
ARASH JAMSHIDPEY, University of Ottawa

Stochastic DCJ jump process on signed permutation groups and the validity of the median as an approximation to the true ancestor

A genome can be represented by its set of gene adjacencies. Each gene is denoted by a signed number where the sign indicates the orientation. A double-cut-and-join (DCJ) operation acts on two adjacencies of the genome, xy and wz , where x, y, w, z are heads or tails of some genes, cuts these adjacencies and rejoins the four points in two possible ways, either xw, yz or xz, yw . The *dcj* distance between two genomes is the minimum number of DCJs necessary to convert one into the other. A DCJ jump process on the space of all multichromosomal genomes is a continuous time version of a DCJ random walk where at random Poisson times we uniformly randomly choose one of all possible DCJs and let it act on the current genome state.

Suppose we have k independent DCJ jump processes, X_t^1, \dots, X_t^k , all starting at the same genome X_0 . Our goal is to study the median value of $G_t := \{X_t^1, \dots, X_t^k\}$, namely $\min_M \sum_i dcj(M, X_t^i)$. We prove that as the number of genes n goes to ∞ , the median value approximates the total divergence time kt as well as the total *dcj* distance of the true ancestor to G_t , namely $\sum_i dcj(X_0, X_t^i)$, if the number of rearrangements $t < n/4$. Furthermore, we investigate when this approximation does not hold for $t > n/4$. We make use of a new algebraic representation of DCJ which enables us to show that the state space of the process has a group structure.