АКАДЕМИЯ НАУК СССР
СИБИРСКОЕ ОТДЕЛЕНИЕ
АВТОМЕТРИЯ

№ 6 1968

УДК 519.281

Т. М. АЛИЕВ, Г. В. ЕВЗЛИНА, А. А. ТЕР-ХАЧАТУРОВ,
Ю. В. ЩЕРБИНИН
(Сумгаит)

ИССЛЕДОВАНИЕ ПОГРЕШНОСТЕЙ СТАТИСТИЧЕСКОГО ИНТЕГРИРОВАНИЯ ПРИ КОРРЕЛЯЦИОННОЙ СВЯЗИ ГЕНЕРИРУЕМЫХ СЛУЧАЙНЫХ ВЕЛИЧИН*

Статистическое интегрирование, под которым подразумевается интегрирование по методу Монте-Карло, является одним из новых направлений в измерительной технике, имеющим целью построение принципиально новых, более простых и современных по структуре аналого-дискретных измерительно-вычислительных устройств [1, 2].

При выполнении требований равномерности и независимости случайных чисел погрешность статистического интегрирования зависит только от числа испытаний и определяется по формуле

$$\varepsilon < \frac{3}{2} N^{-\frac{1}{2}}, \qquad (1)$$

где N — число испытаний.

Вероятность того, что значение интегрируемой функции больше случайной величины, описывается выражением

$$P\{x > y\} = \int_0^1 G(x/x)\, d\, F(x), \qquad (2)$$

где $G(x/x)$ — вероятность того, что случайная величина меньше x при условии, что значение интегрируемой функции равно x.

Если интегрируемая случайная величина имеет функцию распределения $G(x)$ и при этом интегрируемая функция и генерируемая случайная величина независимы, то справедливо равенство $G(x/x) = G(x)$.

Если же генерируемая случайная величина имеет функцию распределения $G(x) = x$, тогда

$$P\{x > y\} = \int_0^1 x\, d\, F(x) = M_x,$$

где M_x — математическое ожидание исследуемой функции.

---

* Представленный материал является развитием работы, доложенной на VII Всесоюзной конференции по автоматическому контролю и методам электрических измерений в сентябре 1965 года в Новосибирске.

Из изложенного становится очевидной причина возникновения погрешности при отклонении функции распределения случайной величины от равномерной, и определение такой погрешности не вызывает затруднений.

Значительные трудности вызывает определение погрешности статистического интегрирования при наличии корреляционной зависимости между последовательно генерируемыми случайными величинами.

Степень корреляционной зависимости генерируемых случайных величин в основном определяется скоростью их выдачи, и, следовательно, это условие накладывает определенные ограничения на быстродействие статистического интегратора. Покажем это на примере двух способов генерирования случайных величин. В первом случае (рис. 1) генерируемая случайная величина представляет собой значения некоторого физического случайного процесса, выдаваемые с некоторой частотой $f = \frac{1}{\Delta t}$, где Δt — интервал времени между двумя последовательно выдаваемыми случайными величинами. Очевидно, что с повышением f увеличивается степень корреляции случайных величин.

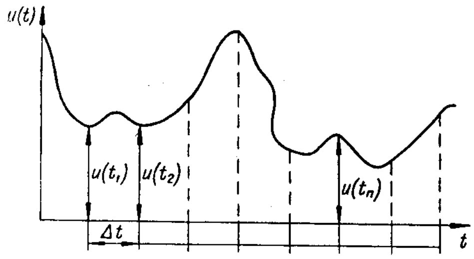

Рис. 1.

Во втором случае (рис. 2) случайная величина представляет собой временной интервал между передним фронтом импульса случайной частоты и началом того периода постоянной частоты, в течение которого он выпал.

При этом корреляционная зависимость генерируемых случайных временных интервалов растет с увеличением математического ожидания случайной частоты выпадения импульсов.

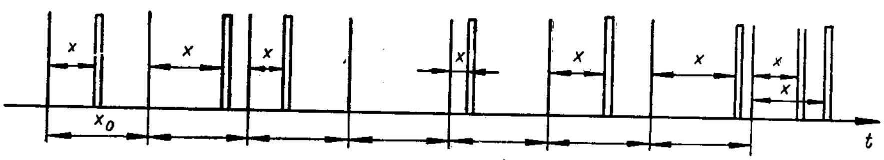

Рис. 2.

