



**This electronic thesis or dissertation has been  
downloaded from Explore Bristol Research,  
<http://research-information.bristol.ac.uk>**

*Author:*

**Kloucek, Maximilian B**

*Title:*

**Inferring criticality in neural networks**

**General rights**

Access to the thesis is subject to the Creative Commons Attribution - NonCommercial-No Derivatives 4.0 International Public License. A copy of this may be found at <https://creativecommons.org/licenses/by-nc-nd/4.0/legalcode>. This license sets out your rights and the restrictions that apply to your access to the thesis so it is important you read this before proceeding.

**Take down policy**

Some pages of this thesis may have been removed for copyright restrictions prior to having it been deposited in Explore Bristol Research. However, if you have discovered material within the thesis that you consider to be unlawful e.g. breaches of copyright (either yours or that of a third party) or any other law, including but not limited to those relating to patent, trademark, confidentiality, data protection, obscenity, defamation, libel, then please contact [collections-metadata@bristol.ac.uk](mailto:collections-metadata@bristol.ac.uk) and include the following information in your message:

- Your contact details
- Bibliographic details for the item, including a URL
- An outline nature of the complaint

Your claim will be investigated and, where appropriate, the item in question will be removed from public view as soon as possible.

UNIVERSITY OF BRISTOL

DOCTORAL THESIS

---

INFERRING CRITICALITY IN NEURAL  
NETWORKS

---

*Author:*

MAXIMILIAN BENEDIKT KLOUCEK

*Supervisors:*

DR. FRANCESCO TURCI  
DR. THOMAS MACHON  
DR. C PATRICK ROYALL

*A dissertation submitted to the University of Bristol in accordance with  
the requirements for award of the degree of Doctor of Philosophy in the  
Faculty of Science, H. H. Wills Physics Laboratory.*

March 2023

40166 Words





# Abstract

It is often the case that while the specific interactions between individual constituents of complex systems are unknown, their correlations are measurable. Reconstructing the strength of the interactions from these correlations is known as an inverse problem. We focus on reconstructing interactions from functional magnetic resonance imaging (fMRI) studies of the human brain, where different regions of the brain are modelled as binary variables (on vs off). The overarching aim of this work was to analyse these datasets from the perspective of statistical physics and to understand whether the human brain exists at an order-disorder transition, i.e. a critical point. This was motivated by a growing body of evidence which suggests that many complex biological systems are tuned towards criticality. We construct equilibrium statistical physics models of the data via a machine learning scheme termed Inverse Ising inference. We initially show that typical estimators used for inverse Ising inference, such as pseudo-likelihood maximization (PLM), are biased by testing the inference on simulated data. Understanding the performance of the inference on small sample sizes is crucial to interpreting models fitted from real data, as the amount of data available in experimental studies is limited. Using the Sherrington-Kirkpatrick (SK) model as a benchmark, we show that PLM displays large biases in the critical regimes close to order-disorder transitions, which may alter the qualitative interpretation of the inferred model. The bias causes models inferred through PLM to appear closer to criticality than one would expect from the data. We introduce data-driven methods to correct this bias and explore their application in a small fMRI dataset. Our results indicate that additional care should be taken when attributing criticality to real-world datasets, as limited dataset sizes overstate the criticality of the inferred model. We also apply PLM to a large publicly available fMRI dataset from the human connectome project and find that the resting state network of the human brain corresponds to a near-critical paramagnetic state point. The inferred PLM model contains a highly structured coupling network with a heavy, power-law-like tail and we show that negative couplings play a vital role in mediating correlations within the network. We find that coupling networks are sparser than correlation networks, and suggest that inverse methods such as PLM should replace standard correlation-based approaches to network reconstruction in neuroimaging.



# Acknowledgements

I would like to thank Francesco Turci, Thomas Machon and C. Patrick Royall for their supervision and input into this research project. I would also like to thank the Bristol Centre for Functional Nanomaterials (BCFN) for their funding, and for the training opportunities I gained as part of this centre for doctoral training. Thank you most of all to my partner Florence, and to my friends and family, for supporting me through this challenging period. The PhD has been the most difficult experience of my life, and I could not have done it without them. I also want to thank all the friends I made along the way, especially Laurent Vaughan, Marcos Villeda Hernandez and Teodoro Garcia Milan, for the laughs and the burritos, and Yushi Yang, Fergus Moore, Levke Ortlieb, Katherine Skipper, Abraham Mauleon Amieva and Ioatzin Rios de Anda for making G39 such a wonderful place to work in. I especially need to thank Yushi Yang for introducing me to the physics of complex systems, and for showing me a path I could follow for my own research. Thank you to Joshua Robinson and Yushi Yang for creating and updating the LaTeX template I used to format this thesis. I also cannot overstate how important Ioatzin Rios de Anda was in my success, thank you for holding up the group when there was no one else to lean on. Thank you to Léo and Sarah for all the fun we've had together, you are an absolute joy to know! Flo, thanks for picking me up when everything was crumbling, and for providing the energy and light for our life when I couldn't. You truly do mean the world to me.



# Author's declaration

I declare that the work in this dissertation was carried out in accordance with the requirements of the University's Regulations and Code of Practice for Research Degree Programmes and that it has not been submitted for any other academic award. Except where indicated by specific reference in the text, the work is the candidate's own work. Work done in collaboration with, or with the assistance of, others, is indicated as such. Any views expressed in the dissertation are those of the author.

SIGNED: .....**MBK**.....      DATE: .....*20/03/23*.....

## Publications:

1. M. B. Kloucek et al., *Biases in Inverse Ising Estimates of Near-Critical Behaviour*, Jan. 2023, <http://arxiv.org/abs/2301.05556> - to appear in Physical Review E (PRE)



# Contents

<b>Abstract</b>	<b>iii</b>
<b>Acknowledgements</b>	<b>v</b>
<b>Author’s declaration</b>	<b>vii</b>
<b>List of Figures</b>	<b>xv</b>
<b>List of Tables</b>	<b>xvi</b>
<b>List of Abbreviations</b>	<b>xviii</b>
<b>List of Symbols</b>	<b>xix</b>
<b>1 Introduction</b>	<b>1</b>
1.1 A Physicist’s Introduction to the Brain . . . . .	2
1.1.1 Neurons are Binary Variables . . . . .	2
1.1.2 Functional Magnetic Resonance Imaging . . . . .	3
1.2 Objectives and Thesis Structure . . . . .	5
<b>2 Statistical Mechanics of Binary Variables</b>	<b>8</b>
2.1 A Brief Introduction to Statistical Mechanics . . . . .	8
2.2 Critical Points and Phase Transitions . . . . .	10
2.2.1 Order Parameters and Susceptibilities . . . . .	11
2.2.2 Finite-Size Effects . . . . .	12
2.2.3 Scale-Free Distributions . . . . .	12
2.2.4 Universality . . . . .	13
2.3 The Brain, Spin-Glasses and Complex Networks . . . . .	13
2.3.1 Complex Networks . . . . .	13
2.3.2 Complex Brain Networks . . . . .	15
2.3.3 Brains as Spin-Glasses . . . . .	16
2.4 The Sherrington-Kirkpatrick model . . . . .	16
2.4.1 The Spin-Glass Order Parameter . . . . .	17
2.4.2 Phase Diagram . . . . .	18
<b>3 Inverse Ising Inference</b>	<b>21</b>
3.1 Maximum Entropy Models . . . . .	21
3.1.1 Recovering the Canonical Ensemble . . . . .	23
3.1.2 Recovering the Grand Canonical Ensemble . . . . .	23



3.1.3	Recovering the Generalised Ising Model . . . . .	23
3.2	Inferring Ising Models . . . . .	24
3.2.1	Exact Likelihood Maximisation: Boltzmann Learning . . . . .	25
3.2.2	Statistical Physics Based Approximations . . . . .	26
3.2.3	Pseudo-likelihood Maximisation . . . . .	28
3.3	Network Reconstruction . . . . .	29
3.4	Maximum Entropy Models of Real Data . . . . .	32
3.4.1	Neural Activity (2 Spin States) . . . . .	32
3.4.2	Protein-Protein Interactions (21 Spin States) . . . . .	33
3.4.3	Flocks of Birds (Continuous Spin States) . . . . .	33
3.4.4	Financial Markets (2 Spin States) . . . . .	33
3.4.5	Summary . . . . .	33
<b>4</b>	<b>Criticality in Neuroscience</b>	<b>36</b>
4.1	Criticality as an Organising Principle . . . . .	36
4.1.1	Fine Tuning . . . . .	36
4.1.2	Dynamic Systems and Self-Organised Criticality . . . . .	37
4.1.3	Extended Critical Regimes . . . . .	38
4.2	Evidence of Criticality in Neuroscience . . . . .	39
4.2.1	Statistical Criticality of Maximum Entropy Models . . . . .	40
4.2.2	Dynamical Criticality and Neural Avalanches . . . . .	41
4.3	Maximum Entropy Models and Proximity to Criticality . . . . .	41
4.3.1	Inference Errors in Simulated Systems . . . . .	41
4.4	A Taste of Things to Come . . . . .	42
<b>5</b>	<b>Pseudo-likelihood Maximisation in the Sherrington-Kirkpatrick Model</b>	<b>45</b>
5.1	Problem Setup . . . . .	45
5.1.1	Model Generation . . . . .	46
5.1.2	Data Generation . . . . .	46
5.1.3	Relevant Observables . . . . .	47
5.2	PLM for Example State Points . . . . .	49
5.2.1	Single Realisation Investigation . . . . .	50
5.2.2	Error Distribution and Row Dependence . . . . .	53
5.3	Overview of Inference Errors in the SK Model . . . . .	55
5.3.1	Error Phase Diagram . . . . .	55
5.3.2	Error Dependence on Temperature . . . . .	57
5.4	PLM Performance with Limited Data . . . . .	58
5.4.1	Inferring Temperature from Finite Sized Datasets . . . . .	58
5.4.2	Inferring Susceptibility from Finite Sized Datasets . . . . .	60
5.4.3	Formalising Temperature Dependence on Sample Size . . . . .	61
5.5	Explaining the Sample Size Bias . . . . .	63
5.5.1	Bias of Maximum Likelihood Estimates . . . . .	63
5.5.2	Bias of Average Properties in PLM . . . . .	63
5.5.3	Small Sample Sizes and Logistic Regression . . . . .	67
5.6	Inference for Varied System Sizes . . . . .	67
5.6.1	Phase Diagrams . . . . .	68
5.6.2	Error across the P-SG transition . . . . .	70
5.6.3	Error Dependence on System Size . . . . .	72

5.6.4	Sub-Sampling to Find the Bias . . . . .	72
5.6.5	Bias Dependence on System Size . . . . .	75
5.6.6	Explaining the Error-Bias Discrepancy . . . . .	76
5.7	Conclusion . . . . .	77
<b>6</b>	<b>Correcting the Bias of Pseudo-Likelihood Estimates</b>	<b>80</b>
6.1	Motivation and Success Metrics . . . . .	80
6.2	Explicit Corrections . . . . .	81
6.2.1	Correction 1: Sub-Sampling Temperature Estimate . . . . .	81
6.2.2	Correction 2: Enforcing Self-Consistent Correlations . . . . .	82
6.2.3	Correction 3: Jackknife Resampling . . . . .	82
6.2.4	Comparing Sub-Sampling and Self-Consistency Corrections . . . . .	83
6.2.5	Relating the Self-Consistency Correction to the First-Order Bias . . . . .	85
6.3	Implicit Correction . . . . .	88
6.3.1	Firth’s Penalized Logistic Regression . . . . .	88
6.4	Comparing the Self-Consistency Correction to Firth’s Penalized Regression . . . . .	90
6.4.1	Implications for inference around criticality . . . . .	92
6.5	Conclusion . . . . .	92
<b>7</b>	<b>Mindfulness Meditation: a Small Sample Size Study</b>	<b>95</b>
7.1	The Mindfulness Meditation Dataset . . . . .	95
7.2	Small Sample Bias in Real Data . . . . .	96
7.2.1	Impact on State-Point Classification . . . . .	97
7.2.2	Impact on Claims of Criticality . . . . .	98
7.2.3	Ubiquitous Criticality is a Small Sample Size Effect . . . . .	98
7.3	Structure and Statistical Mechanics of the noMM Resting State Network . . . . .	99
7.3.1	Couplings and Correlations . . . . .	100
7.4	Conclusions . . . . .	104
<b>8</b>	<b>The Human Connectome Project: a Large Sample Size Study</b>	<b>106</b>
8.1	The HCP Resting-State Dataset . . . . .	106
8.2	The Resting-State Coupling Network . . . . .	108
8.2.1	Sub-Sampling Analysis . . . . .	108
8.2.2	Overview of the Inferred Connectivity . . . . .	109
8.2.3	Characterising the Inferred Coupling Distribution . . . . .	111
8.2.4	Characterising the Positive Tail . . . . .	112
	Method 1: Simple Binning and Fitting . . . . .	113
	Method 2: A Maximum Likelihood Method . . . . .	114
	Fitting Results Summary . . . . .	116
8.3	Criticality of the Inferred PLM Model . . . . .	117
8.3.1	Fictive Temperature Sweep . . . . .	117
8.4	Thresholding and Network Construction . . . . .	120
8.4.1	Statistics of Thresholded Coupling Networks . . . . .	121
	Assessing the two thresholding schemes in the limit $\delta \rightarrow 0$ . . . . .	122
	Assessing the two thresholding schemes when $\delta \neq 0$ . . . . .	123
8.4.2	Thresholds and Phase Transitions . . . . .	124
	Fictive Temperature Sweeps for Symmetric Thresholds . . . . .	124
	Fictive Temperature Sweeps for Positive Thresholds . . . . .	125

8.4.3	Thresholding Analysis Summary . . . . .	127
8.5	Conclusion . . . . .	128
<b>9</b>	<b>Conclusions</b>	<b>131</b>
	<b>Appendices</b>	<b>134</b>
<b>A</b>	<b>Determining the Linear Bias Regime</b>	<b>135</b>
A.1	Statistically Identifying a Constant Gradient . . . . .	136

# List of Figures

1.1	Activation profile of a neuron . . . . .	2
1.2	How to extract binary data from neuroimaging studies . . . . .	4
2.1	Parameter matrix of the 2D Ising model with periodic boundary conditions . . . . .	10
2.2	Finite size effects in the 2D Ising model . . . . .	12
2.3	Example network topologies . . . . .	14
2.4	Phase diagram of the zero-field Sherrington-Kirkpatrick model . . . . .	18
3.1	Comparison of inverse Ising inference methods in the SK model . . . . .	28
3.2	Comparison between correlations and interactions for a neuroimaging dataset. . . . .	31
4.1	Overview of the contact process . . . . .	37
4.2	Griffiths phases in complex networks and Ising models . . . . .	39
4.3	Examples of evidence for statistical criticality in neural ensembles . . . . .	40
5.1	Auto-correlation times for 4 example SK state points . . . . .	50
5.2	Paramagnetic example state point inference overview . . . . .	51
5.3	Paramagnetic spin-glass transition example state point inference overview . . . . .	51
5.4	Spin-glass example state point inference overview . . . . .	52
5.5	Ferromagnetic example state point inference overview . . . . .	52
5.6	Expanded Inference errors in the ferromagnetic phase . . . . .	53
5.7	Inference error probability distributions for the SK example state points . . . . .	54
5.8	Row error dependence . . . . .	55
5.9	Zero-field SK phase diagram for $N = 200$ . . . . .	56
5.10	Correlations and inference error in the SK model as functions of the temperature at fixed $\mu = 0.1$ for $N = 200$ . . . . .	57
5.11	Inferred temperature as a function of input temperature . . . . .	59
5.12	Impact of inference error on critical properties of the inferred models . . . . .	60
5.13	Saturation of the inferred temperature as a function of sample number . . . . .	61
5.14	Inference error and learning-difficulty parameter as functions of the temperature . . . . .	62
5.15	Inferred temperature dependence on sample number for strongly coupled state points . . . . .	62
5.16	Summary of biased estimates in pseudo-likelihood estimation . . . . .	64
5.17	SK phase diagrams for varying $N$ . . . . .	69
5.18	Auto-correlation time and error as functions of the temperature for varying system sizes . . . . .	70
5.19	Susceptibility and error as functions of the temperature for varying system sizes . . . . .	71
5.20	Location of susceptibility peak and error minimum as functions of system size . . . . .	72
5.21	Dependence of the minimum error in the SK phase diagram on the system size . . . . .	73

