АКАДЕМИЯ НАУК СССР
СИБИРСКОЕ ОТДЕЛЕНИЕ
АВТОМЕТРИЯ

№ 3 1966

УДК 681.142.82

В. И. ЮШИН

(Новосибирск)

ОПТИМАЛЬНЫЕ ИНТЕРВАЛЫ ОСРЕДНЕНИЯ ПРИ ИЗМЕРЕНИИ СТАТИСТИЧЕСКИХ ХАРАКТЕРИСТИК НЕСТАЦИОНАРНОГО ПРОЦЕССА ПО ОДНОЙ РЕАЛИЗАЦИИ

Рассмотрены среднеквадратичные ошибки измерения математического ожидания и дисперсии одного класса нестационарных случайных процессов по способу скользящего осреднения одной реализации. Получены выражения для оптимальных (по критерию минимума этих ошибок) интервалов осреднения.

Большой объем вычислительных операций при определении статистических характеристик нестационарных случайных процессов с осреднением по множеству реализаций заставляет искать менее громоздкие способы их измерений. Наиболее простой из них — способ сглаживания одной реализации — впервые предложен В. С. Пугачевым [1]. Один из вариантов такого способа, названного способом измерения кратковременных корреляционных функций, реализован в известном корреляторе Беннета [2, 3]. Последовательность подобных кратковременных оценок, полученных по разным участкам одной и той же реализации, называется текущей оценкой (текущим математическим ожиданием, текущей дисперсией, текущей корреляционной функцией и т. д.). Интервал осреднения при измерении текущих статистических характеристик выбирается небольшим, с тем чтобы избежать значительных ошибок, вызванных нестационарностью. Однако чрезмерное уменьшение интервала приводит к возрастанию ошибок, связанных с недостаточным объемом выборки.

Вопросы выбора интервала осреднения рассмотрены в [1, 4], где получены выражения, позволяющие оценить сверху среднеквадратичные ошибки текущих математического ожидания, дисперсии и корреляционной функции в зависимости от характера случайного процесса. В качестве априорной информации о случайном процессе используются максимальные отклонения истинных статистических характеристик от линейной функции. Подобный подход ориентирует в известном смысле на максимальную погрешность измерения, не учитывая сравнительный вес максимальных значений, и, кроме того, не дает непосредственной оценки для наилучшего интервала осреднения. Между тем очевидно, что существует некоторая оптимальная величина интервала осреднения, при которой среднеквадратичные ошибки, вызванные, с одной стороны, нестационарностью, а с другой — конечностью объема выборки, в сумме дадут минимум. Такие соотношения для текущего математического ожидания и текущей дисперсии и выводятся в настоящей работе, причем в качестве исходной информации используются только статистические характеристики.

Постановка задачи

Будем рассматривать такое множество $Z(t)=\{z(t)\}$ нестационарных случайных процессов $z(t)$, математические ожидания $\mu(t)$, дисперсии $\delta(t)$ и корреляционные функции $\varkappa(t_1 t_2)$ которых являются реализациями некоторых стационарных случайных процессов $m_Z(t)$, $D_Z(t)$ и $K_Z(t_1 t_2)$:

$$m_Z(t)=\{\mu(t)\}; \tag{1}$$

$$D_Z(t)=\{\delta(t)\}; \tag{2}$$

$$K_Z(t_1 t_2)=\{\varkappa(t_1 t_2)\}. \tag{3}$$

Поскольку каждый случайный процесс $z(t)$ характеризуется множеством своих реализаций $\{\zeta(t)\}$, мы будем каждое из множеств $\{\zeta(t)\}_i$ называть i-м подмножеством в отличие от основного множества $Z(t)=\{z(t)\}$. Примем следующие допущения относительно объединения и пересечения подмножеств:

