

УДК 681.3.06:629.12.004.67

ОЦЕНКА РАЗМАХА ВЫБОРКИ ДАННЫХ ПРИ ОПТИМИЗАЦИИ СИСТЕМ ОБРАБОТКИ ИНФОРМАЦИИ И СУДОРЕМОНТА

Байздренко А.А., Горпинченко А.В., Чабаненко П.П.

Дано решение задачи оценки характеристик экстремальных значений выборки n данных через параметры распределения исследуемой случайной величины. Получены расчетные зависимости применительно к системам судоремонта и обработки сигналов в РЛС. Приводится пример оценки точности решений по разработанным математическим моделям. Применение полученных результатов позволяет сократить объем необходимых статистических данных в n раз.

При разработке автоматизированных систем сбора и обработки информации, оптимизации производственных процессов возникает задача оценки параметров распределения экстремальных членов выборки. Так, в конвейерном производстве осуществляется параллельное выполнение операций, времена их выполнения случайны и распределены по близким законам. Конвейерный цикл должен обосновываться с учетом распределения как максимального, так и минимального времени выполнения операции. При обработке радиолокационных сигналов важно рационально использовать ресурсы памяти с тем, чтобы охватить весь диапазон изменения сигнала без излишних затрат ресурсов системы его обработки. Это также связано с оценками характеристик выборочных максимума и минимума, а также размаха выборки.

Закон распределения экстремальных значений выборки получен в асимптотической форме для широкого класса распределений случайных величин $f(x)$ в виде двойного показательного закона [1,2] – распределения Фишера–Типпета I типа с функцией распределения максимума

$$F_M(x) = \exp\{-e^{(a-x)b}\} \quad (1)$$

и плотностью вероятности

$$w_M(x) = b \cdot \exp\{(a-x) \cdot b - e^{(a-x)b}\},_{-\infty < x < \infty}, \quad (2)$$

где a – параметр сдвига (мода); b – параметр масштаба.

Для применения отмеченного распределения нет необходимости в детальном знании закона, которому подчиняется случайная величина X , порождающая выборку n данных $\{X_1, \dots, X_n\}$.

Платой за это оказывается необходимость оценки параметров a и b по статистическим данным большого объема. Так, при $n = 200$ и требовании примерно такой же достоверности оценки параметров a, b распределения (2), что и достоверность оценки параметров распределения $f(x)$ величины X , нужно получить $n \times n = 4 \cdot 10^4$ данных. Действительно, в выборке фикси-

рованного объема n только одно максимальное значение $X_{\max}$, а нужно получить выборку для них такого же объема, что и для X .

Если распределение $f(x)$ неизвестно, то распределение (2) применимо в отмеченной ситуации с предварительной оценкой его параметров по статистическим экстремальным данным, например, при расчете гидротехнических сооружений, при наличии данных по максимальным расходам воды и т.п. [3].

Однако в тех случаях, когда распределение $f(x)$ известно, а выборки $\{X_{1\max}, \dots, X_{n\max}\}$ нет, подобная традиционная схема представляется расточительной. В этих случаях целесообразно установить связь параметров распределения $\omega(x)$ экстремальных значений с параметрами распределения $f(x)$ и решать задачу аналитически. Это позволит сократить объем эксперимента в n раз по сравнению со статистической оценкой.

Рассмотрим путь решения этой задачи. Для выборочного максимума $X_{\max}$ функция распределения [3]

$$F_M(x) = P(X_{\max} \leq x) = F^n(x), \quad (3)$$

где n – объем выборки; $F(x)$ – функция распределения X .

Для выборочного минимума $X_{\min}$ функция распределения имеет вид:

$$F_{\mu}(x) = 1 - F_M(-x). \quad (4)$$

Поэтому достаточно найти $F_M(X)$.

Поскольку

$$F(x) = 1 - \int_x^{\infty} f(x) dx = 1 - \psi(x), \quad (5)$$

то

$$F_M(x) = [1 - \psi(x)]^n. \quad (6)$$

Известно также, что

$$\lim_{n \rightarrow \infty} \left(1 - \frac{k}{n}\right)^n = e^{-k}, \quad (7)$$

где $k = \text{const}$ (не зависит от n).

Следовательно, решение задачи включает последовательность следующих операций:

1) поиск приемлемого приближения $\tilde{\psi}(x)$ для $\psi(x)$ как функции от $f(x)$ в неинтегральной форме;

2) приведение $\tilde{\psi}(x)$ к виду $\exp(-z)/n$ при $z = z(\theta_k, x, n)$, где θ_k – k -мерный параметр распределения $f(x) = f(\theta_k, x)$, а изменение n компенсируется изменением x так, что z не зависит от n ;

3) упрощение $z(x)$ с приведением ее к виду $\exp[(x-a)b]$, где $a = a(\theta_k, n)$, $b = b(\theta_k, a)$.

Не останавливаясь на технике вывода требуемых аналитических зависимостей, приведем полученные нами окончательные результаты для двух широко распространенных распределений случайных величин.