5.22	Inferred temperature as a function of sample size for the minimum error state points of the SK phase diagram . . . . .	74
5.23	Demonstration of sub-sampling method to find the bias strength . . . . .	74
5.24	Bias dependence on system size . . . . .	75
5.25	Linearity of the re-scaled error and bias . . . . .	77
6.1	Saturation of susceptibility with sample size . . . . .	82
6.2	Effect of the self consistency correction on the critical properties of the inferred model	83
6.3	Comparison of explicit bias corrections . . . . .	84
6.4	Comparing of jackknife and self-consistency corrections on the inferred coupling distribution from a small dataset . . . . .	86
6.5	Comparison of parameter shifts from jackknife and self-consistency corrections as sample size increases . . . . .	87
6.6	Deviation of jackknife shifts from the self-consistency correction analysed by input parameter magnitude . . . . .	88
6.7	Comparing Firth’s penalised logistic regression to PLM . . . . .	89
6.8	Comparing impact of Firth’s correction and the self-consistency correction on the inferred temperature and susceptibility estimates . . . . .	91
7.1	Auto-correlation of the mindfulness meditation (small sample size) dataset . . . . .	96
7.2	Coupling distributions for the mindfulness meditation dataset . . . . .	97
7.4	Sub-sampling analysis of the noMM condition in the mindfulness dataset as a function of inverse sample number . . . . .	97
7.3	Sub-sampling analysis of the noMM condition in the mindfulness dataset . . . . .	98
7.5	Fictive temperature sweep for mindfulness dataset . . . . .	99
7.6	Coupling matrix inferred from the noMM condition dataset . . . . .	100
7.7	Correlation matrix for the noMM condition dataset . . . . .	101
7.8	Trajectory visualisations for different state-points . . . . .	101
7.9	Empirically measured covariance vs simulated covariances at different fictive temperatures for the noMM PLM model . . . . .	102
7.10	Adjacency matrices produced by thresholding covariance matrices at different fictive temperatures . . . . .	103
8.1	The multi-modal parcellation of the human cerebral cortex . . . . .	107
8.2	Auto-correlation showing the dynamics of the HCP dataset . . . . .	108
8.3	Sub-sampling analysis of the HCP dataset . . . . .	109
8.5	Hemisphere parameter correlations . . . . .	109
8.4	Coupling matrix inferred from the HCP dataset . . . . .	110
8.6	Distribution of HCP couplings for the intra- and inter-hemisphere networks . . . . .	111
8.7	Coupling distributions of the HCP dataset and the no-mindfulness dataset . . . . .	112
8.8	Analysis of the positive tail of the inferred couplings . . . . .	113
8.9	Summary of the analysis of the heavy tail via direct binning and linear fitting . . . . .	114
8.10	Summary of the analysis of the heavy tail via the maximum likelihood method . . . . .	115
8.11	Fictive temperature sweep for the PLM model of the HCP dataset . . . . .	118
8.12	Example trajectories of the HCP PLM model at different fictive temperatures . . . . .	119
8.13	Matrix of the Pearson’s correlation coefficients for the HCP aggregate dataset . . . . .	120
8.14	Statistical physics observables for thresholded inferred models . . . . .	122
8.15	Coupling distribution of symmetrically thresholded models . . . . .	124

8.16	Order parameter dependence on the fictive temperature for a range of symmetrically thresholded models . . . . .	125
8.17	Susceptibility dependence on the fictive temperature for a range of symmetrically thresholded models . . . . .	126
8.18	Coupling distribution of positively thresholded models . . . . .	126
8.19	Order parameter dependence on the fictive temperature for a range of positively thresholded models . . . . .	127
8.20	Susceptibility dependence on the fictive temperature for a range of positively thresholded models . . . . .	128
A.1	Inverse temperature vs inverse sample size for varying $N$ . . . . .	135
A.2	Statistical method to determine whether observables depend linearly on the inverse sample number . . . . .	136
A.3	Constant bias gradients identified using the statistical method . . . . .	137

# List of Tables

5.1	Summary statistics for 4 example SK state points . . . . .	49
6.1	Comparison of errors in temperature and susceptibility estimates . . . . .	92
8.1	Fitted power law parameters of the positive tail of the HCP couplings . . . . .	113

# List of Algorithms

1	Boltzmann Learning . . . . .	26
---	------------------------------	----



# List of Abbreviations

**fMRI** Functional magnetic resonance imaging 3

**BOLD** Blood oxygen level dependent 5

**ROI** Region of interest 5

**SK** Sherrington-Kirkpatrick 6, 8, 16

**PLM** Pseudo-likelihood maximisation 6, 28

**HCP** Human Connectome Project 6

**P** Paramagnetic 11

**F** Ferromagnetic 11

**SG** Spin-glass 16

**i.i.d** independent and identically distributed 17

**MEM** Maximum entropy model 24

**MC** Monte Carlo 25, 46

**FC** Functional connectivity 29

**SOC** Self-organised criticality 37

**GSI** generate-simulate-inference 48

**MLE** Maximum likelihood estimate 63

**SS** Sub-sampling 81

**SC** Self-consistency 82, 97

**MM** Mindfulness meditation 95

**noMM** No mindfulness meditation 95

**L** Left 107

**R** Right 107

# List of Symbols

<b>Notation</b>	<b>Description</b>	<b>Page List</b>
$B$	the number of samples in each dataset	5, 8
$N$	the number of spins / regions of interest	5, 8
$s_i$	the state of the $i^{th}$ spin in the system	8
$h_i$	the external field acting on the $i^{th}$ spin in the system	9
$J_{ij}$	the pair-wise couplings between spins $i$ and $j$	9
$\mathbf{h}$	the vector of external fields	9
$\mathbf{J}$	the matrix of couplings	9
$\boldsymbol{\theta}$	the parameter matrix containing all fields and couplings of the Ising Hamiltonian	9
$\beta$	the inverse temperature	9
$Z$	the partition function	9
$k_B$	the Boltzmann constant	9
$T$	the temperature	9
$F$	the free energy	9
$\langle A \rangle$	the thermodynamic average of an observable $A$	9
$J$	the constant couplings strength of the Ising model	10
$T_c$	the critical temperature	11
$m$	the ferromagnetic order parameter / the magnetisation	11
$\chi$	the global magnetic susceptibility	11
$T_R$	the reduced temperature	12
$\mathbf{A}$	the adjacency matrix	15
$k_i$	the degree of the $i^{th}$ node in the network	15
$\mu_J$	the extensive mean of the coupling distribution	17
$\sigma_J$	the extensive standard deviation of the coupling distribution	17
$\mu$	the intensive mean of the coupling distribution	17
$\sigma$	the intensive standard deviation of the coupling distribution	17
$m_i$	the local magnetisation of the $i^{th}$ spin	17
$q$	the spin-glass order parameter / the overlap	17

<b>Notation</b>	<b>Description</b>	<b>Page List</b>
$\chi_{SG}$	the spin-glass susceptibility	18
$C_{ij}$	the covariance of spins $i$ and $j$	18
$t$	time	24
$\{\mathbf{s}\}^B$	the full dataset consisting of $B$ observations from $N$ spins	24
$\langle A \rangle^e$	the empirical mean of an observable $A$ in the dataset	24
$\mathbf{m}^e$	the vector of empirical local magnetisations	24
$\mathbf{C}^e$	the empirical covariance matrix	24
$\mathcal{L}$	the log-likelihood function	25
$\mathbf{h}^*$	the inferred field vector of the pairwise maximum entropy model	25, 45
$\mathbf{J}^*$	the inferred coupling matrix of the pairwise maximum entropy model	25, 45
$\mathbf{J}^{\text{nMF}}$	the naive mean-field estimate of the coupling matrix	27
$\mathcal{L}_r$	the row dependent log-likelihood of pseudo-likelihood maximisation (PLM)	28
$h_r$	the field in the $r^{\text{th}}$ row of the PLM parameter matrix	28
$\mathbf{J}_r$	the couplings in the $r^{\text{th}}$ row of the PLM parameter matrix	28
$R_{ij}$	the Pearson correlation coefficient of spins $i$ and $j$	29
$\boldsymbol{\theta}^0$	the input parameter matrix	45
$\mathbf{h}^0$	the input field vector	45
$\mathbf{J}^0$	the input coupling matrix	45
$\boldsymbol{\theta}^*$	the inferred parameter matrix of the pairwise maximum entropy model	45
$T_f$	the fictive temperature	46
$t_{eq}$	the equilibration time of a Monte Carlo simulation	47
$t_{pr}$	the production time of a Monte Carlo simulation	47
$t_s$	the sampling interval of a Monte Carlo simulation	47
$p_{mc}$	the Monte Carlo sampling meta-parameters: $\{t_{eq}, t_{pr}, t_s\}$	47
$C^2$	the temperature independent spin-glass susceptibility measure	48
$C_t$	the auto-correlation function	48
$\Delta t$	the time delay	48
$\tau$	the auto-correlation time	48
$\varepsilon$	the inference error	48
$T^0$	the measured input temperature	48
$\sigma_J^0$	the measured standard deviation of the input couplings	48
$T^*$	the measured inferred temperature	48
$\sigma_J^*$	the measured standard deviation of the inferred couplings	48
$T^{B \rightarrow \infty}$	the infinite sample size temperature prediction	61

<b>Notation</b>	<b>Description</b>	<b>Page List</b>
$\tilde{B}$	the bias strength	61
$\gamma$	the exponent of the power law	112
$x_{\min}$	the minimum value denoting the start of heavy-tailed behaviour	113
$R^2$	the coefficient of determination	114
$D$	the Kolmogorov-Smirnov distance metric	115
$\delta$	the threshold	121



# Chapter 1

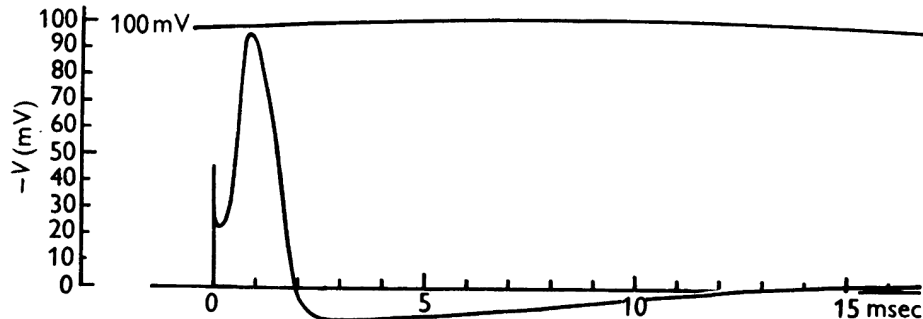
## Introduction

As a result of the increasing availability of high-quality biological data, a growing body of evidence suggests that some of the most interesting properties of living systems - such as memory or the ability to solve problems - are *emergent* properties which arise out of the collective dynamics of relatively simple individual constituents [2, 3]. This sentiment of “more is different” [4] is perhaps most apparent in the context of ensembles of neurons. The dynamical activation profile of single neurons, and the non-linear interaction by which they stimulate other neurons, have been known to some extent since the 1950s [5]. Yet, it appears entirely hopeless to try and explain the diverse goal-oriented macroscopic behaviours (cognitive, emotional or otherwise) exhibited by living beings in terms of the simple dynamical rules governing individual neurons. Ultimately, these behaviours must result from emergent phenomena caused by the cooperative behaviour of the underlying neuronal collection.

Not only does the experimental evidence point to emergent behaviour, but it suggests that many biological systems exist near a *critical point*, that is, an order-disorder phase transition [6, 7]. Criticality, the term used to describe the phenomenology near such transitions, is believed to provide biological systems with a range of advantageous properties, from maximising their sensitivity to inputs [8] to enabling them to exhibit an extensive range of dynamic responses [9, 10]. The apparent ubiquity of close-to-critical behaviour in biology has led some to conjecture that evolutionary pressures may be responsible for fine-tuning biological systems towards a critical point, and that criticality serves as a template from which complex behaviours arise across a range of length scales. While the idea of criticality being a generic feature of biological systems is still in its infancy, a growing pool of empirical results suggests it as a robust hypothesis specifically in neuronal systems [7]. Indeed, the “distance” between the state of the system and its supposed critical point is becoming an increasingly relevant biological variable [11] and, in neuroscience for instance, is being considered as a guiding route for clinical work [12].

This brings us to the work we report in this thesis. Inspired by recent findings [13], we sought to investigate the criticality of coarse-grained brain imaging data using methods from statistical physics, as statistical physics is the branch of physics that deals with the study of emergence in complex ensembles. Specifically, we construct equilibrium statistical physics models

**Figure 1.1:** The dynamic activation profile of a neuron, where  $V$  denotes changes in the membrane potential of the cell. Upon receiving a stimulus at time  $t = 0$ , the neuron rapidly depolarises before returning to its resting potential. The state of the neuron is commonly described as binary (i.e. on or off) as this “spike” in activity occurs in a short time window ( $\sim 1$  millisecond). The plotted activation profile is from experiments on the squid giant axon, reproduced from [5].



that are parameterised to represent real data via a machine learning method named inverse Ising inference [14–17]. Although our primary goal was to analyse real data, we identified that sufficient methodological gaps existed in the current literature and that the performance of inverse Ising inference was not well characterised on datasets of limited size. Understanding the bias of the inference when sample sizes are small is vital to correctly interpret fitting results from real datasets. Our work is therefore composed of an in-depth analysis of the predominant technique used to perform inverse Ising inference (so-called pseudo-likelihood maximisation), as well as results gathered by applying this technique to both small and large brain imaging datasets.

## 1.1 A Physicist’s Introduction to the Brain

The aim of this interdisciplinary research project was to build physics-based models that represent experimental data from neuroscience. As we chose to specifically focus on data collected from neuroscience (from functional magnetic resonance imaging to be exact), we will begin by providing “a physicist’s introduction to the brain”. While leaving out an egregious amount of detail, this will hopefully allow us to establish the surface-level knowledge required to interpret the rest of our work.

