

## Distribution of the number of carrier genotypes in Mendelian models

A. Lefebvre<sup>1</sup>, G. Nuel<sup>2</sup>.

(1) Laboratoire Jacques-Louis Lions (LJLL, UMR 7598), 75005, Paris, France, [alexandra.lefebvre@math.cnrs.fr](mailto:alexandra.lefebvre@math.cnrs.fr).

(2) Laboratoire de Probabilité, Statistiques et Modélisation, (LPSM, UMR 8001), 75005, Paris, France.

Mendelian models are based on Bayesian networks for them to be particularly suited for analyzing a family history of disease. They allow indeed for modeling the structure dependency between genotypes (usually latent) and phenotypes (usually observed) of family members. That structure dependency can be used to reduce the computational complexity of an inference via the sum-product algorithm [1, 2, 3].

Given a genetic model (allele frequencies of a major gene, mode of inheritance, disease-specific penetrance per genotype, etc.), genetic counselors are usually interested in computing the marginal posterior probability of carrying a deleterious allele for an individual of interest and his/her resulting probability of developing the disease in the future. Along with these marginal posterior probabilities, one may also be interested in familial risks, in particular the distribution of the number  $N$  of carriers within a family.

In this work, we address this question following the idea of introducing probabilistic relationships between variables through polynomials for computing generating functions in probabilistic graphical models [4,5]. In particular we introduce specific polynomials for computing the probability generating function (pgf) of  $N$ . From the pgf of  $N$ , one can derive marginal posterior carrier probabilities for an individual or a group of individuals conditional on the family history and  $N=k$  carriers.

We illustrate the method over various simulated family histories using different genetic models estimated in the framework of the breast/ovarian and the Lynch syndrome. We show the interest of considering the distribution of the number of carriers in posterior risks inference for highlighting at-risk individuals and helping clinicians in prioritizing genetic investigations.

### References

- [1] R. C. Elston and J. Stewart. *A general model for the genetic analysis of pedigree data*. Human heredity, 21(6):523–542, (1971).
- [2] S. L. Lauritzen and N. A. Sheehan. *Graphical models for genetic analyses*. Statistical Science, pages 489-54, (2003).
- [3] D. Koller and N. Friedman, *Probabilistic graphical models: principles and techniques*. MIT press, (2009).
- [4] R. G. Cowell. *Calculating moments of decomposable functions in Bayesian networks*. Research Report 109, Department of Statistical Sciences, University College London, (1992).
- [5] A. Lefebvre and G. Nuel. *A sum-product algorithm with polynomials for computing exact derivatives of the likelihood in Bayesian networks*. In Proceedings of Machine Learning Research, International Conference on Probabilistic Graphical Models, pages 201–212, (2018).