A stochastic optimisation algorithm for pairwise likelihood estimation of ordinal factor models

Giuseppe Alfonzetti[1], Ruggero Bellio[2], Yunxiao Chen[3], Irini Moustaki[3].

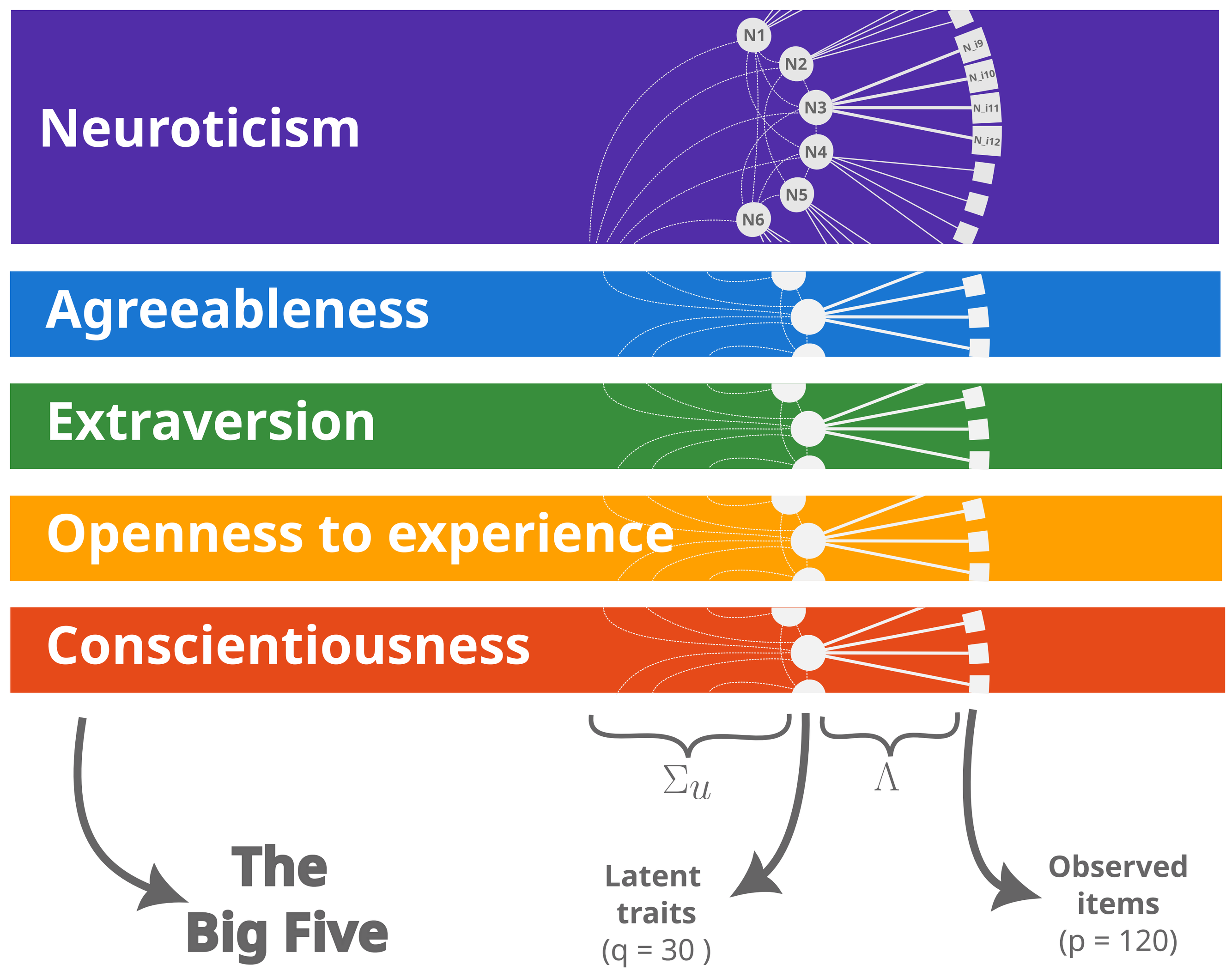
[1]University of Padua, [2]University of Udine, [3]London School of Economics.

Summary

Motivated by a big web-based personality test, we develop a stochastic optimisation algorithm that exploits pairwise likelihood structure to scale ordinal factor models to large datasets.

The Big Five Personality test

- Large-scale web-based test designed to measure 5 personality areas [1]: Neuroticism (N), Agreeableness (A), Extraversion (E), Openness to experience (O) and Conscientiousness (C).
- Each area can be further split in 6 personality facets, for a total of 30 mutally correlated latent traits to account for.
- The dataset consists of answers to 120 items on a five-category rating scale, observed on more than 600 thousands units.