### 1.1.1 Neurons are Binary Variables

The elementary constituents of the nervous system (which for our purposes is synonymous with the brain) are *neurons*. Neurons (like all cells) have an electrical potential difference across the cell membrane which separates the interior of the cell from the extracellular fluid. Unlike most cells, however, neurons are also excitable, so external inputs can cause the membrane potential of the neuron to rapidly deviate from its baseline. This allows neurons to propagate electric signals. In Nobel Prize-winning work, Hodgkin and Huxley [5] introduced a quantitative model that described the initiation and propagation of these potential changes in terms of ion currents across the cell membrane. We show one such example potential for a single neuron in Fig. 1.1. Upon receiving an input stimulus at time  $t = 0$ , the neuron rapidly depolarizes, before returning to its original baseline voltage. The initial period ( $t = 0$ ms to  $t = 2$ ms) of this profile is commonly referred to

as the action potential, as a neuronal spike, or as the neuron firing. The long-time behaviour ( $t > 2\text{ms}$ ) where the membrane potential is below the baseline is referred to as the refractory period, during which the neuron cannot be re-excited. The entire spiking process occurs very rapidly (on a millisecond time scale), and so it has long been standard to model the activity of each neuron as an “all-or-none” process [18], i.e. as a binary variable which can either be on or off. Indeed we see this reflected in the terminology (i.e. spiking, firing, etc) commonly used to describe the state of each neuron.

Ensembles of binary variables have been studied extensively in statistical physics. Historically, this research aimed to understand the collective properties of atomic magnetic moments (i.e. *spins*) in metals<sup>1</sup>. Binary representations of these spins (i.e. where each spin can only be oriented in one of two directions, usually referred to as up and down) have been shown to provide great pedagogical descriptions of the phenomenology of magnetism [19], with the famous 2-dimensional Ising spin model being one of the archetypal models with which statistical physics and phase transitions are understood [20]. Although developed in the context of magnetism, these spin models are sufficiently general to describe the interactions of many systems. As such, results and techniques developed in the analysis of magnetism can naturally be extended to the analysis of ensembles of neurons in the brain. Indeed this connection has been recognised since the 1980s [21, 22], where computational features such as a content addressable memory were shown to emerge in simple spin-based statistical physics models. We will follow this line of thinking and model the brain as an ensemble of binary variables.

### 1.1.2 Functional Magnetic Resonance Imaging

The brain of a healthy human adult consists of approximately 100 billion neurons, each of which can be connected to around 1000 other neurons<sup>2</sup> [24], forming a complex network [25]. One of the key experimental challenges of neuroscience is to establish this connectivity, and more so, how the structure of the neural network is related to its function [26]. As one might expect, experimentally characterising the structural connectivity of the brain with cellular resolution is a gargantuan task. To date, the largest living organism for which we have measured a complete neuron-resolved structural connectivity network is the nematode *Caenorhabditis elegans* (a roundworm), which only has 302 neurons [27]. Moreover, we must also understand neural dynamics (i.e. neural correlations in time) to begin to understand the function of each neuron or grouping of neurons [2]. Therefore, even if we could characterise all synaptic connections within the brain, we may never understand the function of each component without time-resolved data on how signals propagate through the network [28].

One approach to tackling these issues is to employ so-called neuroimaging methods. In these, neural activity is probed non-invasively by measuring proxy quantities that can be related back to the activation of neurons. Many neuroimaging modalities exist, but in this work, we focus exclusively on data collected via functional magnetic resonance imaging (fMRI)<sup>3</sup>. As the name suggests, fMRI works by measuring the *magnetic resonances* of biological tissues, as hydrogen nuclei in different tissues have different inher-

<sup>1</sup> The simplest example of such a phenomenon is ferromagnetism, where all magnetic spins align in the same direction to generate a macroscopic magnetic field.

<sup>2</sup> The points at which neurons connect to one another are called synapses. There are therefore approximately 100,000 billion synaptic connections within our brains.

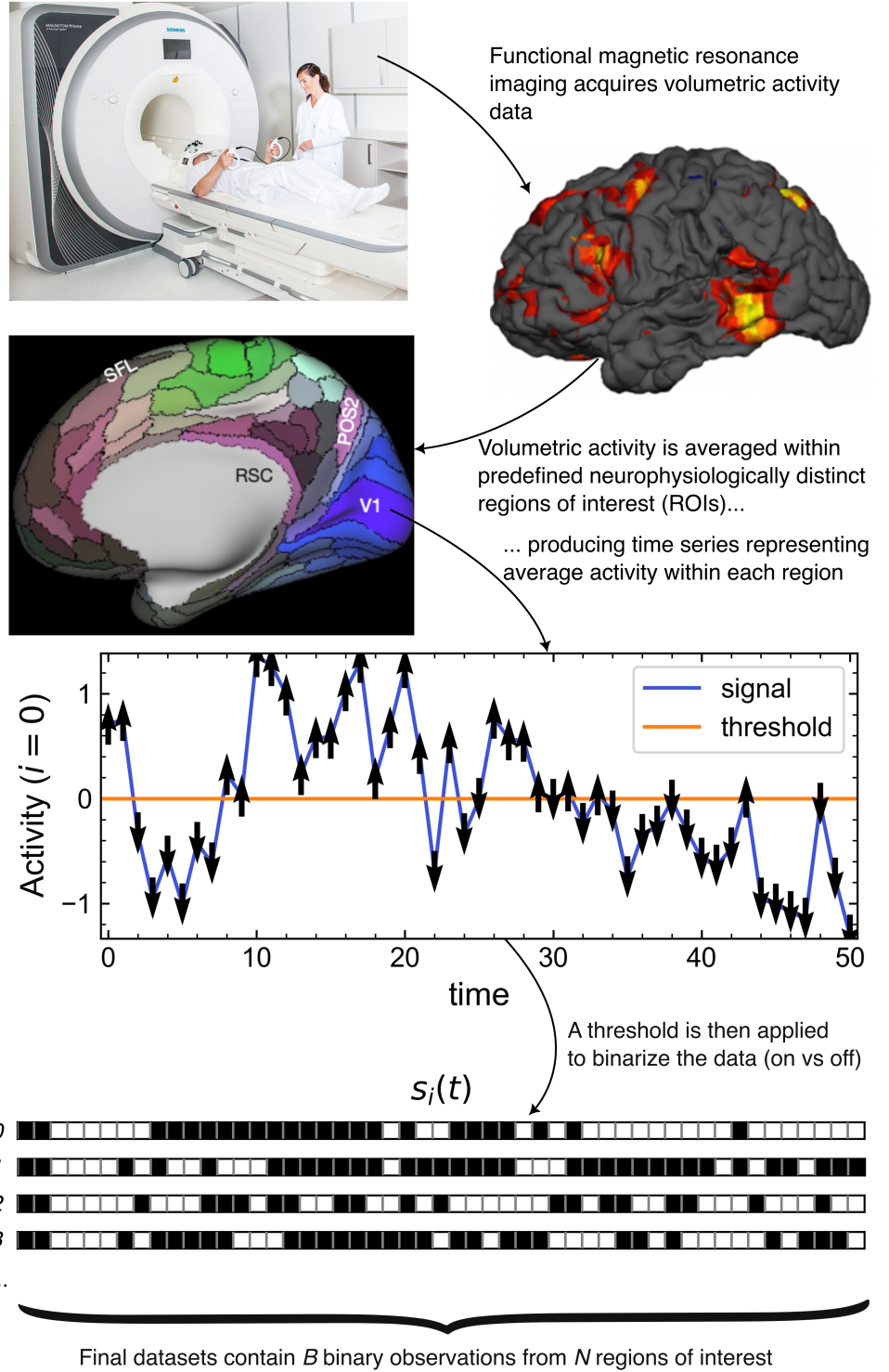
<sup>3</sup> Other notable examples include electroencephalography (EEG) and magnetoencephalography (MEG) [29].



**Figure 1.2:** Schematic representation of how data collected from fMRI studies is transformed into the binary datasets we work with. fMRI is a non-invasive neuroimaging technique, which collects volumetric time series data. Each voxel typically represents the activity of hundreds of thousands of neurons. The volumetric data is then segmented into different neurologically relevant regions of interest (ROIs), e.g. V1 labels the first visual region of the brain. This process is termed parcellation, and the parcellation atlas shown here is adapted from Ref. [23]. All ROIs are spatially continuous but represent non-uniform volumes. This produces  $N$  continuous activity time series of length  $B$ , one for each ROI, with ROIs labelled by  $i$  and time by  $t$ . Each series is then thresholded to produce the binary time series which we ultimately analyse. We threshold the activity about 0, so that the binary ROI states are “activated” ( $s_i = +1$ , shown in black) if the activity within the ROI is above average, and deactivated ( $s_i = -1$ , shown in white) if it is below average. The goal of our work is to infer the network of interactions between ROIs from these binary activation time series. Images of the fMRI machine<sup>a</sup>, and volumetric activity<sup>b</sup> were sourced from the internet.

<sup>a</sup>fMRI scanning image, accessed (27/02/2023)

<sup>b</sup>volumetric activity, accessed (27/02/23)



ent magnetic resonances. fMRI specifically uses the fact that deoxygenated haemoglobin is paramagnetic, while oxygenated haemoglobin is diamagnetic. The contrast between these two magnetic states then allows the flow of oxygenated blood through the brain to be tracked [30]. The brain does not store energy locally, and so the so-called blood oxygen level-dependent (**BOLD**) signal indirectly represents neural activity by assuming that in-use (i.e. active) areas of the brain require additional nutrients (in the form of glucose, oxygen, etc) to operate [31], leading to changes in blood flow to those areas. The BOLD signal is a *proxy* measure for neural activity as blood oxygenation levels are also impacted by other physiological processes, such as respiration and cardiac oscillations [32]. We will disregard these details, as it is generally accepted that fluctuations in the BOLD signal do indeed correlate with neural activity [33, 34], and assume that the data we work with has either been pre-processed to account for these effects or that they do not play a significant role (either way we leave this problem to the neuroscientists).

fMRI acquires 3-dimensional volumetric data, with the sides of each imaging voxel usually being around 5mm in length. The fMRI signal, therefore, does not represent the activity of individual neurons, but rather the average activity of  $\mathcal{O}(1,000,000)$  neurons within the imaging voxel. Samples are collected once every 1 to 3 seconds (i.e. with a sampling frequency of around 0.5Hz). It is common practice to further reduce the complexity of the data by segmenting the volumetric scan into pre-defined spatially continuous regions of interest (**ROIs**) [23] and tracking the average signal within each of these. We then threshold the activity time series from each ROI to produce binary representations of the state of each region. These binary fMRI signals now represent whether or not an entire *region* of the brain is active or inactive. These steps are illustrated in Fig. 1.2. The final datasets we analyse contain  $B$  binary observations (off or on) from  $N$  regions of interest.

## 1.2 Objectives and Thesis Structure

The overarching aim of this thesis is to investigate criticality in human resting-state fMRI data in the context of statistical physics. We do this by using an inference (machine learning) method named inverse Ising inference, which allows us to reconstruct pairwise interactions from binary datasets. Models inferred via inverse Ising inference correspond to equilibrium statistical physics models which match the pairwise correlations of the data. We will establish if these models are close to a critical point, and how systematic errors within the inference technique might bias our conclusions.

There are three strands of literature underpinning our research, which we detail in Chapters 2 to 4. We introduce the foundations of equilibrium statistical physics as this is the framework we use to assess if our inferred models are critical in Chapter 2. We know that the brain contains both positive (excitatory) and negative (inhibitory) connections. This competition (termed *quenched disorder* in the physics literature) leads to the development of a *spin-glass* phase at low temperatures. Recent [13] and historic [22] work both connect brain data to this phase and so we also introduce the

phenomenology of spin-glasses in the archetypal Sherrington-Kirkpatrick (SK) [35] model in this chapter. We later use the SK model to characterise the inference and to establish the types of errors we might encounter. In chapter 3, we detail the inference method we use. We show that equilibrium models provide the least biased descriptions of the data [36], and that making different observations from the data corresponds to selecting different model classes. We will measure up-to-pairwise correlations, which corresponds to constraining our models to pairwise interactions. We compare different methods with which inverse Ising inference can be performed, and highlight that so-called pseudo-likelihood maximisation (PLM) [15] performs best overall. We thus use PLM exclusively throughout this work. The pairwise couplings inferred from PLM also allow us to gain insight into the network topology of the fMRI data. In chapter 4, we discuss evidence supporting the “critical brain hypothesis”, and provide an overview of how biological systems might organise towards this special point.

In the remaining chapters (5 - 8), we present the results of our investigation. We begin by performing pseudo-likelihood maximisation on simulated data from across the SK phase diagram in chapter 5. This was motivated by previous observations that PLM inference errors depend on how close-to-critical a given dataset is [16]. We validate these results for a range of system sizes and find that errors are minimal near, but offset from, the phase transitions of the SK model. Moreover, we find that PLM models inferred from fluctuating data (i.e. data with similar dynamics to the fMRI signal) are biased towards the critical point and that the severity of this effect is state-point and dataset size dependent. We show that average quantities of the model, such as the temperature, depend linearly on the inverse sample number ( $1/B$ ). We link this to standard results for the bias of maximum likelihood parameter estimates [17, 37]. We conclude by cautioning against claims of criticality when using PLM to analyse datasets with small sample sizes, as the apparent criticality of these could be the result of the bias entirely. Due to these results, we introduce methods to correct the bias in chapter 6. Our proposed corrective procedures allow us to establish a lower-bound estimate for the temperature of the data, and significantly reduce the inference error when the sample number is small. We conclude chapter 6 by recommending that an analysis of the bias should always be performed (e.g. by sub-sampling the data) if one wishes confidently assert that their inferred model is critical. Otherwise small sample size effects cannot be discounted as the source of the apparent criticality.

In chapter 7, we perform a case study of a small fMRI dataset to demonstrate how small sample size effects might impact real-world analyses. Two biological conditions (mindfulness meditation vs no mindfulness meditation) are compared. We investigated whether or not a) the control (no mindfulness condition) corresponds to a critical point, and b) whether practising mindfulness meditation shifts the state of the brain towards criticality. We show that applying PLM without consideration of the bias leads to incorrect conclusions on both fronts and that neither condition corresponds to a critical state point. In chapter 8, we apply PLM to one of the largest openly available fMRI datasets, the young adult study from the human connectome project (HCP). We sub-sample the HCP dataset and establish that this dataset is sufficiently large to disregard the effects of the bias. We find

that the majority of connections in the brain are negative (inhibitory), but that the overall distribution of couplings is skewed, with the positive tail of couplings following a power-law distribution. We find that these negative couplings are essential in mediating correlations within the network. We investigate the criticality of the inferred model and find that the HCP dataset corresponds to a super-critical state point close to a statistical critical point. We further show that inverse methods such as PLM provide insights into the functional connectivity of the brain beyond those of traditional correlation-based methods employed within neuroscience. We summarise our results and provide an overall conclusion in chapter 9.

## Chapter 2

# Statistical Mechanics of Binary Variables

We will be modelling the brain as a collection of binary variables. Specifically, we will infer equilibrium statistical physics models that best represent the brain data. The aim of this chapter is to establish the foundations of this theory. A number of excellent textbooks exist which provide detailed overviews of statistical mechanics [38–40]. We focus on presenting the key concepts required to interpret our work. We introduce the notion of criticality and phase transitions by making reference to the well-studied two-dimensional (2D) Ising model on a regular lattice. We know neural ensembles contain couplings which are both inhibitory (positive) and excitatory (negative). Competition between positive and negative couplings leads to the formation of so-called spin-glasses, the phenomenology of which we introduce through the archetypal Sherrington-Kirkpatrick (SK) model. We expect neural couplings to form complex networks and devote time to introducing these.

