В. С. ГОРБЕНКО

СИСТЕМАТИЧНА ПОХИБКА ПРИ ЕКСПЕРИМЕНТАЛЬНОМУ ВИЗНАЧЕННІ ФУНКЦІЙ РОЗПОДІЛУ ІМОВІРНОСТІ, ОБУМОВЛЕНА КІНЦЕВОЮ ТРИВАЛІСТЮ ВИБІРКОВИХ ДАНИХ

При експериментальному визначенні одновимірних функцій розподілу імовірностей стаціонарних випадкових процесів часто використовується дискретна або періодична вибірка. Експериментально визначаються величини [1]

$$F^*(x) = \frac{n_1^*}{N}; \tag{1}$$

$$W^*(x) = \frac{n_2^*}{\Delta x N}, \tag{2}$$

де N — загальне число вибіркових значень реалізації випадкового процесу $\varepsilon(t)$ на інтервалі часу вимірювання T;

n_1^* — число вибіркових значень, які відповідають вимозі $\varepsilon \leqslant x$;

n_2^* — число вибіркових значень, які відповідають вимозі

$$x - \frac{\Delta x}{2} < \varepsilon \leqslant x + \frac{\Delta x}{2};$$

Δx — інтервал групування (ширина «вікна» вимірювача).

Для стаціонарного випадкового процесу математичне сподівання (1) і (2) (при $\Delta x \to 0$) відповідно дорівнюють $F(x)$ і $W(x)$.

У більшості існуючих приладів дискретна або періодична вибірка виконується не у вигляді послідовної в часі фіксації значень реалізації випадкового процесу в момент часу it_0 (t_0 — період вибірки; $i = 1, 2, \ldots, N$), а у вигляді послідовності прямокутних імпульсів тривалістю τ_0 з амплітудами, в першому наближенні рівними значенням реалізації в інтервалах $it_0 < t < it_0 + \tau_0$.

В дійсності вершина вибіркового імпульсу повторює траєкторію випадкового процесу на інтервалі вибірки. Внаслідок цього при сортуванні та обчисленні вибіркових імпульсів операції (1) і (2) виконуються не точно, що викликає появу систематичних похибок.

Обмежимося розглядом такого випадку, коли похідна неперервної реалізації випадкового процесу на інтервалі вибірки не змінює свій знак, що можна припустити при достатньо малих τ_0. На рис. 1 показаний такий випадок при визначенні імовірності $P^*(\varepsilon > x_0) = \frac{n}{N}$. Тут n^* — число вибіркових значень, що перевищують рівень аналізу x_0. Рівень x_0 перевищують всі позначені на рис. 1 імпульси, крім імпульсу під номером 5. Ці імпульси і будуть зараховані лічильною схемою. Припустимо, що ідеальний вибір імпульсів, що перевищують рівень x_0, повинен виконуватися по передніх фронтах імпульсів. Тоді лічильною схемою будуть правильно зареєстровані імпульси під номерами 1, 2, 4, 6 і 7 і неправильно буде зареєстрований імпульс під номером 3. Імпульс 3 по амплітуді перевищує рівень x_0, але передній фронт його нижче рівня x_0.

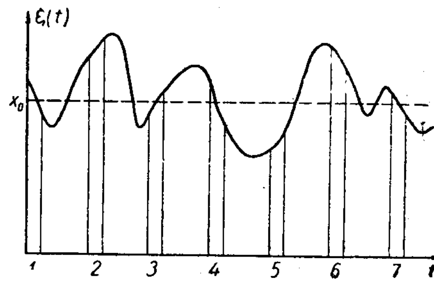

Рис. 1.

На основі такої моделі правильного і помилкового рахунку імпульсів обчислимо систематичну похибку при експериментальному визначенні функції $P(\varepsilon > x_0)$.

Імовірність події, що передній фронт імпульсу тривалістю τ_0 нижче, а задній фронт вище рівня x_0,

$$P(x_0, \tau_0) = P(x < x_0, x_{\tau_0} > x_0) = \int_{-\infty}^{x_0} \int_{x_0}^{\infty} W(x, x_{\tau_0})\, dx\, dx_{\tau_0}, \tag{3}$$

де $W(x, x_\tau)$ — двовимірна густина імовірності процесу.

Сподіване число правильно зарахованих імпульсів

$$n = P(\varepsilon > x_0) N = (1 - F(x_0)) N \tag{4}$$

і неправильно зарахованих імпульсів

$$n_н = P(x_0, \tau_0) N. \tag{5}$$

Загальне число зарахованих імпульсів

$$n_\Sigma = n + n_н. \tag{6}$$

Тоді математичне сподівання емпіричної функції $P_e^*(\varepsilon > x_0)$

$$P_e(\varepsilon > x_0) = \frac{n_\Sigma}{N} = \frac{n}{N} + \frac{n_н}{N} = P(\varepsilon > x_0) + P(x_0, \tau_0) =$$

$$= 1 - F(x_0) + P(x_0, \tau_0). \tag{7}$$

Систематична похибка ΔP при визначенні $P(\varepsilon > x_0)$

$$\Delta P = P_e(\varepsilon > x_0) - P(\varepsilon > x_0) = P(x_0, \tau_0); \tag{8}$$

відповідно систематична похибка ΔF при визначенні $F(x)$

$$\Delta F = -P(x_0, \tau_0). \tag{9}$$

Виходячи з допустимих значень ΔP або ΔF, можна визначити максимально допустиме значення тривалості вибіркового імпульсу, розв'язуючи (3) відносно τ_0.