$$\bigcup_{i=1}^{\infty}\{\zeta(t)\}_i=\{z(t)\}=Z(t); \tag{4}$$

$$\bigcap_{i=1}^{\infty}\{\zeta(t)\}_i=0, \tag{5}$$

за исключением реализаций нулевой вероятностной меры. Физически допущения (4) и (5) означают, что любая реализация $\zeta(t)$ любого случайного процесса $z(t)$ является в то же время реализацией случайного процесса $Z(t)$ и что различные процессы $\{\zeta(t)\}_i$ и $\{\zeta(t)\}_k$, где $i\neq k$, не имеют общих реализаций. В соответствии с разделением основного множества на подмножества будем различать операцию M_0 математического ожидания по основному множеству и операцию M_i математического ожидания по i-му подмножеству. Определим математическое ожидание $m_Z(t)$, дисперсию $D_Z(t)$ и корреляционную функцию $K_Z(t_1 t_2)$ каждого из процессов $z(t)$ как

$$m_Z(t)=M_i[Z(t)]; \tag{6}$$

$$D_Z(t)=M_i\{[Z(t)-m_Z(t)]^2\}; \tag{7}$$

$$K_Z(t_1 t_2)=M_i\{[Z(t_1)-m_Z(t_1)]\,[Z(t_2)-m_Z(t_2)]\}. \tag{8}$$

Предполагается, что наблюдатель получает одну реализацию $\zeta(t)$ из множества $Z(t)$, но неизвестно, из какого именно подмножества (априорные вероятности принадлежности $\zeta(t)$ к тому или иному подмножеству одинаковы), и принимает в качестве оценки истинных статистических характеристик $m_Z(t)$, $D_Z(t)$, $K_Z(t_1t_2)$ значения текущих статистических характеристик $m_Z^T(t)$, $D_Z^T(t)$, $K_Z^T(t_1t_2)$, полученных в соответствии с алгоритмами скользящего осреднения:

$$m_Z^T(t) = \frac{1}{T} \int\limits_{t-\frac{T}{2}}^{t+\frac{T}{2}} Z(s)\,ds; \tag{9}$$

$$D_Z(t) = \frac{1}{T} \int\limits_{t-\frac{T}{2}}^{t+\frac{T}{2}} [Z(s) - m_Z(s)]^2\,ds; \tag{10}$$

$$K_Z^T(t_1t_2) = \frac{1}{T} \int\limits_{\frac{t_1+t_2-T}{2}}^{\frac{t_1+t_2+T}{2}} [Z(s_1) - m_Z(s_1)]\,[Z(s_2) - m_Z(s_2)]\,d\left(\frac{s_1+s_2}{Z}\right). \tag{11}$$

Критерием точности текущих оценок является среднеквадратичная по основному множеству ошибка

$$\overline{\varepsilon_a^2} = M_0\left\{\left[a^T(t) - a(t)\right]^2\right\}, \tag{12}$$

где a — соответствующая статистическая характеристика (6)—(8).

В настоящей работе ставится задача нахождения выражений для ошибки (12) и определения таких значений $T=T_0$, при которых $\overline{\varepsilon_a^2}=\min$. Задача решается для математического ожидания и дисперсии. Измерение текущей корреляционной функции не рассматривается, так как этот случай близок к измерению дисперсии.

Оптимальный интервал осреднения при измерении текущего математического ожидания

Среднеквадратичная ошибка текущего математического ожидания, определяемая формулами (12), (9) и (6), равна

$$\overline{\varepsilon_a^2} = M_0\{[m_Z^T(t) - m_Z(t)]^2\} = M_0\{m_Z^T(t)]^2 - 2m_Z^T(t)\,m_Z(t) - m_Z^2(t)\}. \tag{13}$$