### 2.1 A Brief Introduction to Statistical Mechanics

The overarching aim of statistical mechanics is to explain how interactions between microscopic constituents (e.g. spins) lead to macroscopic properties (e.g. spontaneous magnetisation). Here we will only describe *equilibrium* statistical mechanics, that is, we will assume that our systems are in thermal equilibrium; they are at a constant temperature without external perturbations. We will consider a system of  $N$  interacting spins  $s_i \in \pm 1$ , indexed by  $i = 1, \dots, N$ . These spins can describe any arbitrary binary process, e.g. flipping between up and down states for real electronic spins or switching from on to off in the neuronal ensembles. We associate the microscopic state  $s_i = +1$  with up or on, while the state  $s_i = -1$  describes a spin pointing down or being off. The exact theory of statistical mechanics exists only in the *thermodynamic limit*, when  $N \rightarrow \infty$ , and for now we assume that this is satisfied. When  $N$  is small so-called finite-size effects [41–43] are introduced. The spins fluctuate in time, and for each of the  $N$  labelled regions, we have time series of length  $B$ . The state of the entire spin vector at a time  $t'$ ,  $\mathbf{s}(t = t')$ , is called a *configuration*, and the full dataset of  $B \times N$  observations will either be referred to as a *trajectory* or as the *dataset*. These spins

interact with each other, and (considering up-to-pairwise interactions) each configuration is associated with an energy or Hamiltonian,

$$\mathcal{H}(\mathbf{s}|\mathbf{h}, \mathbf{J}) = - \sum_i h_i s_i - \frac{1}{2} \sum_{i \neq j} J_{ij} s_i s_j. \quad (2.1)$$

Here the parameters  $h_i$  are local external fields acting on each spin  $s_i$ , which bias the spin towards pointing up or down. Each  $J_{ij}$  represents a pair-wise coupling between the spins  $i$  and  $j$  and describes how much influence two spins have on each other. The summation  $i \neq j$  runs over all non-matching pairs of  $i$  and  $j$ . The total energy of (2.1) is reduced when  $J_{ij} > 0$  and the spins  $i$  and  $j$  are aligned. The length- $N$  vector  $\mathbf{h}$  contains each of the external fields, while  $\mathbf{J}$  is a symmetric  $N \times N$  matrix of couplings with a 0-diagonal (as there are no self-interactions). The energy of the system is therefore a combination of two sets of quantities: the spin configuration  $\mathbf{s}$  at time  $t'$  and the constant parameters  $\mathbf{h}$  and  $\mathbf{J}$  which describe a given model. To simplify our notation further we now introduce the parameter matrix  $\boldsymbol{\theta}$ , which is a symmetric matrix containing all the model parameters, with  $\theta_{ii} = h_i$  and  $\theta_{ij} = J_{ij}$  as all  $J_{ii} = 0$ . When in equilibrium, the spins will fluctuate and explore different configurations with probabilities given by the Boltzmann distribution

$$P(\mathbf{s}|\boldsymbol{\theta}) = \frac{1}{Z} \exp\{-\beta \mathcal{H}(\mathbf{s}|\boldsymbol{\theta})\}, \quad (2.2)$$

where  $\beta = 1/k_B T$  is the inverse temperature and  $Z$  the partition function. Note that we set the Boltzmann constant  $k_B = 1$  throughout (or equivalently report all temperatures  $T$  in units of  $k_b$ ). The temperature introduces disorder (entropy) into the system, causing more fluctuations. The partition function, a type of normalisation constant, depends on all micro-states of the system and is given by

$$Z = \sum_u \exp\{-\beta \mathcal{H}(\mathbf{s}_u|\boldsymbol{\theta})\}, \quad (2.3)$$

where  $u$  indexes one of the  $2^N$  possible configurations of the system. The partition function is related to the (Helmholtz) *free energy*  $F$  of the system via:

$$F = -T \ln Z. \quad (2.4)$$

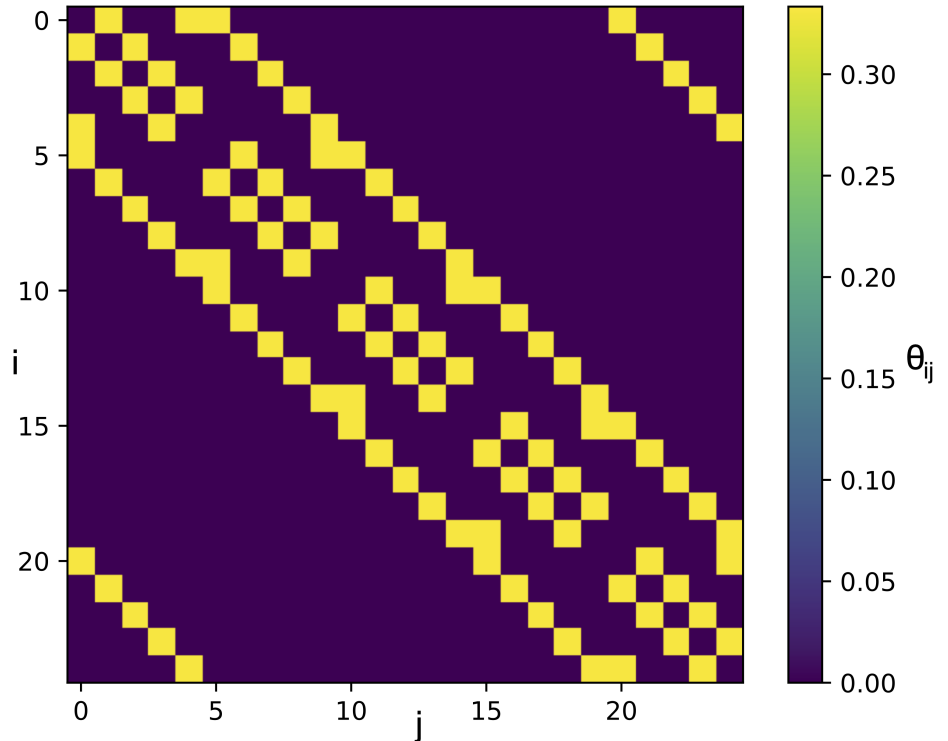
As previously noted, the spin configuration itself fluctuates in time, and macroscopic properties of the equilibrium system correspond to averages with respect to (2.2). We denote this so-called *thermodynamic average* with triangular brackets so that some measured macroscopic quantity  $A$  is given by

$$\langle A \rangle = \sum_u P(\mathbf{s}_u) A(\mathbf{s}_u). \quad (2.5)$$

If we introduce some field  $a$  that can be used to observe  $A$ , we can additionally write that

$$\langle A \rangle = \frac{1}{\beta} \frac{\partial \ln Z}{\partial a}, \quad (2.6)$$

**Figure 2.1:** Visualisation of the parameter matrix  $\theta$  of the 2D Ising model with periodic boundary conditions in zero external fields. The diagonal elements of  $\theta$  contain the fields, i.e.  $\theta_{ii} = h_i$ . The off-diagonal elements contain the couplings, so that if  $i \neq j$ ,  $\theta_{ij} = J_{ij}$ . In this case, the nearest-neighbour lattice structure of the 2D Ising model produces a regular parameter matrix where each spin is connected to 4 other spins with constant coupling strength. The external field is zero,  $\mathbf{h} = 0$ , and the constant coupling strength is  $J = 1/T = 1/3$ . This is for a system of size  $N = 25$ .



that is to say, observables such as  $A$  are first-order derivatives of the free energy. Note that calculating any quantity with (2.5) implicitly requires the partition function over all possible micro-states to be computed. This is often not possible and it is common practice in experiments and simulations to approximate the thermodynamic average with the sample average as:

$$\langle A \rangle \approx \frac{1}{B} \sum_{t=1}^B A(t). \quad (2.7)$$

This relation assumes that the system is *ergodic*, meaning that the system would explore all possible accessible micro-states (configurations) given infinite time. Low-temperature phases, and spin-glasses in particular, often break this ergodicity [44], making it difficult to sample them. The general aim of statistical mechanics is to calculate macroscopic expectations using (2.5) when theoretical calculation is possible or (2.7) when some representative sample of observations of the system is available.

## 2.2 Critical Points and Phase Transitions

Depending on the temperature  $T$  and specifics of  $\theta$ , the spin system may explore a range of macroscopic *phases*<sup>4</sup>. We quantify each phase by defining relevant *order parameters* for the system, which take unique values in each phase. To further explore these concepts, let us introduce one of the most well-studied models in physics: the Ising model [20]. The Ising model describes an ensemble of interacting spins placed on a lattice (of some dimension) which interact only with their nearest neighbours through a constant interaction of strength  $J$ . This system is described by the generalised

<sup>4</sup> Similarly to how water might exist as a solid, liquid or a gas (along with some less well known other phases [45]).



Hamiltonian of (2.1), where we now encode the topology of the problem within  $\mathbf{J}$  by setting  $J_{ij} = 0$  when  $i$  and  $j$  are not neighbours,  $J_{ij} = J$  when  $i$  and  $j$  are neighbours. An example of a parameter matrix for such a 2D lattice is shown in Fig. 2.1. Alternatively, we may simplify the Hamiltonian to

$$\mathcal{H}(\mathbf{s}|h, J) = -h \sum_i s_i - \frac{J}{2} \sum_{(i,j)} s_i s_j, \quad (2.8)$$

now only summing over nearest-neighbour pairs  $(i, j)$ . In the following discussion, we will specifically focus on the 2-dimensional Ising model, as this model has an analytical solution [46] and undergoes a phase transition at a finite, non-zero temperature. For now let us also assume that  $J > 0$ , which as noted previously, means it is energetically favourable for spins to align with each-other<sup>5</sup>.

<sup>5</sup> If we took  $J < 0$  the ground state would be an anti-aligned anti-ferromagnet.

### 2.2.1 Order Parameters and Susceptibilities

The 2D Ising model undergoes a second-order phase transition at a critical temperature  $T_c \approx 2.269J$  [38, 46]. Above this temperature, the system exists in a disordered paramagnetic (P) phase, dominated by random thermal fluctuations, and there is no preferred direction along which the spins point. Yet as the system is cooled below  $T_c$ , the orientational symmetry of the spins is broken, leading to the formation of an ordered aligned ferromagnetic (F) phase. In the case of the Ising model, a suitable order parameter describing this transition is the magnetisation  $m$ ,

$$\langle m \rangle = \frac{1}{N} \sum_i \langle s_i \rangle = \frac{1}{N\beta} \frac{\partial \ln Z}{\partial h}, \quad (2.9)$$

which describes the global alignment of the spins.  $h$  is a constant external field acting on all spins. The order parameter  $m$  varies smoothly i.e. *continuously* as one crosses the critical temperature; from  $m = 0$  in the high-temperature disordered phase, to  $m = 1$  in the low-temperature ordered phase. This transition from disorder to order is a collective process; all spins must act cooperatively to produce the overall global phase change that is observed. This implies that at the critical point, system-wide correlations must emerge, or in other words, the correlation length between spins must diverge. These long-range correlations are a core feature characterising the critical state, and are intimately related to the main phenomenology by which critical points are characterised: the divergence of *second-order* derivatives of the free energy, such as the magnetic susceptibility  $\chi$ :

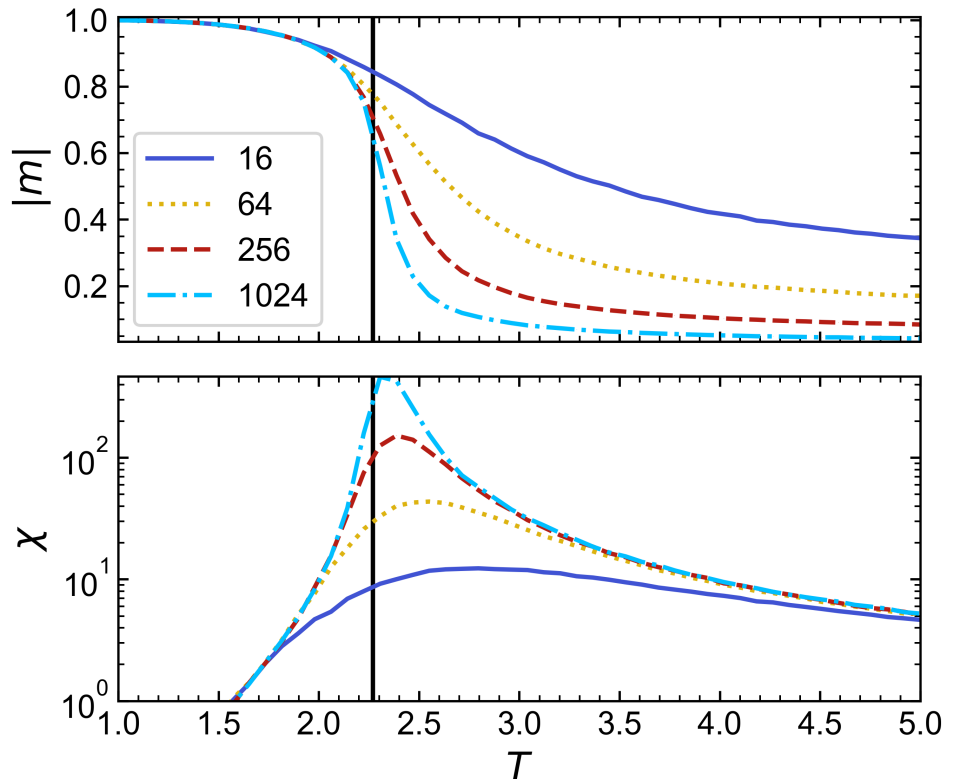
$$\chi = \frac{\partial \langle m \rangle}{\partial h} = \frac{1}{N\beta} \frac{\partial^2 \ln Z}{\partial h^2} = N\beta[\langle m^2 \rangle - \langle m \rangle^2]. \quad (2.10)$$

The point at which  $\chi$  diverges is termed the *critical point*. We show the dependence of  $m$  and  $\chi$  on  $T$  for finite  $N$  simulations of the simple 2D Ising model in Fig. 2.2. At the critical point, quantities like  $\chi$  diverge as *power-laws* in the reduced temperature:

$$\chi(T_R) \sim T_R^{-\gamma}, \quad (2.11)$$



**Figure 2.2:** Finite size effects in the 2D Ising model. The top panel shows the ferromagnetic order parameter  $|m|$  as a function of the temperature  $T$ , bottom panel the susceptibility  $\chi$ . Each line corresponds to a different system size, with the number of spins  $N$  shown in the legend. The black vertical line corresponds to  $T_c$  in the  $N \rightarrow \infty$  thermodynamic limit. As  $N$  increases the peak of  $\chi$  increases, sharpens and shifts towards the black line. This is due to finite-size effects.



where  $T_R = (T - T_c)/T_c$  is the reduced temperature, describing proximity to the critical point, and the power-law exponent  $\gamma$  is the critical exponent associated with the susceptibility [40]. For simplicity, only the magnetic susceptibility was considered in the above discussion. Other key divergent quantities include the specific heat, scaling as  $C \sim T_R^{-\alpha}$ , and the correlation length, scaling as  $\xi \sim T_R^{-\nu}$ . The exponents of the critical divergences may be symmetric about  $T_R = 0$  (i.e. the critical point) or diverge with varying exponents as the critical point is approached from above or below.

### 2.2.2 Finite-Size Effects

The power-law divergences of the derivatives of the free energy discussed above are only truly defined in the thermodynamic limit,  $N \rightarrow \infty$ . When  $N$  is finite, so-called *finite-size effects* cause the transitions to become rounded and shifted [41, 42]. Continuing with the 2D Ising model example, in the finite system we observe a rounded peak in  $\chi$ , which occur at temperatures shifted away from  $T_c(N = \infty)$ . We show this behaviour for the simple 2D Ising model in Fig. 2.2. The size and extent of this peak depend on  $N$ , and investigating how a system's correlation length scales with  $N$  allows one to extrapolate the location of the critical point and exponents to the  $N \rightarrow \infty$  limit, see Refs. [43, 47]. Any critical effects we discover in our finite  $N$  systems will be blurred by finite-size effects.

