

**A comparative evaluation of goodness-of-fit tests for the negative binomial distribution
with application to RNA-Seq**

Data

John Osumba, Collins Odhiambo and Bernard Omolo

Institute of Mathematical Sciences, Strathmore University, Nairobi, Kenya.

University of South Carolina — Upstate, United States of America.

The negative binomial (NB) distribution is considered the most appropriate distribution for modeling over dispersed count data. In this regard, a number of goodness of-fit (GOF) tests have been applied for the NB, but no systematic evaluation of these tests has been done to determine the most powerful tests. In this study, we perform a comparative evaluation of the GOF tests for the negative binomial distribution, based on their power under suitable alternatives, via simulations. The tests considered here include those based on the empirical distribution functions (EDFs), likelihood functions, Kullback-Leibler discrimination information, Laplace transforms, the generalized smooth tests and a combination of tests with complimentary behavior. For illustration and validation, RNA-Seq data from colorectal cancer are used.

Keywords: GOF tests; model selection; negative binomial distribution; power; RNA-Seq data.