Вводя в рассмотрение центрированные случайные процессы

$$Z^0(t) = Z(t) - m_Z(t) \tag{14}$$

и подставляя формулы (14) и (9) в (13), получим

$$\overline{\varepsilon^2} = M_0M_i\left\{\frac{1}{T^2} \int\limits_{t-\frac{T}{2}}^{t+\frac{T}{2}} \int\limits_{t-\frac{T}{2}}^{t+\frac{T}{2}} [Z^0(s_1)\,Z^0(s_2) - Z^0(s_1)\,m_Z(s_2) -\right.$$

$$- Z^0(s_2)\, m_Z(s_1) + m_Z(s_1)\, m_Z(s_2)]\, ds_1 ds_2 -$$

$$- \frac{2}{T} \int\limits_{t-\frac{T}{2}}^{t+\frac{T}{2}} [Z^0(s)\, m_Z(t) + m_Z(s)\, m_Z(t)]\, ds + m_Z^2(t) \Big\}. \qquad (15)$$

После операции математического ожидания сначала по i-му подмножеству, а затем по основному множеству, учитывая

$$M_i [Z^0(s)\, m_Z(t)] = 0; \qquad (16)$$

$$M_i [Z^0(s_1)\, Z^0(s_2)] = K_Z(s_1 s_2), \qquad (17)$$

а также вводя обозначения

$$M_0 [K_Z(t_1 t_2)] = \mathrm{M}_k(s_1 s_2) = \mathrm{M}_k(s_1 - s_2); \qquad (18)$$

$$M_0 [m_Z(s_1)\, m_Z(s_2)] = B_m(s_1 s_2) = B_m(s_1 - s_2) \qquad (19)$$

(здесь $B_m(s_1 - s_2)$ — второй начальный момент по основному множеству математического ожидания $m_Z(s)$,

$\mathrm{M}_k(s_1 - s_2)$ — математическое ожидание по основному множеству корреляционной функции процесса $z(t)$,

получим выражение

$$\overline{\varepsilon^2} = \frac{1}{T} \int\limits_{t-\frac{T}{2}}^{t+\frac{T}{2}} \int\limits_{t-\frac{T}{2}}^{t+\frac{T}{2}} [\mathrm{M}_k(s_1 - s_2) + B_m(s_1 - s_2)]\, ds_1 ds_2 -$$

$$- \frac{2}{T} \int\limits_{t-\frac{T}{2}}^{t+\frac{T}{2}} B_m(s - t)\, ds + B_m(0), \qquad (20)$$

которое, благодаря принятой здесь как допущение симметрии функций $\mathrm{M}_k(s_1 - s_2)$ и $B_m(s_1 - s_2)$, приводится к виду

$$\overline{\varepsilon^2} = \frac{2}{T} \int\limits_0^T \left(1 - \frac{s}{T}\right) \mathrm{M}_k(s)\, ds + \frac{2}{T} \int\limits_0^T \left(1 - \frac{s}{T}\right) B_m(s)\, ds -$$

$$- \frac{4}{T} \int\limits_0^{\frac{T}{2}} B_m(s)\, ds + B_m(0). \qquad (21)$$

Формула (21) в общем виде определяет среднеквадратичную ошибку оценки математического ожидания, полученную методом сглаживания одной реализации. Чтобы минимизировать ошибку $\overline{\varepsilon^2}$, найдем условие

$$\frac{\partial \overline{\varepsilon^2}}{\partial T} = 0, \qquad (22)$$

из которого выводим уравнение

$$\frac{2}{T}\int_0^T\left(1-\frac{s}{T}\right)\mathrm{M}_k(s)\,ds-\frac{1}{T}\int_0^T\mathrm{M}_k(s)\,ds+\frac{2}{T}\int_0^T\left(1-\frac{s}{T}\right)B_m(s)\,ds-$$

$$-\frac{1}{T}\int_0^T B_m(s)\,ds-\frac{2}{T}\int_0^{\frac{T}{2}}B_m(s)\,ds+B_m\left(\frac{T}{2}\right)=0. \qquad (23)$$

Решение этого уравнения относительно T дает оптимальный интервал осреднения. Уравнения (21) и (23) справедливы для произвольного вида симметричных функций $\mathrm{M}_k(s)$ и $B_m(s)$. Ими можно воспользоваться, если известен аналитический вид этих функций. Рассмотрим три примера.