### 2.2.3 Scale-Free Distributions

Power-law distributions such as (2.11) are said to be scale-free, as contributions from all scales are of equal or similar importance. The correlation

length  $\xi \sim T_R^{-\nu}$  in the critical state also follows a power law, meaning that fluctuations in the critical state occur across all length scales, and as a consequence, a system near the critical point will look the same viewed on microscopic, mesoscopic or macroscopic scales. Scale-free power-law distributions turn out to be a ubiquitous property of complex systems, from neuronal activity [48] to human behaviour [49] to the World Wide Web and the Internet [50].

## 2.2.4 Universality

Many diverse systems have been observed to share the same power-law *exponents* at the critical point. This quality is termed universality [40, 47], and models which share the same critical exponents in the thermodynamic limit belong to the same *universality class* [51]. Universality implies that, near the phase transition, the exact microscopic details of a system become unimportant, and global features such as the symmetries of the Hamiltonian, the dimensionality of the problem and the range of the interactions instead predict its physical properties. The Ising model in fact defines one such class - the Ising universality class [52]. Phase transitions in systems as diverse as the liquid-gas transition in simple liquids [38], phase separation in alloys [40], binary fluids [47], colloidal fluids [53, 54] and one component liquid-liquid transitions [55], have all been shown to fall within the Ising universality class. This means that, near the critical point, properties such as the specific heat or the response to external fields of real systems with complex interactions may be correctly predicted from simple toy models such as the Ising model described above, provided one can calculate the critical exponents of the real system.

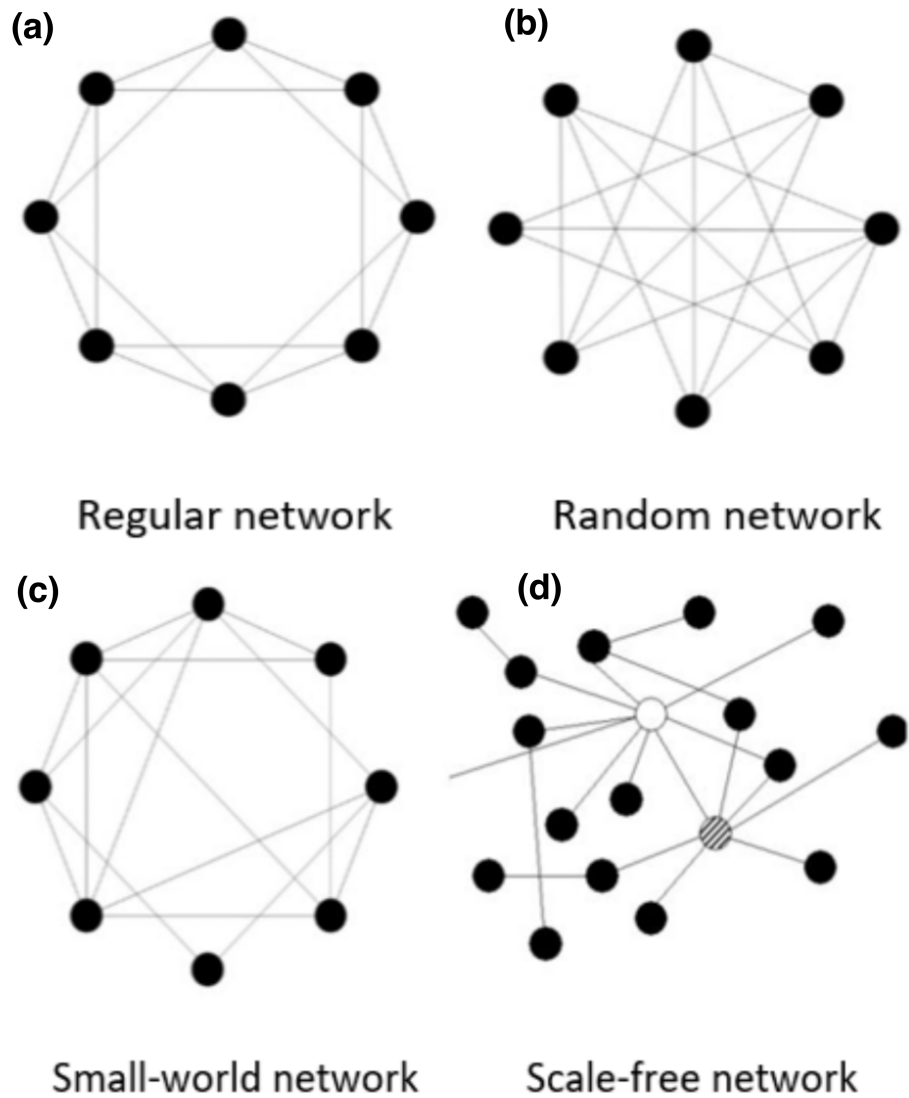
## 2.3 The Brain, Spin-Glasses and Complex Networks

The lattice on which the 2D Ising model exists encodes a specific type of regular (each spin has exactly 4 neighbours) and local (each spin only interacts with its nearest neighbours) network *topology*. Interactions in the brain are both non-local and irregular, indeed we shall see that they appear to form so-called complex networks [25]. In this section, we introduce the basic components required to describe these complex networks and establish how competition between positive and negative interactions on these networks leads to the formation of spin-glasses when they are interpreted in the framework of statistical physics.

### 2.3.1 Complex Networks

Lattice models can be described by structured networks with regular and local interactions. But the vast majority of real-world networks (including brain networks) exist in some other complex arrangement. The study of these *complex networks* [50, 57, 58] exploded around the turn of the century, with two particular topologies, small-world networks [59] and scale-free networks [60], appearing ubiquitously in real-world systems. We sketch these topologies, along with regular and random arrangements, in Fig. 2.3. We

**Figure 2.3:** Example sketches of characteristic network topologies. Nodes are shown by coloured circles, and edges by lines. Panel (a) shows a regular topology. Regular networks have long path lengths (it takes many steps to pass from one node to another) and high clustering coefficients (nodes are locally interconnected). Panel (b) shows a random network. These have short path lengths and low clustering coefficients. Panel (c) shows a small world network, where path lengths are short, and clustering is high. This allows information to rapidly propagate through the network. Panel (d) shows a scale-free network. The degree distribution in scale-free networks follows a power law, causing some nodes to have a very large number of neighbours. These are termed hubs and are shown by white and striped nodes. Information propagates through the network by passing through these hubs. Adapted from [56].



will briefly introduce the key concepts of these networks so we may understand why one might expect the brain to be organised in this way. We begin by establishing the terminology required to describe complex networks:

- A network (or graph) consists of nodes connected by edges.
- The so-called adjacency matrix  $\mathbf{A}$  encodes whether an edge connects two nodes.
- $A_{ij} = 1$  if the nodes  $n_i$  and  $n_j$  are connected,  $A_{ij} = 0$ .
- The network is *un-directed* if  $\mathbf{A}$  is symmetric. We will only consider un-directed graphs here.
- We define the degree of a node  $k_i$  as the sum of all connections;  $k_i = \sum_j A_{ij}$ . The degree measures the “importance” of a node to the network.
- Distances are measured by paths on the network.

We can now think of each of the spins  $s_i$  in our spin ensemble as sitting on one of the nodes  $n_i$  of the network. The couplings  $J_{ij}$  communicate both topology and the strength of the interaction; that is to say, they represent a *weighted* adjacency matrix of the network. For a regular network, such as the 2D lattice studied so far, we find some simple properties; the degree of each node is constant  $k_i = 4$ , and due to the nearest-neighbour locality of the interactions, the average distances between nodes are long. Fully connected networks by definition have high interconnectivity ( $k_i = N - 1$ ) and short path lengths. Small-world and scale-free networks are differentiated from these regular topologies as they have associated degree *distributions*, i.e.  $k_i$  is a random variable.

*Small-world networks* [59] may be generated by taking a lattice-like network and randomly rewiring some of the nearest neighbour connections. Small-world networks show high degrees of clustering and have approximately constant degree distributions and short average path lengths. *Scale-free networks* [60] on the other hand are characterised by degree distributions that follow a power-law,  $P(k_i) \sim k_i^{-\gamma}$ . Some nodes in this network have very few neighbours, while others are connected with the entire network. These highly connected nodes are termed “hubs” and facilitate the flow of information through the network<sup>6</sup>. They also again act to reduce the average path length between two nodes.

### 2.3.2 Complex Brain Networks

Perhaps unsurprisingly, a range of evidence also suggests that the connection structure of the brain is a complex network, with studies showing that different brain networks are both small-world [62–64] and scale-free [65, 66], see [25, 67, 68] for reviews. For neural computation in particular, small world networks are said to enable a range of behaviours including supporting synchronisation [69], rapid learning [70], synchronous processing [71] and the efficient exchange of information [72]. One might then ask how these network structures impact the phase transition we explored in the 2D lattice

<sup>6</sup> Interestingly, the structure of scale-free networks means they are robust to the deletion of random nodes but vulnerable to attacks targeted on the hub nodes [61].

Ising model. Luckily the answer is available; simulations of both small-world [73] and scale-free [74] topologies have shown that both models still undergo a familiar P-F phase transition.

### 2.3.3 Brains as Spin-Glasses

In the simulation of Refs. [73, 74] it was again assumed that  $J_{ij} = J > 0$ , i.e. that all connected couplings were the same, and that these couplings were *excitatory* (causing spins to align). It is well known however, that neural networks require both excitatory and inhibitory connections to function properly [75, 76], with imbalances in the excitatory/inhibitory ratio linked to neuropsychiatric disorders [77]. Our description of the brain must therefore include both excitatory ( $J_{ij} > 0$ ) and inhibitory ( $J_{ij} < 0$ ) couplings. We do not know the exact values of each coupling, all we know is that they are somehow specified by a probability distribution  $P(J_{ij})$ . We will assume that once drawn, each coupling will stay fixed, i.e. that the  $J_{ij}$  do not vary in time. This introduces *quenched disorder* into the system [44], and as we will see briefly, leads to the creation of a new low-temperature phase, the so-called *spin-glass* (SG). The SG phase has many interesting properties [78] and was first linked to neural computation in the 1980s, where it was demonstrated that associative memories could be encoded in the many meta-stable minima of the SG energy landscape [21, 22, 79]. Each spin  $s_i$  now represents the binary state (on/off) of the  $i^{\text{th}}$  neuron or brain region in the network. The components  $h_i$  of the field  $\mathbf{h}$  describe the bias of a particular brain region towards activity or inactivity. The couplings  $J_{ij}$  describe the synaptic strength when modelling individual neurons, or the functional connectivity in coarse-grained neuro-imaging studies. Although the coupling distribution  $P(J_{ij})$  may take any arbitrary form (and encode any arbitrary network topology), the model is still described by the general Hamiltonian of (2.1), which for clarity, we re-print here

$$\mathcal{H}(\mathbf{s}|\mathbf{h}, \mathbf{J}) = - \sum_i h_i s_i - \frac{1}{2} \sum_{i \neq j} J_{ij} s_i s_j. \quad (2.12)$$

The Hamiltonian shown in (2.12) is often also referred to as the Sherrington-Kirkpatrick (SK) Hamiltonian [35, 80], after the pair who first solved and understood the phase diagram of the fully connected (mean-field) spin-glass. We will now provide a surface-level introduction to this model.

## 2.4 The Sherrington-Kirkpatrick model

The SK model [35, 80] was originally proposed as an exactly solvable model with which to explain experimental [81] and theoretical evidence [82] of a disordered magnetic phase, the spin-glass phase, in which competing interactions cause the orientation of spins to be "frozen" in random orientations at low temperatures. The SK model is a so-called "mean-field" or infinite-dimensional model, where every spin interacts with every other spin in the system. Such models are generally easier to investigate mathematically, and while the SK model can be solved analytically [35, 80, 83–86], the mathematics by which this is achieved, namely the replica trick [35, 44, 78], are

beyond the scope of this work and will not be given here. We will use results from the SK model, such as the expected structure of the phase diagram, to contextualise how we might expect models built from brain data to behave. Brain data is dynamic (i.e. it fluctuates) and we thus expect that SK-like models built to represent this data will sit outside of the spin-glass regime, e.g. see [13]. The precise properties of the low-temperature frozen SG phase are therefore of limited relevance for the work presented in this thesis, and we only briefly introduce these for completeness.

As before, our ensemble of spins interacts via the Hamiltonian of (2.12). In general, both  $\mathbf{h}$  and  $\mathbf{J}$  are now treated as random variables drawn from some probability distribution. We will consider the original model formulated by Sherrington and Kirkpatrick [35] in which all fields are zero ( $\mathbf{h} = 0$ ) and the couplings  $J_{ij}$  are independent and identically distributed (i.i.d) random variables drawn from a Gaussian distribution with mean

$$\mu_J = \mathbb{E}[J_{ij}] = \mu/N, \quad (2.13)$$

and standard deviation

$$\sigma_J = \sqrt{\text{Var}[J_{ij}]} = \sigma/N^{1/2}, \quad (2.14)$$

where  $\mu$  and  $\sigma$  are intensive variables. The state of the system is controlled by the dimensionless averages  $\mu/\sigma$  and  $T/\sigma$ , and different phases are explored by tuning these parameter ratios. We note that the distribution  $P(J_{ij})$  need not be Gaussian, indeed a common alternative is to instead take  $J_{ij} = \pm 1$  with equal probability [19, 87], and it has been proven that any distribution which for  $N \rightarrow \infty$  has the same first and second moments as (2.13) and (2.14), and bounded non-infinite higher moments, will produce the same phase diagram [88]. Indeed recent computational studies using empirical brain networks have shown that even these complex networks produce phase diagrams which closely mimic the SK phase diagram [13].

