An Algorithm of Calculation of Products with Very Large Number and Its Application to Calculation of Discrete Probability

Tadashi Nakamura

Okayama University of Science, Department of Mathematical Information Science

1-1 Ridai-cho, Okayama 700-0005, Japan

naktad@mis.ous.ac.jp

Yasuhisa Hirai

Okayama University, Faculty of Education

3-1-1 Tsushima-naka, Okayama 700-8530, Japan

qdca1385@cc.okayama-u.ac.jp

It is well known that, for representation and arithmetic operations of real numbers by computer, IEEE 754 standard for binary floating-point arithmetic was proposed(Kahan and Palmer 1979, Stevenson et al. 1981, IEEE 1985). It is also pointed out there are a few demerits: (1) Data are not convertible between computers with different format;(2) Calculation based on this standard may cause a trouble called overflow/underflow problem. To avoid these troubles, many reseachers proposed a representation of real numbers and arithmetic operations. There are two kinds of representation. The first is a representation based on a modified version of binary floating-point representation(Coonen 1981, Morris 1971, Mastui and Iri 1980). Especially, Mastui and Iri (1980) proposed a closed overflow/underflow-free floating-point number system. Their system with the 64-bit format and the level 0 can represent numbers of which absolute value is, approximately, less than 2.8×10^{16}. The second is a different representation from the floating-point one. Hamada(1981,1984) proposed a data length independent real number representation based on double exponential cut. This is called a universal representation of real numbers(URR). This representation has merit and demerit which the floating-point number system has not(Yokoo 1989). To conquer demerit, they developed URR to a representatin of numbers based on multiple exponential cut(Yokoo 1989, Tomimastu and Kanada 1998). All representation stated above are designed to treat numbers on hardware. For this reason, there is a limit on the format of the data type, that is, the 32-bit format, the 64-bit format, etc.

On the other hand, our proposed representation of numbers is designed to treat numbers on software. We adopt a representation of numbers called a normalized t-digit floating-point numbers in base 2 with exponent of order k, abbrebiated generalized floating-point numbers of order (t,k). Signed zero, signed infinities and zero can be also represented by generalized floating-point number. Note that we can take t and k possibly large so far as the hardware system permit. Four basic arithmetic operations(addition, subtraction, multiplication, devision) are defined for the calss of generalized floating-point numbers. Then, we have a closed overflow/underflow -free number system. Furthermore, this system can treat a very large number, for example, thirty billion-figure numbers in the decimal system without the multiple-precision calculation of numbers .

As an application, we will calculate the value of discrete probability function and the value of discrete probability distribution function by using the representation of generalized floating-point number and arithmetic operations stated above. Especially, we will calculate probability of binomial probability $B(n,p)$ for large value of n and various values of p, and compare numerical results with those of Johnson, Kotz and Kemp(1992) and Ghosh(1980). Detailed numerical results will be given in presentation.

References

Coonen, J. T.(1981), Underflow and denormalized numbers, Computer, Vol. 14, No.3,51 – 62.

Ghosh, B. K.(1980), Two normal approximation to the binomial distribution, Commun. Statist.-Theor. Meth., A9, 427 – 438.

Hamada, H.(1981), Data length independent real number representation based on double exponential cut(in Japanese), Trans. IPS Japan, Vol. 22, No. 6, 521 – 526.

Hamada, H.(1984), An evaluation of universal representation of real numbers: A new representation of numbers(in Japanese), Computer Software, Vol. 1, No.3, 35 – 43.

IEEE standard for binary floating-point arithmetic, ANSI/IEEE Std. 754, 1985.

Johnson, N. L., Kotz, S. and Kemp, A. W.(1992). Univariate discrete distribution, John Wiley&Sons, New York.

Kahan, W. and Palmer, J.(1979), On a proposed floating-point standard, ACM SIGNUM Newsl., Special Issue, 13 – 21.

Morris, R.(1971), Tapered floating point: A new floating-point representation, IEEE Trans. Comput., Vol. C-20, No. 6, 1578 – 1579.

Mastui, S. and Iri, M.(1980), An overflow/underflow-free floating point representation of numbers(in Japanese), Trans. IPS Japan, Vol. 21, No. 4, 306 – 313.

Stevenson, D. et al.(1981), A proposed standard for binary floating-point arithmetic, Draft 8.0 of IEEE, Computer, Vol. 14, No.3, 51 – 62.

Tomimastu, T. and Kanada, Y.(1998), A generalized numerical representaion based on double exponential cut(in Japanese), Trans. IPS Japan, Vol. 39, No. 3, 511 – 518.

Yokoo, H.(1989), A class of number representation system based on multiple exponential cut(in Japanese), IEICE Trans. A, Vol. J72-A, No. 12, 1998 – 2004.