Пример 1. «Средняя» корреляционная функция множества нестационарных процессов есть

$$\mathrm{M}_k(s)=\mathrm{M}_k e^{-\alpha|s|}. \qquad (24)$$

Корреляционная функция множества математических ожиданий этих нестационарных процессов имеет вид

$$B_m(s)=B_m(0)\,e^{-\beta|s|}, \qquad (25)$$

где M_k — «средняя» дисперсия множества процессов $\{z(t)\}$;
$B_m(0)$ — дисперсия математических ожиданий этих процессов.

Подставляя (24) и (25) в (23), получим приближенное выражение для оптимального интервала осреднения $T=T_0$ при $\alpha\gg\beta$

$$T_0\approx\sqrt{\frac{12}{\alpha\beta}\,\frac{\mathrm{M}_k}{B_m(0)}}. \qquad (26)$$

На основании (21) можем записать формулу среднеквадратичной ошибки

$$\overline{\varepsilon^2}\approx 2\frac{\mathrm{M}_k}{\alpha T}+\frac{1}{6}\beta T B_m(0). \qquad (27)$$

Подставляя (26) в (27), найдем минимальную среднеквадратичную ошибку

$$\overline{\varepsilon^2}_{\min}\approx 2\sqrt{\frac{\mathrm{M}_k B_m(0)}{3}\,\frac{\beta}{\alpha}}. \qquad (28)$$

Последняя формула хорошо объясняется физически: из нее следует, что суммарная среднеквадратичная ошибка из-за нестационарности и конечности объема выборки достигает минимума при таком значении интервала осреднения, при котором обе указанные составляющие ошибки сравняются.

Пример 2. «Средняя» корреляционная функция определяется через интервал корреляции T_k формулой

$$\mathrm{M}_k(s) = \begin{cases} \mathrm{M}_k\left(1 - \dfrac{|s|}{T_k}\right); & |s| < T_k; \\ 0; & |s| > T_k, \end{cases} \tag{29}$$

а корреляционная функция $B_m(s)$ — формулой (25). Подставляя (29) и (25) в (21) и (23), находим:

$$\overline{\varepsilon^2} \approx \frac{T_k}{T}\mathrm{M}_k + \frac{1}{6}\beta T B_m(0); \tag{30}$$

$$T_0 \approx \sqrt{6\frac{T_k}{\beta}\frac{\mathrm{M}_k}{B_m(0)}}; \tag{31}$$

$$\overline{\varepsilon^2_{\min}} \approx 2\sqrt{\frac{1}{6}T_k\beta\mathrm{M}_k B_m(0)}. \tag{32}$$

Пример 3. Функция $\mathrm{M}_k(s)$ задана в виде (29), функция $B_m(s)$ — в виде

$$B_m(s) = \begin{cases} B_m(0)\left(1 - \dfrac{|s|}{T_B}\right); & |s| < T_B; \\ 0; & |s| > T_B, \end{cases} \tag{33}$$

причем $T_k \ll T_B$. В этом случае получаем:

$$\overline{\varepsilon^2} \approx \frac{T_k}{T}\mathrm{M}_k + \frac{1}{6}\frac{T}{T_B}B_m(0); \tag{34}$$

$$T_0 \approx \sqrt{6T_k T_B\frac{\mathrm{M}_k}{B_m(0)}}; \tag{35}$$

$$\overline{\varepsilon^2_{\min}} \approx 2\sqrt{\frac{1}{6}\frac{T_k}{T_B}\mathrm{M}_k B_m(0)}. \tag{36}$$

Оптимальный интервал осреднения и среднеквадратичная ошибка измерения текущей дисперсии