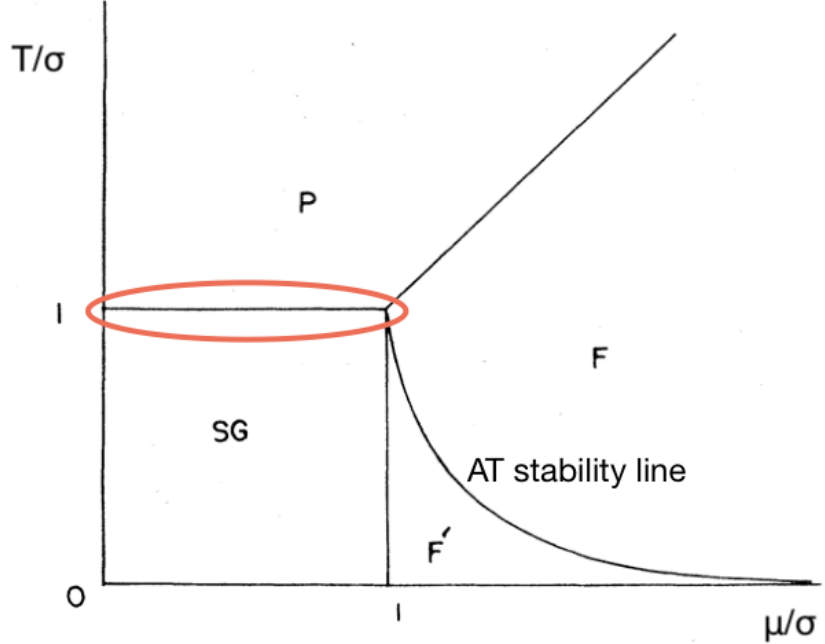
### 2.4.1 The Spin-Glass Order Parameter

We previously introduced the magnetisation order parameter,  $m$ , to describe the paramagnetic (P) - ferromagnetic (F) transition in the Ising model. For clarity,  $m$  is defined as

$$m = \frac{1}{N} \sum_i m_i, \quad (2.15)$$

where  $m_i = \langle s_i \rangle$  are the “local” magnetisation’s of each spin. The ordered (low temperature) F phase is characterised by spins which are globally aligned in the same direction:  $m = 1$ . These spins are frozen, i.e. they do not fluctuate and remain globally ordered. In the P phase, the spins randomly fluctuate, and there is no global alignment:  $m = 0$ . The SG phase however is characterised by spins which are frozen in random orientations and thus  $m = 0$  here also. The magnetisation  $m$  is therefore unable to distinguish between P and SG phases, and we need to introduce another order parameter to describe this transition. We define the spin-glass order parameter, the (Edwards-Anderson) *overlap*  $q$ , as [82]

**Figure 2.4:** Phase diagram of the zero-field SK model in the thermodynamic limit. The model exhibits four phases; these are the paramagnetic (P), ferromagnetic (F), replica symmetry breaking ferromagnetic (F') and low-temperature spin-glass (SG) phases. The AT stability line indicates the onset of replica symmetry breaking. Highlighted in red is the P-SG transition line, along which SK-like models inferred from fMRI neuroimaging data have been recently found to lie [13]. The phase diagram is adapted from [19].



$$q = \frac{1}{N} \sum_{i=1}^N \langle s_i \rangle^2. \quad (2.16)$$

The spin-glass order parameter again also has an associated susceptibility  $\chi_{SG}$ , which we define as [89]

$$\chi_{SG} = \frac{1}{NT^2} \sum_{i,j=1}^N C_{ij}^2, \quad (2.17)$$

and which diverges as spontaneous large correlations develop near the SG phase boundaries. Here

$$C_{ij} = \langle s_i s_j \rangle - \langle s_i \rangle \langle s_j \rangle, \quad (2.18)$$

is the pairwise correlation between spins  $i$  and  $j$  and the sum  $i, j$  in (2.17) runs over all pairs of  $i$  and  $j$ .  $q = 1$ ,  $m = 1$  in the F phase,  $q > 0$ ,  $m = 0$  in the SG phase and  $q = 0$ ,  $m = 0$  in the P phase.

### 2.4.2 Phase Diagram

By tuning the parameter ratios  $\mu/\sigma$  and  $T/\sigma$  one can access each of these phases, and the corresponding phase diagram of the  $N \rightarrow \infty$  SK model is shown in Figure 2.4. Many details have been omitted here, but loosely speaking there are two solutions to the SK model, the replica symmetric solution [35] and Parisi's full replica symmetry breaking (fullRSB) solution [78, 86]. The replica symmetric solution is valid in the paramagnetic (P) phase, and in the ferromagnetic (F) phase above the de Almeida-Thouless (AT) stability line [83]. Upon crossing below the AT line, the replica symmetric solution becomes unstable and gives a negative entropy which is

non-physical for a model of Ising spins. Instead, a new solution (referred to as the fullRSB solution) correctly describes this low temperature / weak coupling strength region of phase space. At low temperatures, the fullRSB solution gives rise to a spin-glass (SG) phase, the states of which are organised hierarchically and which corresponds to the richest possible organisation of phase space [90]. Recent work [13] found Ising models inferred from large-scale neurological imaging data to sit in the P phase along the P-SG phase boundary highlighted in Figure 2.4, and proposed that the brain may derive some of its computational ability by exploiting the complex energy landscape near this transition. Further evidence suggesting proximity to this transition is required to provide merit to this idea, and these findings form the starting point of the work presented in this thesis. In particular, we will end up performing an in-depth analysis of the inference of data from this region of phase space which will allow us to confidently appraise results from real-world datasets.



## Summary of Chapter 2

- We introduced the basics of equilibrium statistical mechanics required to understand and interpret our models. We considered models of binary *spins* exclusively.
- The statistics of an equilibrium system are governed by the *Boltzmann distribution* and the macroscopic properties of a system are given by averages with respect to this distribution. Complex systems can undergo *phase transitions*, which we characterise by measuring *order parameters* and *susceptibilities*. The point where the system undergoes a second-order phase transition is called the *critical point*.

### Order Parameters

Order parameters allow us to characterise the macroscopic state of a system. They change continuously near second-order phase transitions.

### Susceptibilities

Susceptibilities measure the response of a system to external stimuli. They are second-order derivatives of the free energy and diverge near second-order phase transitions.

### Finite Size Effects

Transitions between phases are only truly defined in the *thermodynamic limit*, where the system size tends to infinity. For finite-sized systems, transitions are blurred and shifted, and extended critical-like regimes appear.

- We discussed that connections in the brain form a *complex network* and that the coupling matrix  $\mathbf{J}$  of a generalised Ising model may be interpreted as a weighted adjacency matrix describing the topology of a given model.
- Connections in the brain can be both inhibitory and excitatory, introducing *quenched disorder* into the system. Systems with the quenched disorder tend to form *spin-glasses* at low temperatures.

### The Sherrington-Kirkpatrick Model

The SK model is an archetypal spin-glass model. It is a fully connected model where all spins interact with all other spins, and has an analytical solution. We later use this model to investigate our inference method as we expect models inferred from neural data to explore a similar phase diagram.

## Chapter 3

# Inverse Ising Inference

In this chapter, we will introduce methods which allow one to *learn* (i.e. infer) parameters of equilibrium statistical physics models from real data in a process known as maximum entropy modelling. By measuring pairwise correlations in binary data one reconstructs models akin to the Sherrington-Kirkpatrick model introduced in the previous chapter, i.e. one reconstructs generalised Ising models, where the parameters  $h_i$  and  $J_{ij}$  are random variables drawn from some unknown distribution. Pairwise maximum entropy modelling is therefore also commonly referred to as *inverse Ising inference* when applied to binary data. We introduce a number of solutions to the inverse Ising problem in this chapter. We then detail how pairwise maximum entropy models may be used for network reconstruction, and how networks inferred via such inverse methods are superior to those constructed from correlations. We finish this chapter by providing an overview of how maximum entropy modelling has been previously used to better understand a variety of complex systems.

### 3.1 Maximum Entropy Models

In this section, we will show how statistical mechanics may be re-derived purely from information-theoretic considerations. The motivation for this is simple; doing so frees the theory from being a theory of *physics* describing interactions between real atoms and molecules in a volume, and allows the construction of a more general probabilistic theory that describes the behaviour of any collection of random variables. In this formalism equilibrium statistical physics corresponds to calculating the probability distribution which maximises the Shannon entropy [91] under some constraints set by quantities we measure. By maximising this entropy we chose to represent the data by a probability distribution that is exactly equivalent to the Boltzmann equilibrium distribution obtained in classical statistical mechanics.

This reformulation was first done in 1957 by Jaynes [36, 92], and we will largely be following the same arguments in the section below. We begin by considering a discrete random variable  $x$ , which can take values  $x_u$  for  $u = 1, 2, \dots, U$ , with probability  $P(x_u)$ . The precise distribution of  $P(x)$  is unknown. The only information available about the system are the expectation values of  $v = 1, 2, \dots, V$  observable functions  $f_v(x)$

$$\langle f_v(x) \rangle = \sum_{u=1}^U P(x_u) f_v(x_u), \quad (3.1)$$

with the usual normalisation condition that the probabilities must sum to one

$$\sum_u P(x_u) = 1. \quad (3.2)$$

The Shannon entropy [91] of the probability distribution  $P(x)$  defines the “amount of uncertainty” represented by that probability distribution and is defined as

$$S[P(x)] = -K \sum_u P(x_u) \ln P(x_u), \quad (3.3)$$

where  $K$  is a positive constant. Equation (3.3) is simply the usual expression for the entropy in statistical mechanics, and thus the two may be considered synonymous. Given that we have information about the system in the form of the expectation values of equation (3.1) and the normalisation condition of (3.2), we can calculate the probability distribution  $P$  that maximises  $S$  under the constraints (3.1) and (3.2) through the standard method of Lagrange multipliers [93]. One multiplier  $\lambda_0$  is introduced to account for the normalisation constraint (3.2). Each observation made on the system also introduces another multiplier, such that for each of the  $V$  measurements  $\langle f_v(x) \rangle$  in (3.1) we introduce another multiplier,  $\lambda_v$ . Performing this optimisation gives the maximum entropy probability distribution as:

$$P(x_u) = \exp\left(-\lambda_0 - \sum_{v=1}^V \lambda_v f_v(x_u)\right). \quad (3.4)$$

To determine the multiplying constants one substitutes (3.4) into the constraint conditions (3.1) and (3.2). The Lagrange multiplier related to the normalisation condition (3.2) is given by

$$\lambda_0 = \ln Z(\lambda_1, \dots, \lambda_V), \quad (3.5)$$

where  $Z$  is the so-called partition function

$$Z(\lambda_1, \dots, \lambda_V) = \sum_u e^{-\sum_v \lambda_v f_v(x_u)}, \quad (3.6)$$

which acts as a normalisation constant on  $P(x)$ . Similarly, the multipliers associated with the measurements in (3.1) are given by

$$\langle f_v(x) \rangle = -\frac{\partial \ln Z}{\partial \lambda_v}, \quad (3.7)$$

and we see that the expectation values of the measurement functions are related to derivatives of the partition function. Substituting the probability distribution found in (3.4) into (3.3), we find the solution for the maximum entropy as:

$$S_{\max} = K \left( \lambda_0 + \sum_v \lambda_v \langle f_v(x) \rangle \right). \quad (3.8)$$

### 3.1.1 Recovering the Canonical Ensemble

To demonstrate that the equations presented here are indeed mathematically identical to their counterparts in statistical physics, let us begin by defining a system with an energy given by

$$E_u = E(x_u | \lambda_1, \dots, \lambda_V) = \sum_v \lambda_v f_v(x_u). \quad (3.9)$$

Given that it is only possible to measure the average energy of our system  $\langle E(x) \rangle$ , Eq. 3.9 simplifies to  $E_u = \lambda E(x_u)$ , and the maximum entropy probability of equation (3.4) reduces exactly to the familiar equilibrium Boltzmann distribution [38]

$$P(x_u) = \frac{1}{Z} e^{-\lambda E_u}. \quad (3.10)$$

There is now only one conjugate Lagrange multiplier (associated with the energy) which we identify as the inverse temperature,  $\lambda = \beta = (k_b T)^{-1}$ . Using the relation for the maximum entropy in (3.8) with  $K = k_b$ , one finds the usual expression for the free energy:

$$F = -(1/\beta) \ln Z = \langle E \rangle - TS. \quad (3.11)$$

### 3.1.2 Recovering the Grand Canonical Ensemble

If one imposes additional constraints on the system, say by measuring the average number of particles  $\langle N(x) \rangle$ , the maximum entropy probability distribution changes accordingly

$$P(x_u) = \frac{1}{Z} e^{-[\lambda_1 E(x_u) + \lambda_2 N(x_u)]}, \quad (3.12)$$

with  $\lambda_1 = \beta$  and  $\lambda_2 = -\beta\mu$ , where  $\mu$  is the chemical potential. Calculating the new free energy with this additional constrain using (3.8) gives

$$F = -(1/\beta) \ln Z = \langle E \rangle - TS - \mu \langle N \rangle. \quad (3.13)$$

Equations (3.12) and (3.13) are exactly those of the probability distribution and free energy describing the grand canonical ensemble [38].

### 3.1.3 Recovering the Generalised Ising Model

Equations (3.10) and (3.12) serve as two simple examples of calculating the maximum entropy distribution under given constraints in the information-theoretic formalism. Let us now return to a system of interacting binary spins and consider the constraints which would produce the maximum entropy probability distribution of a generalised Ising model

$$P(\mathbf{s}_u) = \frac{1}{Z} \exp \left\{ - \sum_i h_i s_{i,u} - \frac{1}{2} \sum_{i,j} J_{ij} s_{i,u} s_{j,u} \right\}. \quad (3.14)$$

Note that we have absorbed  $T$  into the definition of the fields and couplings. Equation 3.14 is comprised of two types of conjugate Lagrange multipliers: the fields  $h_i$  which act on individual microscopic spins  $s_i$  and the pairwise

couplings  $J_{ij}$  which act on pairwise correlations between spins  $s_i s_j$ . To introduce these multipliers one must therefore measure the means  $\langle s_i \rangle$  and pairwise-correlations  $\langle s_i s_j \rangle$  of the spin distribution respectively. Systems which follow the generalised Ising Hamiltonian (i.e. SK-like systems) are therefore *pairwise* Maximum Entropy Models (MEMs).

### 3.2 Inferring Ising Models

Measuring the first and second moments of a spin distribution corresponds to selecting a generalised (SK-like) Ising model in the maximum entropy framework. We now explore how to learn the specific model parameters  $\mathbf{h}$  and  $\mathbf{J}$  that best represent a given dataset. This process is known as *pairwise* maximum entropy modelling or inverse Ising inference<sup>7</sup>. We again consider a system of  $N$  interacting binary spins  $s_i \in \pm 1$ ,  $i = 1, \dots, N$ , for each of which we have a time series of length  $B$ . We describe the state of the system at a given time  $t$  as a vector  $\mathbf{s}(t)$ . We label the full dataset of  $B \times N$  observations as  $\{\mathbf{s}\}^B = \{\mathbf{s}(t)\}_{t=1}^B$ . From this dataset, we measure the empirical averages (local magnetisations)

$$m_i^e = \langle s_i \rangle^e, \quad (3.15)$$

and pairwise covariances

$$C_{ij}^e = \langle s_i s_j \rangle^e - \langle s_i \rangle^e \langle s_j \rangle^e, \quad (3.16)$$

where  $\langle A \rangle^e = \frac{1}{B} \sum_{t=1}^B A(t)$  is the *empirical* mean evaluated from the dataset. We know that by measuring the (3.15) and (3.16) we constrain the maximum entropy model describing the data to a generalized equilibrium Ising model. The inverse Ising problem, therefore, consists in determining the vector of fields  $\mathbf{h}$  and symmetric matrix of couplings  $\mathbf{J}$  from the empirical averages  $\mathbf{m}^e$  and correlations  $\mathbf{C}^e$  so that

$$\langle s_i \rangle = \langle s_i \rangle^e, \quad (3.17)$$

and

$$\langle s_i s_j \rangle = \langle s_i s_j \rangle^e. \quad (3.18)$$

Here  $\langle A \rangle$  indicates taking the thermodynamic average of the inferred Ising model, i.e. the average with respect to

$$P(\mathbf{s}|\mathbf{h}, \mathbf{J}) = \frac{1}{Z(\beta, \mathbf{h}, \mathbf{J})} \exp\{-\beta \mathcal{H}(\mathbf{s}|\mathbf{h}, \mathbf{J})\}, \quad (3.19)$$

where

$$\mathcal{H}(\mathbf{s}|\mathbf{h}, \mathbf{J}) = - \sum_i h_i s_i - \frac{1}{2} \sum_{i \neq j} J_{ij} s_i s_j. \quad (3.20)$$

A number of techniques exist to solve the conditions (3.17) and (3.18) and in what follows we provide a historical introduction to these. For notational convenience we now absorb the temperature into the definition of our parameters, i.e. we set  $\mathbf{h} = \beta \mathbf{h}$  and  $\mathbf{J} = \beta \mathbf{J}$ . This is equivalent to working in

<sup>7</sup> A brief note to explain the *inverse* terminology used here. In statistical mechanics, in the *forward problem* the underlying model (e.g. the parameter set  $\boldsymbol{\theta}$ ) is known in full and one aims to calculate averages of some set of observables  $\langle f_r(x) \rangle$  (e.g.  $\langle s_i \rangle$ ) from this known model. The usual objective of statistical mechanics is to solve the forward problem. The aim of the inverse problem is to do the opposite; to estimate the model  $\boldsymbol{\theta}$  which best represents some set of known observations measured from the data.

units of  $T = 1$ . This means a “low temperature” state point is functionally equivalent to a “strongly coupled” state point.

