



Universiteit
Leiden
The Netherlands

Learning under constraints: inference, selection and validation with maximum entropy models

Giuffrida, F.

Citation

Giuffrida, F. (2026, September 10). *Learning under constraints: inference, selection and validation with maximum entropy models*. Retrieved from <https://hdl.handle.net/1887/4309386>

Version: Publisher's Version

License: [Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden](#)

Downloaded from: <https://hdl.handle.net/1887/4309386>

Note: To cite this publication please use the final published version (if applicable).

Learning Under Constraints

Inference, Selection and Validation
with Maximum Entropy Models

Proefschrift

ter verkrijging van
de graad van doctor aan de Universiteit Leiden,
op gezag van rector magnificus prof.dr. S. de Rijcke,
volgens besluit van het college voor promoties
te verdedigen op donderdag 10 september 2026
klokke 11:30 uur

door

Francesca Giuffrida
geboren te Rome
in 1995

Promotores:

Prof. dr. D. Garlaschelli

Prof. dr. P. D. Grünwald

Promotiecommissie:

Prof. dr. ir. S.J. van der Molen

Prof. dr. K.E. Schalm

Prof. dr. W.T.F. den Hollander

Dr. M. Sales-Pardo (Rovira i Virgili University)

Prof. dr. T. P. Peixoto (Interdisciplinary Transformation University Austria)

Contents

1	Introduction	7
1.1	Maximum entropy models	11
1.2	Model selection	13
1.3	Hypothesis testing	15
1.4	Pattern validation	18
2	Description length of maximum entropy models	23
2.1	Introduction	24
2.2	The Minimum Description Length principle	26
2.3	Description length of canonical and microcanonical models	28
2.4	Applications to maximum-entropy models	33
2.5	NML-based VS Bayesian approach to model selection	38
2.6	Conclusion	41
2.A	Derivation of the microcanonical distribution from entropy maximization	42
2.B	Asymptotic expansions	43
2.C	NML and Bayes	45
3	Testing maximum entropy models with e-values	49
3.1	Introduction	50
3.2	Introduction to E-variables	51
3.3	Application to maximum entropy models	59
3.4	Application to contingency tables and related models	71
3.5	Conclusion	84
3.A	Redundancy and regret	84
3.B	Microcanonical test	85
3.C	Microcanonical approximation	87
3.6	Proof of Theorem 3.3.1	89

3.D	Pseudo approximation through the high resolution limit	91
3.E	Bound on regret	92
3.F	Testing with the NML universal distribution	93
3.G	Supplementary tables and figures	96
4	E-values for contingency tables	103
4.1	Introduction	104
4.2	Preliminaries and notation	108
4.3	Calculating and analyzing e-variables for a single table	114
4.4	Comparison of Turner’s and conditional e-variables in sequential settings	116
4.5	Conclusion	119
5	Validation of higher-order interactions: application to brain data	125
5.1	Introduction	126
5.2	Results	127
5.3	Methods	135
5.4	Discussion and conclusion	143
5.5	Data availability	146
5.A	Canonical null models for hypothesis testing	146
5.B	Approximations of the Poisson-binomial distribution	154
5.C	Atlas	160
6	Conclusion	163
6.1	Model selection and ensemble non-equivalence	163
6.2	E-values and MDL	164
6.3	Conditional and microcanonical e-values	165
6.4	Refining higher-order validation under null models	166
	Bibliography	168
	List of publications	181
	Samenvatting	182
	Summary	185
	Acknowledgements	188