Для гамма-распределения плотность вероятностей

$$f(x) = f(\alpha, \beta; x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}, \quad x > 0, \quad (8)$$

где α – параметр формы, β – параметр масштаба.

Параметры распределения $\omega(x) = \omega(a, b; x)$ случайных величин $X_{\max}$, $X_{\min}$ находятся следующим образом:

– моды $a_{\max} = x_2$, $a_{\min} = x_1$ – в результате решения уравнения

$$\beta x - (\alpha - 1) \ln x = \ln \left[n \cdot \frac{\beta^{\alpha-1}}{\Gamma(\alpha)} \right]; \quad (9)$$

– параметры масштаба – по формуле

$$b = \beta - \frac{\alpha - 1}{a} \quad (10)$$

с подстановкой $a_{\max}$ для $b_{\max}$ и $a_{\min}$ для $b_{\min}$.

Здесь индексы “max”, “min” употреблены для указания соответствия параметров a , b распределениям величин $X_{\max}$, $X_{\min}$.

Плотность вероятностей m -распределения Накагами

$$f(x) = f(m, \Omega; x) = \frac{2m^m}{\Gamma(m)} \cdot \frac{x^{2m-1}}{\Omega^m} e^{-\frac{m}{\Omega} x^2}, \quad x > 0, \quad (11)$$

где m – параметр формы, Ω – параметр масштаба.

Для распределений экстремальных значений:

– моды находятся как соответствующие корни $a_{\max} = x_2$, $a_{\min} = x_1$ уравнения

$$\frac{m}{\Omega} x^2 - 2(m-1) \ln x = \ln \left[\frac{n}{2\Gamma(m)} \left(\frac{m}{\Omega} \right)^{m-1} \right]; \quad (12)$$

– параметры масштаба – по формуле

$$b = 2 \left(\frac{m}{\Omega} a - \frac{m-1}{a} \right) \quad (13)$$

с подстановкой $a_{\max}$ для $b_{\max}$ и $a_{\min}$ для $b_{\min}$.

Отметим, что с ростом n ошибки в оценке параметров a и b уменьшаются и становятся незначительными. При малых n они могут не удовлетворять требованиям точности. В этих случаях возможна коррекция приведенных зависимостей.

Выражения (9), (10) применялись для вычисления моментных характеристик времени выполнения параллельных операций при судоремонте в случае допущения о подчинении времени выполнения каждой операции гамма – распределению; выражения (12), (13) – для обоснования динамического диапазона системы обработки сигналов РЛС при облучении морских объектов. Точность вычислений оценивалась сравнением с результатами численного интегрирования с использованием исходных формул (3) и (4), требующего больших вычислительных затрат.

Результаты исследования свидетельствуют о достаточной для практики точности расчетов. На рисунке 1 изображены зависимости математических ожиданий m времени выполнения параллельных операций от их числа n (m_q – численное интегрирование, m_A – аналитический расчет) и дисперсий D этого времени от n (при такой же индексации).

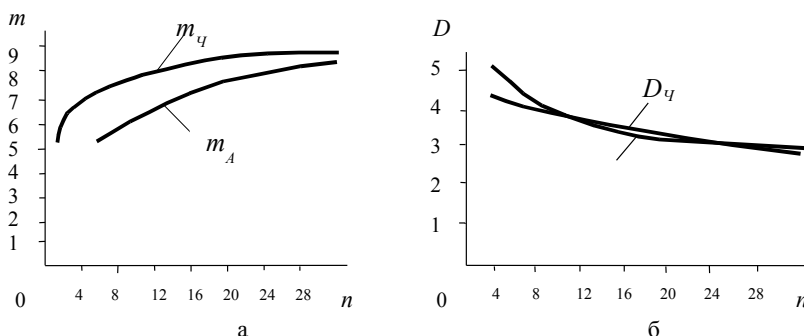


Рисунок 1 – Зависимости математических ожиданий m времени выполнения параллельных операций от их числа n и дисперсий D этого времени от n при $\alpha=4$ и $\beta=1$

Размах выборки определим как $\Delta x_p = x_2 - x_1$, где x_1 и x_2 – решение уравнений $F_M(x_2) = P$ и $F_\mu(x_1) = 1 - P$ соответственно, а P – вероятность попадания максимального (минимального) выборочного значения в интервал (x_1, x_2) . При оптимизации производственных процессов Δx_p может быть принят в качестве целевой функции с минимизацией ее значения.

Библиография

- 1 Справочник по специальным функциям / Под ред. М. Абрамовича, И. Стеган. — М.: Наука, 1979. — 832 с.
- 2 Хастингс Н. Справочник по статистическим распределениям / Н. Хастингс, Дж. Пикок; Пер. с англ. А. К. Звонкина. — М.: Статистика, 1980. — 95 с.
- 3 Смирнов Н.В. Курс теории вероятностей и математической статистики / Н.В. Смирнов, И.В. Дунин-Барковский. — М.: Наука, 1969. — 512 с.

Поступила в редакцию 30.03.2000 г.