### 3.2.1 Exact Likelihood Maximisation: Boltzmann Learning

Motivated by the emergent computational properties (content addressable memories etc) of SK-like models [21, 79, 80], Ackley *et al.* [94] introduced a simple maximum likelihood estimator with which to solve the inverse Ising problem. To construct this, we assume that the  $B$  samples of our dataset  $\{\mathbf{s}\}^B$  are drawn in an independent and identically distributed (i.i.d.) fashion from Eq. 3.19, so that we may write the log-likelihood of observing the dataset as

$$\begin{aligned} \mathcal{L}(\{\mathbf{s}\}^B|\mathbf{h}, \mathbf{J}) &= \frac{1}{B} \sum_{t=1}^B \ln [P(\mathbf{s}(t)|\mathbf{h}, \mathbf{J})] \\ &= -\ln Z + \sum_i h_i \frac{1}{B} \sum_t s_i(t) + \frac{1}{2} \sum_{i \neq j} J_{ij} \frac{1}{B} \sum_t s_i(t) s_j(t) \\ &= \sum_i h_i m_i^e + \frac{1}{2} \sum_{i \neq j} J_{ij} (m_i^e m_j^e + C_{ij}^e) - \ln Z. \end{aligned} \tag{3.21}$$

The set of parameters  $\{\mathbf{h}^*, \mathbf{J}^*\}$  which maximise Eq. 3.21 is known as the maximum likelihood estimate (MLE) of the inference problem [17]<sup>8</sup>. To find these one must obtain the gradients of  $\mathcal{L}$  with respect to the fields and couplings. Due to the linearity of (3.21) this is fortunately rather simple, and we find that

$$\begin{aligned} \frac{\partial \mathcal{L}(\{\mathbf{s}\}^B|\mathbf{h}, \mathbf{J})}{\partial h_i} &= \langle s_i \rangle^e - \frac{\partial \ln Z(\mathbf{h}, \mathbf{J})}{\partial h_i} = \langle s_i \rangle^e - \langle s_i \rangle, \\ \frac{\partial \mathcal{L}(\{\mathbf{s}\}^B|\mathbf{h}, \mathbf{J})}{\partial J_{ij}} &= \langle s_i s_j \rangle^e - \frac{\partial \ln Z(\mathbf{h}, \mathbf{J})}{\partial J_{ij}} = \langle s_i s_j \rangle^e - \langle s_i s_j \rangle. \end{aligned} \tag{3.22}$$

The MLE solution is obtained when the gradients of (3.22) are zero (i.e. when  $\mathcal{L}$  is maximised), in which case the conditions (3.17) and (3.18) of the inverse Ising problem are also satisfied. The parameters  $\{\mathbf{h}^*, \mathbf{J}^*\}$  which maximise the likelihood defined in Eq. 3.21 therefore provide the solution to the inverse Ising problem. When the number of spins  $N$  is very small (typically a few tens), it is computationally feasible to enumerate  $Z$  and perform this optimization analytically. However, as calculating  $Z$  requires summing over all  $2^N$  possible configurations of the system, the problem becomes rapidly intractable with increasing  $N$ , and a range of alternative methods have been proposed to perform the maximisation. The oldest of these is so-called Boltzmann learning [17, 94, 96] which uses Monte Carlo (MC) simulations [97, 98] to approximate the gradients of (3.22) and update them via a simple gradient ascent scheme [99, 100]. In this scheme, one updates the parameter estimates via

<sup>8</sup> Maximising the likelihood is equivalent to minimizing the Kullback–Leibler (KL) divergence [95], which characterises the difference between two probability distributions.

$$\begin{aligned} h_i^{n+1} &= h_i^n + \alpha [\langle s_i \rangle^e - \langle s_i \rangle(\mathbf{h}^n, \mathbf{J}^n)], \\ J_{ij}^{n+1} &= J_{ij}^n + \alpha [\langle s_i s_j \rangle^e - \langle s_i s_j \rangle(\mathbf{h}^n, \mathbf{J}^n)], \end{aligned} \tag{3.23}$$

where  $\alpha$  is a small constant known as the *learning rate* and  $n$  labels the current step of the iterative parameter update process. A pseudo-algorithm describing the complete Boltzmann learning scheme is shown in Algorithm 1.

---

**Algorithm 1:** Boltzmann Learning

---

1. Initialise MLE parameters:  $\{\mathbf{h}^*, \mathbf{J}^*\} \leftarrow \{\mathbf{h}^n, \mathbf{J}^n\}$
  2. Estimate  $\langle s_i \rangle$  and  $\langle s_i s_j \rangle$  at  $\{\mathbf{h}^n, \mathbf{J}^n\}$  through MC sampling
  3. Apply parameter update rule (3.23) giving  $\{\mathbf{h}^{n+1}, \mathbf{J}^{n+1}\}$
  4. Set  $\{\mathbf{h}^n, \mathbf{J}^n\} \leftarrow \{\mathbf{h}^{n+1}, \mathbf{J}^{n+1}\}$  and repeat from step 2 until parameters converge to constant value (within some tolerance)
  5. Return  $\{\mathbf{h}^*, \mathbf{J}^*\}$
- 

It is theoretically possible to estimate  $\{\mathbf{h}, \mathbf{J}\}$  with unbounded accuracy using Boltzmann learning (i.e. the method is asymptotically exact). This highlights an interesting property of (3.21): the averages  $\langle s_i \rangle^e$  and  $\langle s_i s_j \rangle^e$  form sufficient statistics [101–103] for the maximum likelihood estimate, and one cannot obtain better estimates of  $\{\mathbf{h}, \mathbf{J}\}$  by using all the  $B \times N$  available data points than by just measuring  $\langle s_i \rangle^e$  ( $N$  observations) and  $\langle s_i s_j \rangle^e$  ( $N(N-1)/2$  independent observations) [15]. While the accuracy may be unbounded, the update steps (Step 3, Algorithm 1) require re-evaluating the partition function through MC simulations, which is a computationally hard problem. More generally, it has been shown that methods which rely on sufficient statistics (such as Boltzmann learning) may gain accuracy at the cost of computational complexity [104], and in real applications, this has meant that Boltzmann learning is limited to small ( $N \sim 120$  [105]) system sizes. A range of approximate solutions to the inverse Ising problem have been developed to overcome this limitation, which we will now go on to discuss.

### 3.2.2 Statistical Physics Based Approximations

In this section, we introduce some analytical approximations with which to solve the inverse Ising problem. These are predominantly based on the plethora of results developed to understand the statistical physics of spin-glasses throughout the 20<sup>th</sup> century [78, 89]. Although these approximations are computationally fast, they generally produce larger errors than the exact method of Boltzmann learning, and it is well known that they provide poor performance in low-temperature (strong-coupling) or small-sample size regimes (see [16, 106, 107] for reviews) when compared to Boltzmann learning. We will not use any of these techniques in this thesis and simply highlight them here for completeness. The foundation of these methods is

to construct mean-field theories [38] of varying complexity, and to use these to estimate  $\langle s_i \rangle$  and  $\langle s_i s_j \rangle$ . Some of the most popular mean-field approaches (in order of increasing refinement) are naive mean field (nMF) methods [108, 109], Thouless-Anderson-Palmer (TAP) inversion [109, 110], Bethe approximations [111], susceptibility propagation [112–114], small correlation expansions [115] and cluster expansions [116, 117]. Bar the cluster expansion method (which is computationally intense [118]), all these mean-field reconstructions are known to fail near the transition to the low-temperature phase [107, 112, 117], and we predominately regard them as exploratory methods with which to obtain simple and reasonably representative initial estimates of the model parameters [107]. These can then be used to initialise the exact Boltzmann learning scheme or other more complex inference schemes so that fewer iterations are required for these to converge.

While we do make use of any of these mean-field methods in this work, we will provide a brief introduction to the simple nMF approximation to establish how these methods are formulated. One begins by defining a set of mean-field equations for the external fields of the generalized Ising model, which in the nMF case is [39, 89]

$$h_i = \tanh^{-1}(m_i^e) - \sum_k J_{ik} m_k^e. \quad (3.24)$$

As we are building a mean-field theory for a specific dataset we make explicit reference to the empirical magnetisations  $m_i^e$ . Using the linear response (fluctuation-dissipation) theorem [39], one may relate the pairwise correlations  $C_{ij}^e$  to derivatives of (3.24) [109, 119], such that

$$C_{ij}^e = \frac{\partial m_i^e}{\partial h_j} \quad \text{and} \quad ((C^e)^{-1})_{ij} = \frac{\partial h_i}{\partial m_j^e}, \quad (3.25)$$

where  $((C^e)^{-1})_{ij}$  are the elements of the inverse correlation matrix. Differentiating (3.24) w.r.t.  $m_j^e$  gives

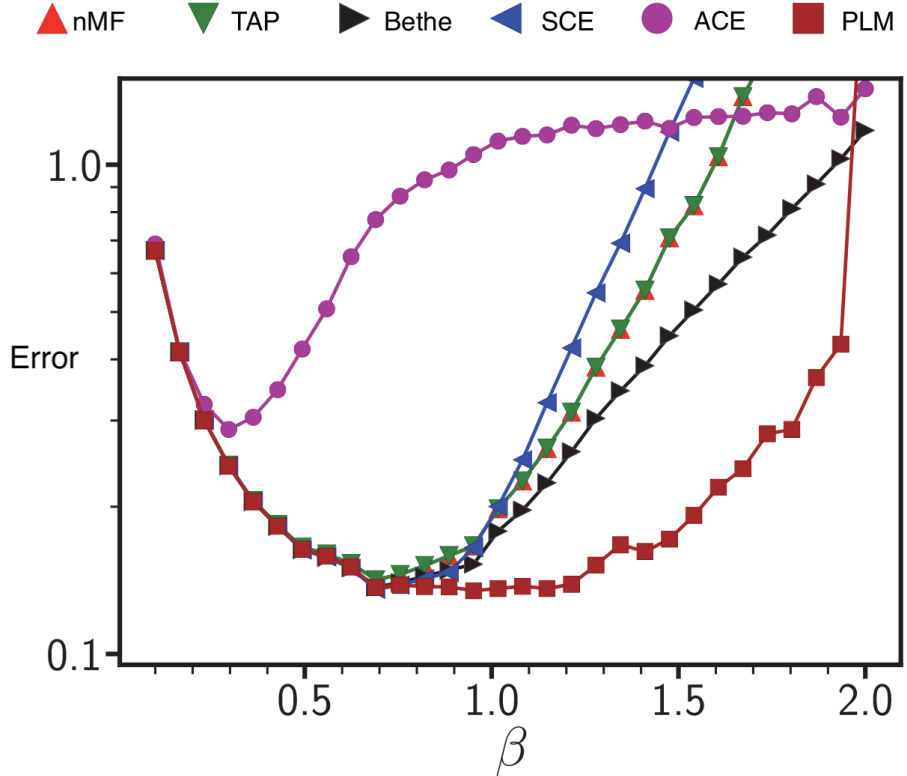
$$\begin{aligned} ((C^e)^{-1})_{ij} &= \frac{\partial h_i}{\partial m_j^e} = \frac{1}{1 - (m_i^e)^2} \delta_{ij} - \sum_k J_{ik} \delta_{kj}, \\ &= \frac{1}{1 - (m_i^e)^2} \delta_{ij} - J_{ij}. \end{aligned} \quad (3.26)$$

As there are no-self couplings in the model (we only calculate  $J_{ij}$  for  $i \neq j$ ),  $\delta_{ij}/(1 - (m_i^e)^2) = 0$  for all valid couplings. Thus, the nMF estimate for the couplings matrix  $\mathbf{J}^{\text{nMF}}$  is simply

$$\mathbf{J}^{\text{nMF}} = -(\mathbf{C}^e)^{-1}. \quad (3.27)$$

To calculate the nMF estimates of  $\mathbf{h}$  and  $\mathbf{J}$  one simply has to invert  $\mathbf{C}^e$ , use (3.27) to estimate  $\mathbf{J}$  and then substitute these couplings into (3.24) to obtain  $\mathbf{h}$ . More advanced mean-field reconstructions are formulated identically, but replace the initial mean-field equations (3.24) with more refined estimates, such as the TAP mean-field equation [110]. In the next section we detail our method of choice, so-called pseudo-likelihood maximisation, which has been shown to outperform the above approximate schemes across a range of





**Figure 3.1:** Comparison of different inverse Ising solvers for data generated from an SK model of size  $N = 64$ .  $\beta$  is the inverse temperature. Different coloured lines correspond to different methods. See the text for a description of how nMF, TAP and Bethe methods are constructed. SCE indicates the small correlation expansion [115], ACE the adaptive cluster expansion [116] and PLM the pseudo-likelihood maximisation [15, 120] method we use. We see PLM performs better or as well as all other methods for all  $\beta$ . The spin-glass phase is at high  $\beta$ . This figure is adapted from Ref. [16].

systems and coupling strength regimes [15, 16], see Fig. 3.1 for an example in the SK model.

### 3.2.3 Pseudo-likelihood Maximisation

A powerful alternative solution, and the inverse Ising solver we choose for our analysis, is pseudo-likelihood maximisation (PLM) [15]. In PLM one approximates the inverse Ising log-likelihood (3.21) by a set of  $N$  pseudo-log-likelihoods  $\mathcal{L}_r$  [120]

$$\mathcal{L}_r(h_r, \mathbf{J}_r | \{\mathbf{s}\}_B) = \frac{1}{B} \sum_{t=1}^B \ln P_{\{h_r, \mathbf{J}_r\}}(s_r(t) | \mathbf{s}_{\setminus r}(t)), \quad (3.28)$$

which depend only on the parameter  $h_r$  and the  $r^{\text{th}}$  row of entries  $\mathbf{J}_r = \{J_{rj}\}_{j \neq r}$  to the coupling matrix. Here we additionally introduced the conditional probability distribution

$$P_{\{h_r, \mathbf{J}_r\}}(s_r | \mathbf{s}_{\setminus r}) = 1 / (1 + e^{-2s_r[h_r + \sum_{j \neq r} J_{rj} s_j]}), \quad (3.29)$$

corresponding to the probability of observing the  $r^{\text{th}}$  spin in state  $+1$  or  $-1$  given all other  $N - 1$  spins. Note that in the limit of  $B \rightarrow \infty$  the PLM approach to inverse Ising inference is exact. Each of the  $\mathcal{L}_r$  can be maximised independently for each spin, making the problem highly suitable for parallelisation. Moreover, the structure of the pseudo-likelihood means that each PLM optimisation is formally identical to logistic regression for which efficient computational algorithms exist<sup>9</sup>. As such, PLM

<sup>9</sup> For this work, we perform each logistic regression using the `sklearn.linear_model.LogisticRegression` classifier from the *Scikit-learn* [121] Python package with the L-BFGS-B optimizer [122–124].

depends on “all the data” [15], i.e. it is not a method based on sufficient statistics and does not require the partition function to be computed. Note that the coupling matrix inferred using PLM is generally not symmetric. We therefore always perform a post-inference symmetrising step, setting  $\mathbf{J}^* = \frac{1}{2}[\mathbf{J}_{PLM}^* + (\mathbf{J}_{PLM}^*)^T]$  where  $T$  is the transpose of the matrix. PLM reconstructions have been shown to outperform other approximate methods for a range of topologies and coupling strength regimes [15, 16]. Moreover, PLM may be efficiently implemented due to its inherent parallelism and simple logistic regression form, making it possible to apply this method to large system sizes. PLM is therefore considered the state-of-the-art method with which to solve the inverse Ising problem, and this thesis focuses on *performing an in-depth study of PLM and its application to real data*. As a final comment, there is an alternative non-mean-field approximate inverse Ising solver termed minimum probability flow [125]. PLM generally outperforms this method when applied to real neurological data [126] and we will thus not discuss it any further.