Наличие корреляции генерируемых случайных величин приводит к тому, что:

во-первых, вероятность наступления события $|x>y|$ в каждом отдельном испытании не постоянна и не равна M_x, а определяется коррелированностью случайных величин и значением случайной и исследуемой величин в предыдущем испытании (ниже будет показано, что это явление не вносит погрешности в результат измерения, а даже, напротив, уменьшает дисперсию результатов статистического интегрирования);

во-вторых, функция распределения генерируемых случайных величин зависит от номера испытания и, следовательно, от времени. В этом случае естественно предположить, что вероятность события $|x>y|$ будет определяться не только величиной математического ожидания исследуемой функции, но и характером протекания ее во времени. Тогда

$$P\{x > y\} = \int_0^1 G(x/x)\, d F(x) \neq M_x,$$

где $P\{x>y\}$ — интегральная вероятность события $x>y$ серии N статистических испытаний. При этом $G(x/x)$ определяется не только величиной x, но и моментом времени, когда интегрируемая функция приняла значение x, т. е. $G(x/x) \neq x$.

С увеличением корреляции случайных величин событие $x>y$ становится все менее случайным. В предельном случае, когда коэффициент автокорреляции генерируемых случайных величин равен единице, а исследуемая функция представляет собой линейную функцию $x=kt$, при условии $M_x > M_y$ вероятность события $x>y$ становится равной единице, а при $M_x < M_y$ вероятность $x>y$ равна нулю.

Теперь становится очевидной методика определения погрешности измерения среднего значения конкретной реализации случайного процесса в зависимости от степени корреляции генерируемых случайных величин.

Интеграл $\int_0^1 G(x/x)\, d F(x)$ заменим суммой n членов, представляющих собой произведение вероятности i-го значения исследуемой функции на вероятность того, что генерируемая случайная величина меньше, чем i-е значение исследуемой функции, при условии, что последнее равно i-му:

$$\int_0^1 G(x/x)\, d F(x) = \sum_{i=1}^{n} G(x_i/x_i)\, P(x_i). \qquad (3)$$

Значение $P(x_i)$ определяется как частность x_i в исследуемой реализации, значение $G(x_i/x_i)$ — как математическое ожидание вероятностей того, что случайная величина меньше x_i при условии, что интегрируемая функция приняла значение, равное x_i.

Для определения $G(x_i)_n$ на n-м шаге испытаний воспользуемся математическим аппаратом исследования марковских процессов, поскольку генерируемые случайные величины составляют марковскую цепь.

Генерирование случайных величин y схематически можно представить как серию последовательных испытаний, заключающихся в случайном выборе из некоторой генеральной совокупности, представляющей собой множество всех возможных значений y. Очевидно, что при наличии коэффициента автокорреляции случайных величин y вероятность выбора y_j в $(s+1)$-м испытании зависит от того, какое значение y было выбрано в s-м испытании, и не зависит от того, какие значения y выбирались в испытаниях, предшествующих s-му.

Воспользуемся иной терминологией и будем рассматривать эту серию последовательных испытаний как некоторую физическую систему S, которая в каждый момент времени может находиться в одном из состояний $x_1, x_2, \ldots, x_n$ и изменяет свое состояние только в момент $t_1, t_2, \ldots, t_n$.

Полная вероятностная картина возможных изменений, осуществляющихся при переходе от одного испытания к непосредственно следующему, задается матрицей перехода.

$$\Pi_1 = \begin{pmatrix} P_{11} & P_{12} & \dots & P_{1k} \\ P_{21} & P_{22} & \dots & P_{2k} \\ \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots \\ P_{k1} & P_{k2} & \dots & R_{kk} \end{pmatrix}, \tag{4}$$

где P_{ij} — вероятность перехода, т. е. вероятность перехода системы S в состояние j, если она в состоянии i.

Для получения полной вероятностной картины возможных изменений на каждом шаге серии N испытаний необходимо определять вероятности перехода из состояния i в 1-м испытании в состояние j через n испытаний. Обозначим эту вероятность $P_{ij}(n)$. Матрица перехода через n испытаний соответствует

$$\Pi_n = \begin{pmatrix} P_{11}(n) & P_{12}(n) & \dots & P_{1k}(n) \\ \dots & \dots & \dots & \dots \\ P_{k1}(n) & P_{k2}(n) & \dots & P_{kk}(n) \end{pmatrix} \tag{5}$$

Согласно теории марковских цепей,

$$\Pi_n = (\Pi_1)^n; \ \lim_{n \to \infty} P_{ij}(n) = P_j.$$

Естественно заключить, что чем сильнее статистическая зависимость между генерируемыми случайными величинами y, тем медленнее $P_{ij}(n)$ приближается к своему пределу P_j.