Pairwise Likelihood

The pairwise log-likelihood [2] is constructed exploiting the underlying response variable parameterisation, with:

$$p\ell(\theta)=\sum_{j<j'}\ell_{jj'}(\theta)=\sum_{j<j'}\sum_{s_j,s_{j'}} n^{jj'}_{s_js_{j'}}\log\pi^{jj'}_{s_js_{j'}} \quad (1)$$

where $n^{jj'}_{s_js_{j'}}$ is the observed frequency of the specific bivariate pattern on columns j, j', while its marginal probability is defined as

$$\pi^{jj'}_{s_js_{j'}}=\int_{\tau_{s_j-1}}^{\tau_{s_j}}\int_{\tau_{s_{j'}-1}}^{\tau_{s_{j'}}}\phi_2(y_j^*,y_{j'}^*;\rho^{y^*}_{jj'})dy_j^*dy_{j'}^*; \quad (2)$$

where $\rho^{y^*}_{jj'}$ is the model correlation between variables j and j', computed via $\lambda_j^T\Sigma_u\lambda_{j'}$.

Pros and cons

+ It substitutes large-dimensional integration problems with bivariate ones;
+ It reduces data by sufficiency;
- Its computational cost grows with $O(p^2)$;

Details on the Underlying response variable (URV) parameterisation

Assumptions:

1. Ordinal responses, $Y_j = s_j \in \{0,\ldots,c_j-1\}$.
2. Data are partial observation of an underlying response normally distributed y^*, such that

$$Y_j = s_j \iff \tau^{(j)}_{s_j-1} < y^*_j < \tau^{(j)}_{s_j},$$

and $y^*=\Lambda u+\delta$ with $\Sigma_\delta = I_p - \mathrm{diag}(\Lambda\Sigma_u\Lambda^T)$ and $\delta\sim\mathcal{N}_p(0,\Sigma_\delta)$.

Parameters:

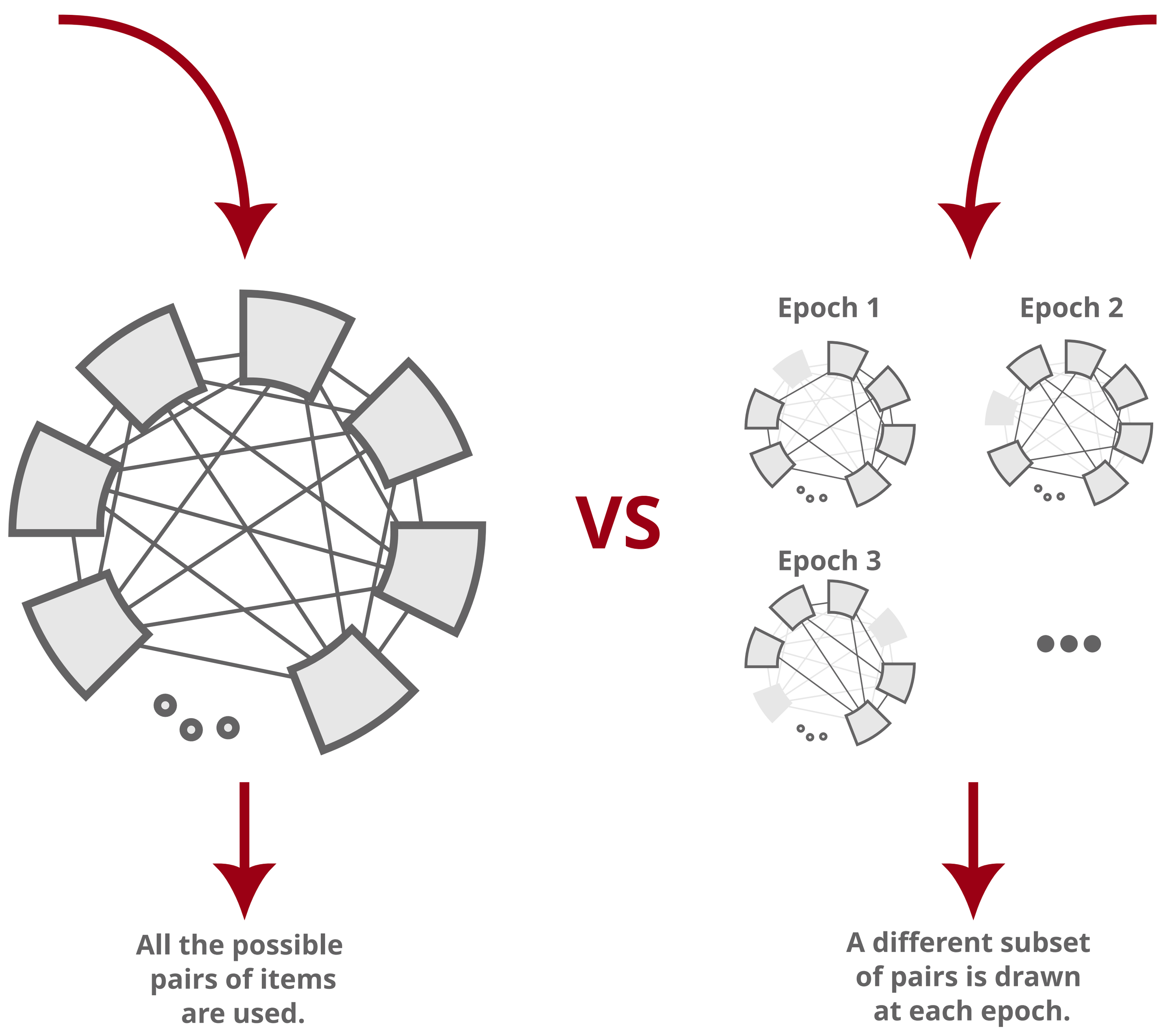
- Loading matrix Λ, latent correlation matrix Σ_u and thresholds vector τ, where $\tau=(\tau^{(1)^T},\ldots,\tau^{(p)^T})^T$ with $\tau^{(j)}=(\tau^{(j)}_0,\ldots,\tau^{(j)}_{c_j-2})\in\mathbb{N}^{c_j-1}$;
- $\theta\in\mathbb{R}^d$ collects parameters from Λ, Σ_u and τ.

