТ 26.203-81. Комплексы измерительно-вычислительные. Признаки классификации. Общие требования. М.: Изд-во стандартов, 1982. 9 с.

2. Соболев В.С. Проблемы метрологического обеспечения сложных измерительных процедур, реализуемых ИИС на базе ПрИС //Тез.докл. Всесоюз. н/т конф. "ИИС-87". Ч. I. Ташкент: Таш. ПИ, 1987, с. 24.

3. *Miller R.G. The jackknife. – A review // Biometrika. – 1974. – 61. – pp. 1–15.*
4. Эфрон Б. Нетрадиционные методы многомерного статистического анализа //Сб. статей. /Пер. с англ. под ред. Ю.П.Адлера. М.: Финансы и статистика, 1988. 264 с.

УДК 519.722

Д.Б.Аратский, Е.А.Солдатов, В.Р.Фидельман

СПЕКТРАЛЬНЫЙ АНАЛИЗ НЕГАУССОВЫХ СЛУЧАЙНЫХ ПРОЦЕССОВ НА ОСНОВЕ ПРИНЦИПА МАКСИМАЛЬНОЙ ЭНТРОПИИ

(г. Горький)

Современное спектральное оценивание является мощным средством исследования статистической структуры случайных процессов. Известно, что одной из наиболее информативных характеристик случайного процесса является спектральная плотность мощности (СПМ) /1/. Во многих прикладных задачах научных исследований приходится иметь дело со спектральной обработкой негауссовых случайных процессов, поскольку к ним можно отнести любой квазипериодический сигнал, а также самоизлучающиеся сигналы сложных механических систем /2/. При этом априорная информация о процессе представляет собой набор отсчетов автокорреляционной функции (АКФ) и отличных от нуля высших моментных функций (ВМФ). Поскольку для полного статистического описания и изучения спектрального состава негауссова случайного процесса требуется знание бесконечного ряда его моментных функций, а на практике доступным измерению оказывается всегда их конечное число, имеющиеся данные являются принципиально неполными.

Пусть априорная информация представляет собой конечный набор отсчетов АКФ и ВМФ случайного стационарного негауссова процесса. Требуется на основе этой заведомо неполной информации построить достоверную оценку СПМ высокого разрешения, которая обеспечивала бы исчерпывающее для имеющегося объема данных описание спектрального состава процесса.