Зная вероятности P_{11}, P_{12}, P_{13}, ..., P_{1k}, можно определить $G(x_i/x_i)$ на каждом шаге испытаний как сумму соответствующих P_{1i}:

$$G(x_i/x_i) = \bar{P}_{11} + \bar{P}_{12} + \dots + \bar{P}_{1i} = \sum_{i=1}^{i} \bar{P}_{1i}, \tag{6}$$

где $\bar{P}_{1i}$ — среднее значение P_{1i} на всех шагах испытаний, где исследуемая величина равнялась x_i.

Разность между $\int\limits_0^1 x\, dF(x)$ и $\sum\limits_{i=1}^{n} G(x_i/x_i)\, P(x_i)$ представляет собой систематическую погрешность статистического интегрирования, связанную с корреляцией генерируемых случайных величин.

Элементы первичной матрицы перехода Π_1 определяются тем или иным способом, в зависимости от метода генерирования случайных величин.

Для описанного выше варианта генерирования случайных временных интервалов

$$P(y_i) = P_{1i} = e^{-\lambda x_i} - e^{-\lambda (x_i + \Delta x)} + \frac{e^{-\lambda}}{1 - e^{-\lambda}} (e^{-\lambda x_i} - e^{-\lambda (x_i + \Delta x)}),$$

где λ — математическое ожидание случайной, распределенной по Пуассону частоты выпадения импульсов; Δx — величина дискретности, равная $\frac{1}{n}$.

Как уже упоминалось, корреляция генерируемых случайных чисел приводит к тому, что вероятность наступления события $x>y$ в каждом отдельном испытании не постоянна и не равна M_x, а зависит от степени корреляции как генерируемой случайной величины, так и исследуемой функции, а также от значений генерируемой случайной величины и интегрируемой функции в предшествующем испытании.

В этом случае распределение частоты наступления события будет отличаться от обычного биномиального. Математическое ожидание такого распределения равно

$$\mu_1 = n\overline{p},$$

где $\overline{p}$ — среднее арифметическое всех P_j, а дисперсия

$$\mu_2 = n\overline{p}\,\overline{q} - n\sigma_p^2,$$

где σ_p^2 — дисперсия p.

Сравнение этих результатов с соответствующими результатами для биномиального распределения показывает, что дисперсия распределения с различными вероятностями P_j меньше, чем дисперсия биномиального распределения с постоянной вероятностью $p=\overline{p}$, на величину $n\sigma_p^2$.

Этот вывод имеет важное практическое значение. Следствием его является тот факт, что если $p\{x>y\}$ принимает i различных значений в ходе статистических испытаний, то дисперсия частоты события $x>y$ равна $D'\left(\frac{M}{N}\right) = \frac{\overline{pq} - \sigma_p^2}{N}$ и, следовательно, будет меньше $D\left(\frac{M}{N}\right) = \frac{\overline{pq}}{N}$. Таким образом, максимальная погрешность с достоверностью 0,997 в этом случае будет равна

$$\varepsilon \leqslant 3\sqrt{\frac{\overline{pq} - \sigma_p^2}{N}}.$$

Вероятность наступления события $x>y$ остается постоянной в ходе статистических испытаний только в следующих двух случаях: 1) определение M_x эргодического процесса при отсутствии корреляции генерируемых случайных чисел; 2) определение $M(a)$ (a — постоянная величина) при отсутствии корреляции генерируемых случайных величин.

При корреляции генерируемых случайных величин все остальные случаи применения статистического интегрирования следующие.

I. Определение математического ожидания стационарного случайного процесса (значения исследуемого случайного процесса не коррелированы). На каждом шаге испытаний

$$P_i\{x > y\} = \int_0^1 G(x/y_i)\, dF(x),$$

где $G(x/y_i)$ — вероятность того, что случайная величина меньше исследуемой при следующем условии: в предыдущем испытании случайная величина равнялась y_i; $\sigma_p = \sum_{i=1}^{n} (P_i - \overline{p})^2\, p(y_i)$; $\overline{p}$ — среднее арифметическое P_i, равное математическому ожиданию исследуемого случайного процесса.