Likelihood:

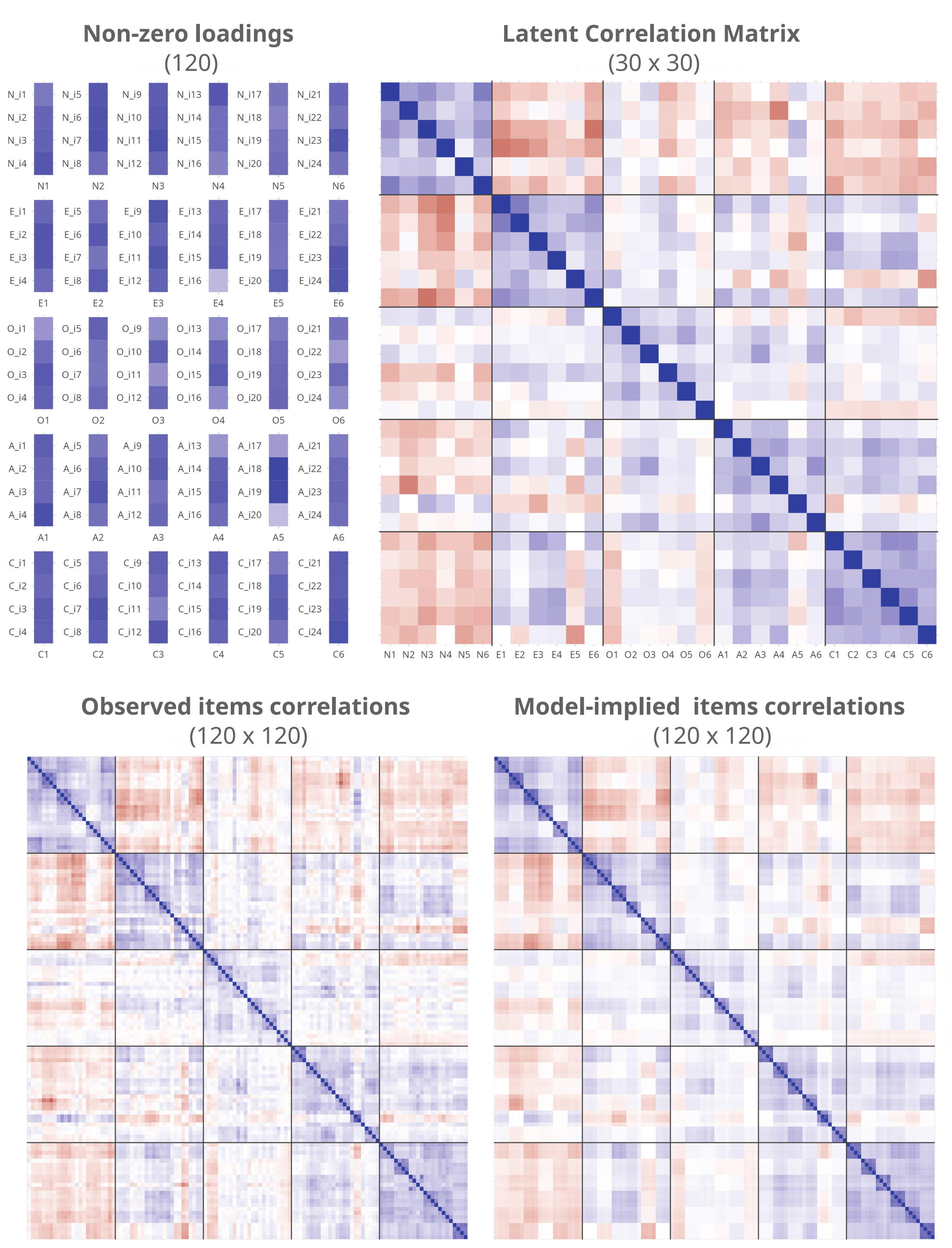
The data marginal likelihood is given by

