11th Conference on Dynamical Systems Applied
to Biology and Natural Sciences DSABNS 2020
Trento, Italy, February 4-7, 2020

INFERENCE OF UNSEEN GENETIC VARIANTS FROM LOCAL PRESENCE/ABSENCE INFORMATION

Anna Fochesato[1,2], Anna Tovo[3,4],
Stefano Lise[5] and Marco Formentin[*3]

[1]Fondazione The Microsoft Research
University of Trento Centre for Computational and
Systems Biology (COSBI), Italy

[2]Department of Mathematics,
University of Trento, Italy

[3]Department of Mathematics "Tullio Levi-Civita",
University of Padova, Italy

[4]Department of Physics and Astronomy "Galileo Galilei",
University of Padova, Italy

[5]Centre for Evolution and Cancer,
The Institute of Cancer Research, London, United Kingdom

marco.formentin@unipd.it (*corresponding author),
fochesato@cosbi.eu

In the last years, the development of new DNA sequencing technologies has provided researchers with a massive amount of data. However, extrapolating useful information and patterns from these data remains a huge challenge. In this context, statistical tools play a central role in giving insights into the genetic human profile and its mutational heterogeneity. The accumulation of DNA variants is recognized to fuel tumor occurrence, thus the need for developing methods to quantify mutation abundance and variability.

To this end, we propose a statistical framework to predict the number of mutations in a DNA sequence or in a whole tumor (global scale) starting from presence/absence information collected in certain samples (local scale). By mapping inference of unseen mutations into the unseen species problem in biodiversity, we could tackle genomic inference with tools from statistical ecology.

In details, we choose a Negative Binomial (NB) family of distributions to describe the probability for a mutation to occur in n samples. The choice of a NB family is due to the following properties that represent the key points of our method. First, NB distributions, for different ranges of parameters, can capture both exponential and power-law tails. Second, under the hypothesis of spatial homogeneity, random samplings of a NB can still be described via a NB. The latter property, named form invariance, results in a computable formula that bridges NB parameters at both local and global scales and that we use to determine an unbiased and consistent estimator of the mutation abundance.

We validate our framework on both DNA single-nucleotide polymorphism and synthetic spatial tumor

 ISBN: 978-989-98750-7-4

growth datasets. We fit the empirical curves at local scale obtaining the local parameters for NB, and we use the form invariance property to retrieve global parameters. The results suggest the stability of the proposed method, displaying a positive correlation between the accuracy of prediction and the local scale size. Future perspectives may include this framework in the context of hybrid modelling. In this way, we could improve our predictions by coupling our approach with a dynamical representation of the tumor growth that tracks the mutation heterogeneity.

Acknowledgements
The work of Anna Fochesato was developed as Master student at Department of Mathematics Tullio Levi-Civita, University of Padova, Italy.

 ISBN: 978-989-98750-7-4