Accepted Manuscript

A genetic algorithm for simulating correlated binary data from biomedical research

Jochen Kruppa, Bernd Lepenies, Klaus Jung

PII: S0010-4825(17)30349-9

DOI: 10.1016/j.compbiomed.2017.10.023

Reference: CBM 2816

To appear in: *Computers in Biology and Medicine*

Received Date: 4 July 2017

Revised Date: 20 October 2017

Accepted Date: 20 October 2017

Please cite this article as: J. Kruppa, B. Lepenies, K. Jung, A genetic algorithm for simulating correlated binary data from biomedical research, *Computers in Biology and Medicine* (2017), doi: 10.1016/j.compbiomed.2017.10.023.

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

ACCEPTED MANUSCRIPT

A genetic algorithm for simulating correlated binary data from biomedical research

Jochen Kruppa[a], Bernd Lepenies[b,c], Klaus Jung[a,c,*]

[a] *Institute for Animal Breeding and Genetics, University of Veterinary Medicine Hannover, Hannover, Germany*
[b] *Immunology Unit, University of Veterinary Medicine Hannover, Hannover, Germany*
[c] *Research Center for Emerging Infections and Zoonoses (RIZ), University of Veterinary Medicine Hannover, Hannover, Germany*

Abstract

Correlated binary data arise in a large variety of biomedical research. In order to evaluate methods for their analysis, computer simulations of such data are often required. Existing methods can often not cover the full range of possible correlations between the variables or are not available as implemented software. We propose a genetic algorithm that approaches the desired correlation structure under a given marginal distribution. The procedure generates a large representative matrix from which the probabilities of individual observations can be derived or from which samples can be drawn directly. Our genetic algorithm is evaluated under different specified marginal frequencies and correlation structures, and is compared against two existing approaches. The evaluation checks the speed and precision of the approach as well as its suitability for generating also high-dimensional data. In an example of high-throughput glycan array data, we demonstrate the usability of our approach to simulate the power of global test procedures. An implementation of our own and two other methods were added to the R-package 'RepeatedHighDim'. The presented algorithm is not restricted to certain correlation structures. In contrast to existing methods it is also evaluated for high-dimensional data.

*Corresponding author. Institute for Animal Breeding and Genetics, University of Veterinary Medicine Hannover, Bünteweg 17p, D-30559 Hannover, Germany
Email address: klaus.jung@tiho-hannover.de (Klaus Jung)

Download English Version:

https://daneshyari.com/en/article/6920683

Download Persian Version:

https://daneshyari.com/article/6920683

Daneshyari.com