Розглянемо окремий випадок — нормальний випадковий процес з середнім значенням нуль, нормований по σ. Підставляємо вираз для двовимірної густини імовірності такого процесу

$$W(x, x_\tau) = \frac{1}{2\pi\sqrt{1 - r^2(\tau)}} \exp\left\{\frac{x^2 + x_\tau^2 - 2xx_\tau r(\tau)}{2(1 - r^2(\tau))}\right\} \tag{10}$$

у (3) і одержуємо, використовуючи розклад (10) в ряд по поліномах Ерміта,

$$P(x_0, \tau_0) = F(x_0)(1 - F(x_0)) - \frac{1}{2\pi} e^{-x_0^2} \sum_{n=1}^{\infty} \frac{H_{n-1}^2(x_0)\, r^n(\tau_0)}{n!}, \tag{11}$$

де

$$F(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{x^2}{2}} dx.$$

При $x_0 = 0$ вираз (11) записується в замкнутому вигляді

$$P(0, \tau_0) = \frac{1}{4} - \arcsin r(\tau_0) = \arccos r(\tau_0) = \arcsin\sqrt{1 - r^2(\tau_0)}, \tag{12}$$

що збігається з виразом, наведеним у [2, стр. 331].

Визначення $P(x_0, \tau_0)$ по формулі (3) трудомістке. Тому розглянемо другий спосіб приблизного визначення ΔF.

Для деякого рівня аналізу x_0 абсолютне значення різниці між переднім і заднім фронтом вибіркового імпульсу при достатньо малих τ_0 приблизно дорівнює деякому випадковому значенню (див. рис. 2)

$$\delta^*(x_0) = \tau_0 \left|\frac{d\varepsilon(t)}{dt}\right|_{x_0}, \tag{13}$$

де $\left|\frac{d\varepsilon(t)}{dt}\right|_{x_0}$ — абсолютне значення похідної випадкового процесу на рівні x_0.

Математичне сподівання цієї різниці

$$\delta(x_0) = M\{\delta^*(x_0)\} = \tau_0 M\left\{\left|\frac{d\varepsilon(t)}{dt}\right|_{x_0}\right\}. \tag{14}$$

Тепер можна визначити сподіване число імпульсів, у яких передній і задній фронти знаходяться в інтервалі $\delta(x_0)$. Для цього

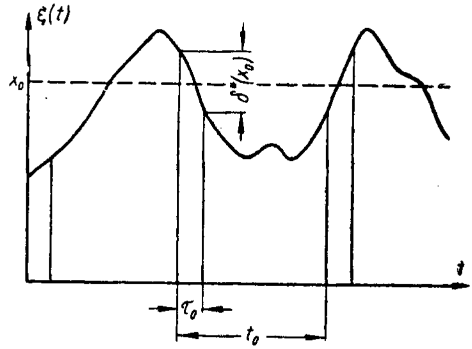

Рис. 2.

загальне число N вибіркових імпульсів потрібно помножити на імовірність попадання вершини імпульсу в інтервал $\delta(x_0)$.

$$n_0(x_0) = N\left[F\left(x_0 + \frac{\delta(x_0)}{2}\right) - F\left(x_0 - \frac{\delta(x_0)}{2}\right)\right] = N\int\limits_{x_0 - \frac{\delta(x_0)}{2}}^{x_0 + \frac{\delta(x_0)}{2}} W(x)\,dx. \tag{15}$$

Для одержання разрахункової формули припустимо, що в інтервалі $x_0 - \frac{\delta(x_0)}{2} < x < x_0 + \frac{\delta(x_0)}{2}$ можна з невеликою помилкою прийняти $W(x) = W(x_0) = \text{const}$. Тоді

$$n_0(x_0) = NW(x_0)\,\delta(x_0) = NW(x_0)\,\tau_0 M\left\{\left|\frac{d\varepsilon(t)}{dt}\right|_{x_0}\right\}. \tag{16}$$

Якщо випадковий процес та його похідна по часу в збіжні моменти часу незалежні, то [2, стр. 343]

$$W(x_0)\,M\left\{\left|\frac{d\varepsilon(t)}{dt}\right|_{x_0}\right\} = \nu_0(x_0) = 2\nu(x_0), \tag{17}$$

де $\nu_0(x_0)$ — сподіване число перетинів рівня x_0 в одиницю часу;