II. Определение математического ожидания нестационарного случайного процесса (значения исследуемого случайного процесса коррелированы). На каждом шаге испытаний

$$P_{ij}\{x > y\} = \int_0^1 G(x/x_j\, y_i)\, dF(x),$$

где $G(x/x_j\, y_i)$ — вероятность того, что случайная величина меньше x

при условии: в предыдущем испытании случайная величина равнялась y_i, а исследуемая — x_j.

В этом случае $\bar{p}$ отличается от математического ожидания исследуемого случайного процесса на величину, определяемую степенью корреляции генерируемых случайных величин и видом обрабатываемой реализации случайного процесса.

III. Производится определение среднего значения временной функции $f(t)$. На каждом шаге испытаний

$$P_{ij}\{f(t) > y\} = G\{f(t)/y_i f(t)_j\},$$

где $G\{f(t)/y_i f(t)_j\}$ — вероятность того, что случайная величина меньше исследуемой функции при условии: исследуемая функция в предыдущем испытании равнялась $f(t)_j$, а случайная величина — y_i.

В этом случае $\bar{p}$ также отличается от среднего значения исследуемой функции на величину, определяемую степенью корреляции случайных чисел и характером исследуемой временной функции.

IV. Определение $M(a)$ методом статистических испытаний (a — постоянная величина). На каждом шаге испытаний

$$P_i\{a > y\} = G(a/y_i),$$

где $G(a/y_i)$ — вероятность того, что случайная величина меньше постоянной величины a, при условии: в предыдущем испытании случайная величина равнялась y_i. В этом случае $\bar{p}$ равно a.

Интересно отметить, что при полной корреляции генерируемых случайных чисел (линейная зависимость) последний случай применения метода статистических испытаний превращается в развертывающее преобразование постоянной аналоговой величины.

Случай IV моделировался на разработанном статистическом интеграторе. Полученные экспериментальные данные подтверждают теоретический расчет дисперсии величины $\frac{M}{N}$ в зависимости от степени корреляции случайных чисел, пропорциональной средней частоте пуассоновского потока импульсов. Так, например, при $\lambda = 10$ дисперсия результатов уменьшилась по сравнению с обычным биномиальным распределением в 1,75 раза.

Кроме того, был проведен математический эксперимент с целью определения систематической ошибки измерения среднего значения функции $x = 0{,}01\ i$ (i — номер испытания) методом статистических испытаний с использованием случайных величин, генерируемых описанным выше способом в виде временных интервалов при средней частоте пуассоновского потока импульсов, равной 10. Число испытаний 100. Предварительно на ЭВМ были определены матрицы переходов для генерируемой случайной величины на каждом шаге. Теоретически определенная систематическая погрешность равна 0,007, а максимальная случайная погрешность — 0,12*. Результаты эксперимента совпали с расчетными данными.

ВЫВОДЫ

При корреляции генерируемых случайных чисел возникает систематическая погрешность определения математического ожидания нестационарного случайного процесса и среднего значения временной функ-

* При обычном биномиальном распределении максимальная случайная погрешность равна 0,15, а систематическая погрешность — нулю.

ции. Величина этой погрешности зависит как от степени корреляции генерируемых случайных величин, так и от вида реализации нестационарного случайного процесса или интегрируемой временной функции.

Систематическая погрешность уменьшается с увеличением числа испытаний, так как элементы матрицы переходов генерируемых случайных величин через определенное число шагов сходятся к безусловной вероятности.

При корреляции случайных генерируемых величин уменьшается дисперсия результатов серии статистических испытаний, а следовательно, и максимальная случайная погрешность.

Из проведенных экспериментов и расчетов можно заключить, что даже большая степень корреляции генерируемых случайных величин не вносит существенной систематической погрешности, а поскольку при этом сильно уменьшается случайная погрешность, то некоторая корреляция случайных величин иногда приводит к увеличению точности измерений.

ЛИТЕРАТУРА

1. М. И. Ланин, С. М. Мандельштам. О применении метода статистических испытаний для измерения среднего значения быстропеременных величин.— Сб. «Применение кибернетики в электронно-измерительной технике». М., 1963.
2. Г. В. Евзлина, Т. М. Алиев, А. А. Тер-Хачатуров, Ю. В. Щербинин. Метод статистических испытаний в задачах автометрии.— Рефераты докладов I Всесоюзного симпозиума по статистическим проблемам в технической кибернетике, ч. II. М., 1967.

Поступила в редакцию
3 июля 1967 г.,
окончательный вариант —
20 февраля 1968 г.