A geometric tutorial on exchangeability and related topics

P.G.L. Porta Mana
<piero.mana@ntnu.no>

Draft of 3 March 2019 (first drafted 23 June 2018)

A geometric tutorial on exchangeability
Note: Dear Reader & Peer, this manuscript is being peer-reviewed by you. Thank you.

> Most books on modern physics assume a knowledge of probability theory and claim to derive probability distributions. A casual observer might be led into supposing that the authors would be reasonably up to date in the subject. In fact, however, they seldom show any knowledge of it that is less than fifty years old
>
> H. Jeffreys (1955 p. 275)

> I believe that the greatest unsolved problem in statistics is communicating the subject to others.
>
> N. Davies (in Copas et al. 1995 p. 445)

1 Introduction

Exchangeable models are a family of probability models that appear over and over in all sciences. Notions and constructs such as: Gibbs ensembles, unsupervised learning, quantum theory, maximum-entropy methods can be understood and studied as particular applications of exchangeable models. A thorough knowledge of these models therefore gives us a unified view of a variety of seemingly disparate topics.

Many scientists therefore use exchangeable models routinely without knowing their technical name, but more importantly without knowing about their special properties. This ignorance is mainly caused by a language barrier: most literature on exchangeable models uses very specialized language strongly connected with topology, measure theory, differential geometry, and other subjects that are usually superficially touched in the teaching of other sciences. Such subjects are of course necessary for a precise foundation of exchangeable models and other probabilistic notions. Yet, many technical concepts underlying the probability calculus have a very intuitive and even visual meaning at bottom.

The present notes try to present exchangeable models and most notions related to them in an intuitive and visual way. Connections will be made to all the notions and construct mentioned above.

$\leqslant$

to build a bridge, explaining the essentials of exchangeable models and some related topics in an intuitive and geometric way, hoping to make the main concepts more accessible to scientists in other disciplines. I'm not a statistician or a specialized scientist, so my attempt will likely leave all parties unsatisfied. I hope it will at least give a hunch that a bridge is worthwhile constructing.

The other topics discussed in this note include:

- the choice of so-called 'parameter priors'
- so-called 'non-parametric models'
- model comparison
- sufficiency and its relation to some toy physical models, like Ising models
- partial exchangeability and its relation to machine learning and neural networks
- the phenomenon of 'overtraining'

The mathematics I'll use is very simple: functional analysis and combinatorics. The geometry will rely on convex spaces. These spaces appear in all of statistics. Their basics are very intuitive, and their visualization often allows us to guess mathematical properties or solutions to problems that are harder to guess algebraically. For this reason this note starts with a summary about them.

2 Exchangeability

There's a deep reason why exchangeable models appear in all sciences: they are connected with *reproducibility*, and therefore also with *induction*.

Reproducibility and circularity.

Exchangeability expresses the way we do induction (doesn't justify it).

Example with pattern: we discard exchangeability (dead hypothesis resurrection) – just as we do with reproducibility.

3 Convex spaces

4 Finitely exchangeable models

5 Exchangeable models and their update

maximum-entropy appearance.

6 Probability distributions for the limit frequencies

7 Model comparison

8 Models with sufficient statistics

9 Partially exchangeable models

to be continued

Bibliography

('de *X*' is listed under D, 'van *X*' under V, and so on, regardless of national conventions.)

Chatfield, C. (1995): *Model uncertainty, data mining and statistical inference*. J. Roy. Stat. Soc. A **158**^{3}, 419–444. See also discussion in Copas, Davies, Hand, Lunneborg, Ehrenberg, Gilmour, Draper, Green, et al. (1995).

Copas, J. B., Davies, N., Hand, D. J., Lunneborg, C. E., Ehrenberg, A. S., Gilmour, S. G., Draper, D., Green, P. J., et al. (1995): *Discussion of the paper by Chatfield [Model uncertainty, data mining and statistical inference]*. J. Roy. Stat. Soc. A **158**^{3}, 444–466. See Chatfield (1995).

Jeffreys, H. (1955): *The present position in probability theory*. Brit. J. Phil. Sci. **5**^{20}, 275–289.