$\nu(x_0)$ — сподіване число викидів, що перевищують рівень x_0 в одиницю часу.

$$n_0(x_0) = N\tau_0\nu(x_0) = 2N\tau_0\nu(x_0). \tag{18}$$

При допущенні, що середнє значення похідної з позитивним знаком дорівнює за абсолютною величиною середньому значенню похідної з негативним знаком

$$M\left\{\frac{d\varepsilon(t)^+}{dt}\right\} = \left|M\left\{\frac{d\varepsilon(t)^-}{dt}\right\}\right|,$$

сподіване число неправильно залічених імпульсів

$$n_н = \frac{n_0(x_0)}{2} = \frac{1}{2}N\tau_0\nu_0(x_0) = N\tau_0\nu(x_0). \tag{19}$$

Враховуючи рівняння (7) — (10), визначимо сподівану систематичну похибку

$$\Delta P = -\Delta F = \frac{n_н}{N} = \frac{\tau\nu_0(x_0)}{2} = \tau_0\nu(x_0). \tag{20}$$

Таким чином, сподівана систематична похибка при експериментальному визначенні $F(x)$ або $P(\varepsilon > x)$ пропорціональна тривалості вибіркового значення і сподіваному числу викидів, які перевищують рівень аналізу в одиницю часу.

Якщо систематична похибка при експериментальному визначенні $F(x)$ не повинна перевищувати ΔF_m, то допустима тривалість вибіркових значень визначиться з нерівності

$$\tau_{0m} < \left|\frac{\Delta F_m}{\nu_m(x)}\right|. \tag{21}$$

Таким чином, задача визначення максимально допустимої тривалості вибіркових значень зводиться до визначення максимального числа викидів в одиницю часу в залежності від параметра x.

Наведемо розрахункові вирази для визначення систематичної похибки при експериментальному визначенні густини імовірності $W(x)$. За допомогою міркувань, як і для випадку вимірювання $F(x)$, можна показати, що математичне сподівання емпіричної густини імовірності

$$M\{W_е^*(x_0)\} = W_е(x_0) = W(x_0) + \frac{\delta(x_0)}{\Delta x}W(x_0) =$$

$$= W(x_0) + \frac{\tau_0 W(x_0)}{2\Delta x}M\left\{\left|\frac{d\varepsilon(t)}{dt}\right|_{x_0}\right\} = W(x_0) + \frac{\nu(x_0)\tau_0}{\Delta x} = W(x_0) + \Delta W. \tag{22}$$

Тут

$$\Delta W = \frac{\nu(x_0)\tau_0}{\Delta x} \tag{23}$$

— систематична похибка, обумовлена кінцевою тривалістю вибіркових даних.

Визначимо максимально допустиму тривалість вибіркових даних при допустимому значенні максимальної похибки ΔW_m

$$\tau_{0m} \leqslant \frac{\Delta W_m \Delta x}{\nu_m(x)} . \tag{24}$$

При використанні багатоканальних приладів для експериментального визначення інтегральних та диференціальних функцій розподілу імовірностей потрібно врахувати можливість помилкової реєстрації одного і того ж імпульсу в кількох суміжних каналах. Для зменшення неправильних реєстрацій потрібно виконати вимогу $\delta(x_0) \ll \Delta x$, а саме: сподіване абсолютне значення різниці між переднім та заднім фронтами вибіркового імпульсу повинно бути набагато менше ширини каналу Δx, в якому виконується підрахунок імпульсів.

ЛІТЕРАТУРА

1. Горбенко В. С. Приборы для определения интегральных и дифференциальных функций распределения вероятностей стационарных случайных процессов.— Известия вузов СССР — Радиотехника, 1962, № 3, 301.

2. Бунимович В. И. Флюктуационные процессы в радиоприемных устройствах. «Советское радио», 1951.

В. С. ГОРБЕНКО

СИСТЕМАТИЧЕСКАЯ ПОГРЕШНОСТЬ ПРИ ЭКСПЕРИМЕНТАЛЬНОМ ОПРЕДЕЛЕНИИ ФУНКЦИЙ РАСПРЕДЕЛЕНИЯ ВЕРОЯТНОСТЕЙ, ОБУСЛОВЛЕННАЯ КОНЕЧНОЙ ДЛИТЕЛЬНОСТЬЮ ВЫБОРОЧНЫХ ДАННЫХ

Краткое содержание

Получены выражения для определения систематической погрешности в зависимости от конечных длительностей выборочных данных при экспериментальном определении функций распределения вероятностей случайных процессов.

V. S. GORBENKO

SYSTEMATIC ERROR CAUSED BY THE PROLONGED DURATION OF THE SAMPLE FOR THE CASE OF THE EXPERIMENTAL PROBABILITY DISTRIBUTIONS

Summary

Relations between the duration of the sample and the systematic error for the case of the experimental probability distributions of the stochastic processed are derived.