$$\ell(\theta;y)=\sum_{i=1}^n\log\int_{\tau^{(1)}_{s_1-1}}^{\tau^{(1)}_{s_1}}\cdots\int_{\tau^{(p)}_{s_p-1}}^{\tau^{(p)}_{s_p}}\phi_p(y_i^*;\Sigma_{y^*})dy_i^*. \quad (3)$$

where $\phi_p(x;\Sigma)$ is the density of a p-dimensional normal distribution evaluated at x, with mean zero and variance Σ.



Big Five data results



Notation

- n units, p observed items, q latent variables;
- Random $n\times p$ matrix of manifest variables Y, with $Y_i=(Y_{i1},\ldots,Y_{ip})\in\mathbb{R}^p$ for $i=1,\ldots,n$. Realizations y, with $y_i=(y_{i1},\ldots,y_{ip})\in\mathbb{R}^p$;
- Random $n\times q$ matrix of latent variables U, with $U_i=(U_{i1},\ldots,U_{iq})\in\mathbb{R}^q$ for $i=1,\ldots,n$. Realizations u, with $u_i=(u_{i1},\ldots,u_{iq})\in\mathbb{R}^q$ and $u_i\overset{iid}{\sim}\mathcal{N}_q(0,\Sigma_u)$;
- Latent covariance matrix $\Sigma_u\in\mathbb{R}^q\times\mathbb{R}^q$. Constrained to be a correlation matrix.
- Loading matrix $\Lambda=(\lambda_1^T,\ldots,\lambda_p^T)\in\mathbb{R}^p\times\mathbb{R}^q$ with $\lambda_j=(\lambda_{j1},\ldots,\lambda_{jq})$, $j=1,\ldots,p$.

Stochastic Optimisation

- Define a stochastic approximation to the gradient via

$$\nabla p\ell(\theta,y)\approx\nabla f(\theta;y,w)=\gamma^{-1}\sum_{j<j'}w_{jj'}\nabla\ell_{jj'}(\theta).$$

- The quantities $w_{jj'}$ are random binary weights such that $w_{jj'}\overset{iid}{\sim}\mathrm{Bernoulli}(\gamma)$.
- Note that, if $\gamma=1$, we retrieve $\nabla f(\theta;y,w)=\nabla p\ell(\theta)$. If $\gamma\neq 1$ we still have $E_w\{\nabla f(\theta;y,w)\}=\nabla p\ell(\theta)$.

The algorithm

The generic t-th epoch alternates [3]:

1. Stochastic step: Sample a new set of weights $w^{(t)}$;
2. Approximation step: Compute $\nabla f(\theta_{t-1};y,w^{(t)})$;
3. Update step: Update θ_t via $\theta_t=\theta_{t-1}+\eta_t\nabla f(\theta_{t-1};y,w^{(t)})$, where $\eta_t=\eta t^{-.5+\epsilon}$, such that $\sum_{t=1}^\infty\eta_t=\infty$ and $\sum_{t=1}^\infty\eta_t^2<\infty$.

At the end of the optimization, trajectories are averaged via $\bar\theta=T^{-1}\sum_t^T\theta_t$.

Main Findings

- The proposed framework allows to arbitrarily decrease the complexity per iteration considering only a subset of the pairs. Practitioners can choose the complexity per iteration according to their hardware/time constraints, similarly to what happens when tuning the size of a mini-batch SGD.
- A large-scale factor model application is provided with the Big Five Personality Test, allowing psychometricians to both calibrate items and estimate the latent correlation structure at the same time.
- Future work will focus on:
  - Extending the algorithm to proximal updates [3];
  - Generalizing to the large class of composite likelihood functions while developing the appropriate inference tools.

References

[1] John, A., Johnson. (2014). Measuring thirty facets of the Five Factor Model with a 120-item public domain inventory: Development of the IPIP-NEO-120, *Journal of Research in Personality*, **51**, 78 – 89.

[2] Katsikatsou, M., Moustaki, I., Yang-Wallentin, F. and Jöreskog, K. G. (2012). Pairwise likelihood estimation for factor analysis models with ordinal data. *Computational Statistics & Data Analysis* **56**, 4243 – 4258.

[3] Zhang, S., Chen, Y. (2022). Computation for Latent Variable Model Estimation: A Unified Stochastic Proximal Framework. *Psychometrika*.