Рассмотрим нестационарный относительно i-го подмножества случайный процесс

$$Y(t) = Z^2(t). \tag{37}$$

Допустим, что

$$M_i[Z(t)] \equiv 0, \tag{38}$$

т. е. нестационарность процесса $Z(t)$ заключается только в его дисперсии

$$D_Z(t) = M_i[Z^2(t)] \tag{39}$$

и корреляционной функции

$$K_Z(t_1 t_2) = M_i[Z(t_1) Z(t_2)]. \tag{40}$$

Корреляционная функция процесса $Y(t)$, взятая по i-му подмножеству

$$K_y(t_1 t_2) = M_i[Y(t_1) Y(t_2)] - M_i[Y(t_1)] M_i[Y(t_2)], \tag{41}$$

является случайной функцией относительно основного множества, однако ее математическое ожидание по основному множеству мы будем полагать, как и в предыдущем разделе, зависящим только от разности аргументов t_1 и t_2:

$$\begin{aligned} \mathrm{M}_{ky}(t_1 t_2) &= M_0[K_y(t_1 t_2)] = \mathrm{M}_{ky}(t_1 = t_2) = \\ &= M_0\{M_i[Y(t_1) Y(t_2)] - M_i[Y(t_1) Y(t_2)]\}. \end{aligned} \tag{42}$$

Случайная относительно основного множества функция $D_Z(t)$, по нашему предположению, является стационарной и имеет математическое ожидание

$$\mathrm{M}_{DZ} = M_0[D_Z(t)], \tag{43}$$

не зависящее от текущего времени, и корреляционную функцию

$$B_{DZ}(t_1 t_2) = B_{DZ}(t_1 - t_2) = M_0[D_Z(t_1) D_Z(t_2) - \mathrm{M}_{DZ}^2], \tag{44}$$

зависящую только от разности аргументов.

Функции $\mathrm{M}_{ky}(s)$ и $B_{DZ}(s)$, определяемые формулами (42) и (44), позволяют непосредственно воспользоваться выражениями (21) для среднеквадратичной ошибки и (23) для оптимальной длины интервала.

Однако определение моментов четвертого порядка обычно затруднительно. К более простым результатам приводит случай, когда процесс $Z(t)$ имеет нормальное распределение. Первое слагаемое в формуле (41) можно выразить с помощью моментов второго порядка функции $Z(t)$ [1]:

$$\begin{aligned} M_i[Y(t_1) Y(t_2)] = M_i[Z^2(t_1) Z^2(t_2)] = D_Z(t_1) D_Z(t_2) + \\ + 2K_Z^2(t_1 t_2). \end{aligned} \tag{45}$$

Второй член формулы (41) на основании (39) равен

$$M_i[Y(t_1)] M_i[Y(t_2)] = D_Z(t_1) D_Z(t_2). \tag{46}$$

Подставляя (45) и (46) в (42), получим

$$\mathrm{M}_{ky}(t_1 - t_2) = 2M_0[K_Z^2(t_1 t_2)]. \tag{47}$$

Таким образом, для того, чтобы вычислить среднеквадратичную ошибку измерения текущей дисперсии нормально распределенного нестационарного случайного процесса и оптимальный интервал осреднения, необходимо иметь в качестве априорной информации математическое ожидание квадрата корреляционной функции этого процесса (47) и корреляционную функцию его дисперсии (44).

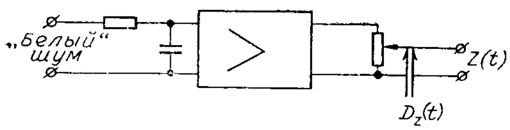
Рассмотрим один частный, но довольно распространенный случай, когда корреляционная функция исходного нестационарного процесса $Z(t)$ имеет вид

$$K_Z(t_1 t_2) = D_Z\left(\frac{t_1 + t_2}{2}\right) e^{-\alpha |t_1 - t_2|}, \tag{48}$$

где $D_Z(t)$ — дисперсия процесса $Z(t)$, зависящая от текущего времени;

α — некоторая положительная постоянная.

