

1159-60-129

Ming-Ying Leung* (mleung@utep.edu). *Scan statistics for finding inversion clusters in genomic sequences.*

Scan statistics have been applied in various biomedical studies to identify genomic regions containing nonrandom clusters of chromosomal translocation breakpoints, viral DNA integration sites, or other specific word patterns in molecular sequences. In these applications, the positions of the sites of interests are modeled as independent and identically distributed random points on the unit interval. The scan statistics are then defined either as the maximum count of points contained in a scanning window of fixed length or the minimum aggregated spacing between a fixed number of consecutive points. We have applied the scan statistics to characterize palindrome clusters in DNA and search for inversion clusters in RNA sequences. An inversion is a string of nucleotide bases followed closely by its inverse complementary sequence with a gap of g intervening nucleotides in between, and a palindrome is a special case of inversions with $g = 0$. Several palindrome clusters found on herpesvirus genomes are associated with viral replication origins. Inversion clusters with g ranging from 0 to 8 have also been identified in genomic RNAs of coronaviruses. Such inversions are essential for RNA secondary structure prediction, which is a computationally intensive problem for long RNA sequences. (Received August 03, 2020)