

Parallel and distributed computing in a statistical system Jasp

Yoshikazu Yamamoto and Junji Nakano

Modern statistical data analysis often requires huge computations. Parallel and distributed computing is the technique to realize them. However, existing software technologies for parallel computing such as OpenMP and MPI are not easy to use for statisticians who are accustomed to “high-level” statistical languages, because such technologies are mainly designed for “low-level” general purpose languages.

Therefore, we have designed and implemented the parallel and distributed computing functions in our statistical analysis Jasp (Java based Statistical Processor).

We designed the parallel computing abilities mainly considering ease of use. Remote Jasp servers can share all global objects which include functions and variables on the Jasp Main Server. We decided not to provide “low level” functions, for example, a separate function for sending data to remote servers. Because many statistical calculations can be divided rather simply and naturally into independent sub-tasks, and it is sufficient to send data implicitly as arguments of the function executed on a remote server.

The parallel computing abilities in Jasp are suitable for utilizing several heterogeneous or homogeneous computers connected by networks, because Jasp is written in the Java language. The Jasp system and parallel computing programs written in the Jasp language can be executed on any computer that supports the Java virtual machine. We have implemented the communication among computers by Java RMI and the load-balancing by JavaSpaces which is the virtual shared memory system using Java.