Discrete Factor Analysis, with an example from Musicology

Rolf Larsson[1] **and Jesper Rydén**[2]

[1] *Uppsala University, Sweden, rolf.larsson@math.uu.se*
[2] *Swedish University of Agricultural Sciences, Sweden, jesper.ryden@slu.se*

In this talk, we present the discrete factor analysis method of [2], and apply it to a data example from Musicology, see [3].

The method of [2] extends the dependent Poisson model of [1] to construct a discrete factor analysis model. The model parameters are estimated by maximum likelihood. The model with the lowest AIC is searched for by employing a forward selection procedure. The probability to find the correct model is investigated in a simulation study.

To be able to properly analyze the Musicology data, we must take over dispersion into account. This we do by replacing the Poisson distribution with the Negative Binomial.

References

[1] Karlis, D. (2003). An EM algorithm for multivariate Poisson distribution and related models. *Journal of Applied Statistics*, **30**, 63 – 77.

[2] Larsson, R. (2020). Discrete factor analysis using a dependent Poisson model. *Computational Statistics*, **35**, 1133 – 1152.

[3] Rydén, J. (2020). On features of fugue subjects. A comparison of J.S. Bach and later composers. *Journal of Mathematics and Music*, **14**, 1 – 20.

Numbers, tables and statistics

Kajsa Møllersen[1]

[1] *UiT The Arctic University of Norway, Norway, kajsa.mollersen@uit.no*

The old "Lofoten Post" building sits at the end of the dock in Svolvær as an empty shell, ready for the winter storms to seal its destiny. Everything that reminds of newspaper production is stripped from the building, only dents from the editor's desk can still be seen on the floor. Wait! A stack of newspapers, filled with tables of fishery, trade, crimes, and everything else that transforms small coastal societies from January to April every year, when the Atlantic Cod comes to spawn in Lofoten. Numbers and tables, tables and numbers.

The stack is not left over from the glory days of *Lofotposten*, but a fresh newspaper printed as part of Lofoten International Art Festival (LIAF) 2019, and the whole building is filled with contemporary art. I have been invited to participate with a talk for the session "Maths, matter & body", followed by a conversation with Toril Johannessen, one of the artists. Her contribution to the main exhibition is works from "Words and Years" (2010-16), a series of graphs based on data from scientific journals and news magazines [1]. The graphs are lovely to look at, their colours are pleasing, the paper they are printed on is thick, almost like cardboard. The information they provide is funny, sometimes sad, and surprisingly interesting. My thoughts go to Florence Nightingale - nurse-statistician with beautiful graphs.

My contribution to LIAF is not numbers and tables or coloured graphs - but the most beautiful part of statistics; mathematics. In my position at the Faculty of health sciences, I do a lot of teaching to students that have weak mathematical background, and have a sincere and deep fear for both proper and complex fractions. I teach statistics without mathematics, not to loose my students into despair and hopelessness. At LIAF, I'm hoping to meet an audience that are curious and comfortable with what they don't understand, and with a sense of aesthetics in general that can open the door into mathematical beauty.

I open my talk with a picture of a Mandelbrot set, all beautiful like a distant galaxy of stars. I quickly let the screen go black, and start talking about Euler's identity. Then the number 0. I slowly build up the proof for $\sqrt{2}$ being irrational, and the talk climaxes with completion of the proof, and I can hear the audience draw a sigh of relief.

Later, when we all just hang around drinking beer - not unlike a small, scientific conference - an artist walks up to me and says "All these tables and numbers - and then statistics was something completely different all along". I feel as if he has understood more about statistics than many of my students will ever do.

References

[1] https://www.toriljohannessen.no/works/words-and-years/

In which order did Platon write Critias, Philebus, Politicus, Sophist, and Timaeus?

Nils Lid Hjort

University of Oslo, Norway, nils@math.uio.no

Perhaps you, like Plato, or Platon, for stylistic or rhetorical reasons, are very careful with your sentence endings. In this case your clausula, the five last syllables, is an instance of L S L S L. There are $2^5 = 32$ variations, of 'light' and 'stress'. Corpora can be carefully read and analysed, with the different types of clausulae tabulated and compared. In probability modelling language, we then get estimates of each work's 32-dimensional clausula probability vector, say $p = (p_1, \ldots, p_{32})$.

For the case of Platon (c. 424-347 B.C.), scholars agree that A = Republic is from his middle period, and B = Laws among his last works; these give rise to firmly different probability vectors p_A and p_B. Experts do not know in which order the Socratic dialogues Critias, Philebus, Politicus, Sophist, Timaeus were written, however. I shall attempt to order these five, inside the time window from A to B, via modelling and estimating the five probability vectors $p_{\rm I}, p_{\rm II}, p_{\rm III}, p_{\rm IV}, p_{\rm V}$ on a bridge, in the world of 32-dimensional probability vectors, from p_A to p_B.

Cox and Brandwood (JRSS B, 1959) analysed such data, with a similar aim, but I shall use modern model selection methods, from Claeskens and Hjort (Model Selection and Model Averaging, Cambridge, 2008) to build more sophisticated models and methods.