Корреляционную функцию подобного вида имеет, например, сигнал на выходе устройства, состоящего из апериодического звена со случайно изменяющимся (по закону $D_Z(t)$) коэффициентом усиления, если на вход этого устройства подан «белый» шум (см. рисунок). Предположим, что корреляционная функция $B_{DZ}(s)$ известна и имеет вид



$$B_{DZ}(s) = B_{DZ}(0)\, e^{-\beta |s|}, \tag{49}$$

где $\beta \ll \alpha$ (см. (48)).

Иными словами, основная энергия процесса $D_Z(t)$ сосредоточена на более низких частотах по сравнению с $Z(t)$. Помимо этого, известна и средняя по основному множеству дисперсия процесса $Z(t)$

$$\mathrm{M}_{DZ} = M_0[D_Z(t)] = \mathrm{M}_k(0) = \mathrm{M}_k. \tag{50}$$

Перепишем формулу (44) с учетом (49) и (50):

$$B_{DZ}(t_1 - t_2) = B_{DZ}(0)\, e^{-\beta |t_1 - t_2|} = M_0[D_Z(t_1)\, D_Z(t_2)] - \mathrm{M}^2_{DZ}. \tag{51}$$

Из уравнений (48), (47) и (51) находим

$$\begin{aligned} \mathrm{M}_{ky}(t_1 - t_2) &= 2M_0\left[D^2_Z\left(\frac{t_1 + t_2}{2}\right) e^{-2\alpha |t_1 - t_2|}\right] = \\ &= 2\,[B_{DZ}(0) + \mathrm{M}^2_{DZ}]\, e^{-2\alpha |t_1 - t_2|}. \end{aligned} \tag{52}$$

Подставляя (51) и (52) в (21) и (23), получаем выражения для среднеквадратичной ошибки

$$\overline{\varepsilon^2} \approx \frac{2}{\alpha T}[B_{DZ}(0) + \mathrm{M}^2_{DZ}] + \frac{\beta}{6} B_{DZ}(0), \tag{53}$$

оптимального интервала осреднения

$$T_0 \approx \sqrt{\frac{12}{\alpha\beta}\,\frac{\mathrm{M}^2_{DZ}}{B_{DZ}(0)}} \tag{54}$$

и минимальной среднеквадратичной ошибки

$$\overline{\varepsilon^2_{\min}} \approx 2\sqrt{\frac{1}{3}\,\frac{\beta}{\alpha}\, B_{DZ}(0)\,[B_{DZ}(0) + \mathrm{M}^2_{DZ}]}. \tag{55}$$

Заключение

Полученные результаты позволяют найти наилучшие в среднем для некоторого множества нестационарных процессов интервалы осреднения для измерения текущих математических ожиданий и дисперсий. Для этого требуется знание таких средних по данному множеству процессов статистических характеристик, как средняя корреляционная функция процесса, средняя корреляционная функция квадрата процесса и корреляционные функции математического ожидания и дисперсии. Для нормальных процессов с экспоненциальными корреляционными функциями грубые оценки всех этих характеристик, как правило, легко доступны. В дальнейшем при повторении измерений, пользуясь методами, изложенными в [1], можно уточнить интервал осреднения применительно к каждому конкретному случаю, а также полученную среднеквадратичную ошибку.

ЛИТЕРАТУРА

1. Б. С. Пугачев. Теория случайных функций и ее применение к задачам автоматического управления. М., Физматгиз, 1962.
2. W. R. Bennet. A Machine for Continuons Display of Short Term Correlation. The Correlatograf.— Bell. Syst. Tech. J., 1953, v. 32, № 5.
3. Б. С. Синицын. Автоматические корреляторы и их применение. Новосибирск, РИО, 1964.
4. Н. А. Лифшиц, В. Н. Пугачев. Вероятностный анализ систем автоматического управления. М., изд-во «Советское радио», 1963.

Поступила в редакцию
25 октября 1966 г.