Ann Inst Stat Math (2017) 69:497–512
DOI 10.1007/s10463-015-0551-8

The *m*th longest runs of multivariate random sequences

Yong Kong[1]

Received: 14 April 2015 / Revised: 19 November 2015 / Published online: 1 February 2016

Abstract The distributions of the *m*th longest runs of multivariate random sequences are considered. For random sequences made up of k kinds of letters, the lengths of the runs are sorted in two ways to give two definitions of run length ordering. In one definition, the lengths of the runs are sorted separately for each letter type. In the second definition, the lengths of all the runs are sorted together. Exact formulas are developed for the distributions of the *m*th longest runs for both definitions. The derivations are based on a two-step method that is applicable to various other runs-related distributions, such as joint distributions of several letter types and multiple run lengths of a single letter type.

Keywords Generating function · Combinatorial identities · Randomness test · Distribution-free statistical test · Runs length test · Biological sequence analysis

✉ Yong Kong
yong.kong@yale.edu

[1] Department of Molecular Biophysics and Biochemistry, W.M. Keck Foundation Biotechnology Resource Laboratory, Yale University, 333 Cedar Street, New Haven, CT 06510, USA