### 3.3 Network Reconstruction

One common aim of fMRI studies is to understand the network structure of the brain. This is usually termed the *functional connectivity* (FC) of the brain, and two regions are said to be functionally connected (i.e they fulfil the same function) if there is some statistical correlation between the activity time series from those regions [25, 67, 127]. Inverse Ising methods provide a new pathway towards establishing functional connectivity, which is represented by the couplings  $\mathbf{J}$  of the inferred model. We now compare how coupling networks constructed by inverse Ising inference differ from the typical correlation-based networks encountered in the neuroimaging literature. While many complex measures of correlation have been defined (see [128] for an exhaustive overview) studies which compare these correlative methods [129–131] find that no singular best method exists for all use cases. As such, the simplest measure of correlation, the linear (Pearson) correlation coefficient

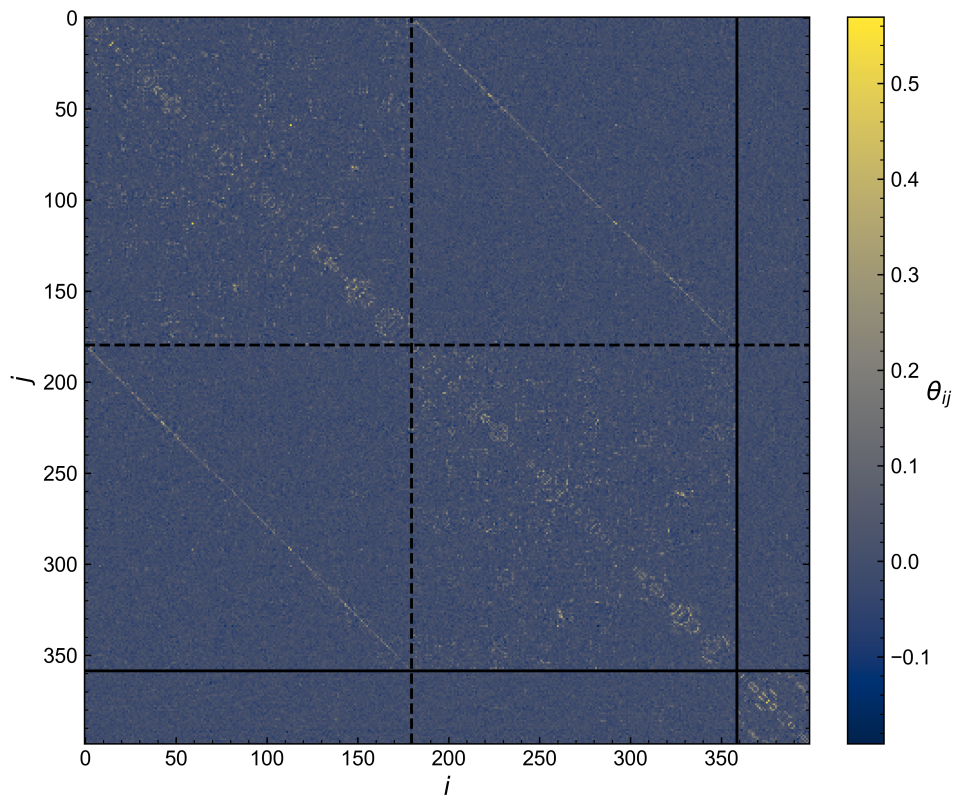
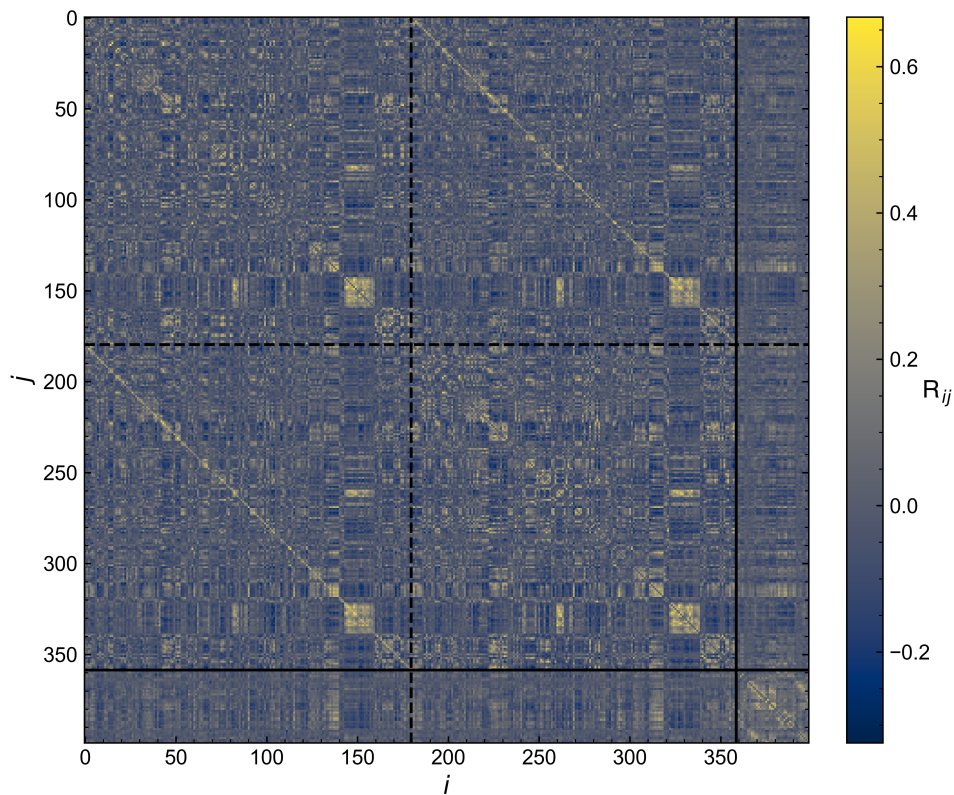
$$R_{ij} = \frac{C_{ij}^e}{\sigma_i^e \sigma_j^e} = \frac{\langle s_i s_j \rangle^e - \langle s_i \rangle^e \langle s_j \rangle^e}{\sigma_i^e \sigma_j^e}, \quad (3.30)$$

still remains widely used to construct the FC of neurological data [62, 66, 132–134]. In (3.30)  $\sigma_i^e$  represents the standard deviation of the  $i^{th}$  region over the time series of the dataset.  $R_{ij}$  varies between values of +1 and -1, with values of +1 implying perfect correlation, -1 perfect anti-correlation and 0 no correlation. Usually,  $R_{ij}$  or  $|R_{ij}|$  is then thresholded via some scheme to construct the adjacency matrix of the graph: i.e. all  $R_{ij} > \delta$  where  $\delta$  is a threshold value are connected. This thresholding process remains contentious [128, 135] as only considering positive correlations, or correlation magnitudes does not allow the excitatory/inhibitory nature of the connection to be assessed. Moreover, negative interactions are known to help modulate brain dynamics, and their functional importance has been demonstrated in neuroimaging studies [132, 136].

Regardless of which thresholding scheme is used, the process of establishing FC described above should immediately trigger some level of concern: *correlations do not imply connections*. Even in the simple nMF inference scheme (3.27) couplings are related to the inverse of correlations. Indeed, let us consider a simple 1D Ising model with only nearest-neighbour interactions (i.e. a chain of spins). At low temperatures this system develops long-range exponentially decaying correlations; a connectivity network of this model defined through correlations would therefore look much more inter-connected than the true, sparse nearest neighbour couplings would suggest. We demonstrate this enhanced connectivity effect for one of our human fMRI datasets in Fig 3.2. More generally, critical points correspond to regimes of large, long-range correlations and we thus expect that even simple (e.g. 2D Ising model) coupling topologies can lead to complex, interconnected correlation topologies. Fraiman *et al.* [137] demonstrate this point beautifully, showing that by tuning the 2D Ising model towards its critical point and defining a functional connectivity network through thresholding  $R_{ij}^c$ , one can mimic the properties of large-scale complex functional brain networks. Simple, local coupling topologies tuned to the critical point can therefore lead to complex correlation networks. This throws into doubt some of the ubiquitous claims of complex networks in neuroscience [25, 67]; could these complex correlations simply be the result of simple connectivity structures? Inverse methods, such as pairwise maximum entropy modelling might provide a cure to this issue.

As previously noted,  $\mathbf{J}$  can be interpreted as the weighted adjacency matrix of a complex network [50, 58]. In practice, however, all Ising models inferred from finite data will appear fully connected, with values of  $J_{ij}$  for truly disconnected couplings following some distribution about zero. Graph reconstruction from inverse Ising models therefore also requires some secondary post-hoc thresholding step. Regardless, PLM with L1-regularisation<sup>10</sup> [120] and post-hoc thresholding has been shown to perform excellently when reconstructing known networks of a variety of topologies [15, 118]. Both post-hoc threshold  $\delta$  and regularisation strength  $\lambda$  are arbitrary hyper-parameters, chosen at the discretion of the user. This introduces some difficulty when approaching real data, as the best values to choose are unknown. Moreover, while  $\delta$  is easily discernible in say the 2D Ising model, an appropriate value of the threshold is much less clear in cases where the coupling distribution is continuous and centred on zero. An improvement to PLM, so-called pseudo-likelihood decimation [138], performs similarly to optimal choices of  $\delta$  and  $\lambda$  and allows difficult-to-learn topologies to be extracted without the need to tune any hyper-parameters. All in all these results demonstrate that PLM and its extensions are excellent network reconstructors. In this work, we will not use decimation or regularisation as we did not want to impose any additional assumptions of sparsity on the data, and we observe that even without regularisation, coupling networks are much sparser than connectivity networks (Fig. 3.2).

<sup>10</sup> L1 regularisation promotes parameter sparsity and is commonly used when a large number of parameters are known (or assumed) to be zero. The regularisation strength  $\lambda$  can be tuned to produce more or less regularisation.



**Figure 3.2:** We compare the empirical correlations  $R_{ij}$  (top) to the inferred couplings  $\theta_{ij}$  (bottom) for one of the fMRI datasets analysed later. Although correlations and interactions are clearly related, the implied connectivity of the interactions is much more sparse than that of the correlations. Networks constructed by thresholding correlations are therefore likely to over-represent the degree of interconnectivity between brain regions.



## 3.4 Maximum Entropy Models of Real Data

So far, we have introduced what maximum entropy models are and how one might go about inferring their parameters. In this section, we highlight some of the experimental systems where the maximum entropy modelling framework<sup>11</sup> has found the most success. We begin by discussing maximum entropy modelling in neural ensembles, where the method was first popularised. We then discuss how extending the number of allowed spin states from 2 to 21 (inferring so-called Potts models), allows contact networks between protein residues to be established. Extending the spins to continuous states (inferring Heisenberg models) allows us to assess interactions between birds in a flock. We also show how MEMs may be used to analyse financial data, to highlight that these methods have applications beyond biological systems.

<sup>11</sup> As a reminder, pairwise MEMs describing binary data are identical to generalised Ising models, and pairwise maximum entropy modelling is synonymous with inverse Ising inference.

### 3.4.1 Neural Activity (2 Spin States)

Following seminal work showing that pairwise MEMs can account for > 90% of the correlations of a network of retinal neurons [14], the maximum entropy modelling framework has found broad adoption across a range of disciplines. In neuroscience, pairwise MEMs are used to understand the interaction network of ensembles of individual neurons (recorded via neuron-resolution multi-electrode arrays), both in live animals [14, 139–141] and in cell cultures [142]. The systems considered in these studies were small ( $N \sim 10$  to 50) and each neuron is modelled as a binary variable (on/off). For larger population sizes ( $N \sim 100$ ), the authors of Ref. [143] claim that pairwise interactions no longer sufficiently capture the behaviour of the neural network, and infer higher order (triplet and quadruple) interactions to correct this. Ref. [105] notes the same, but addresses the divergence from a pairwise description in an alternative way, instead constraining the global activity of the network; i.e. measuring the probability that  $K$  out of the  $N$  neurons activate at the same time [144]. This encodes the global synchrony of the neural population and is similar to constraining the global magnetisation of the network, with “ $K$ -pairwise” models providing excellent agreement with large neural populations ( $N = 120$ ) [105].

Pairwise maximum entropy modelling has also been used to analyse coarse-grained fMRI datasets [13, 145–147]. Both resting-state [145, 146] and task-specific datasets [147] can be analysed. Moreover, one can establish the energy landscape of the MEM, and different brain states can be linked to minima of this rough landscape [146, 147]. System sizes in fMRI analyses range from  $N \sim 10$  [147] to  $N \sim 250$  [13].

In all the above we inherently assumed that the state of the system is stationary and that the fields and couplings do not vary in time (note that this is a necessary assumption of equilibrium models). This may not be the case in real experiments, and Ref. [148] therefore infer MEMs with stimulus-dependent (i.e. time-varying) magnetic fields. Ref. [149] further extend these ideas and introduce a “state-space” model based on pairwise maximum entropy modelling to infer time-varying fields and couplings. This is computationally taxing, and limited to systems of size  $N \leq 60$ . In our work we consider resting-state state neuroimaging data only, which is as-

sumed to be stationary, and thus will not infer time-varying parameters.

### 3.4.2 Protein-Protein Interactions (21 Spin States)

Proteins are macromolecules constructed by folding linear chains of amino acids, and only achieve their desired functional properties if correctly configured into a specific 3-dimensional structure. Understanding how the amino acid chain is folded to form these 3D structures has been one of the primary pursuits of modern biochemistry [150]. In total there are 20 different types of amino acids, which when linked in the protein sequence are referred to as residues. Pairwise MEM can be used to infer 3D structures (more specifically to infer direct residue contacts) by analysing pairwise correlations between residues along the amino acid sequence [151–154]. Each residue can assume 21 states, one for each type of amino acid and one for gaps in alignment, and as such the pairwise MEM describing this dataset is a generalised Potts model [155]<sup>12</sup>. The fields  $h_i$  now describe the preference for a certain residue at the  $i^{\text{th}}$  sequence position, and the couplings  $J_{ij}$  indicate if two types of amino acid are in contact or not. Ref. [156] provides a physics-friendly introduction to the topic, as well as showing that the PLM method may be extended to infer multi-state Potts models.

<sup>12</sup> Potts models are just generalisations of Ising models to many discrete states.

### 3.4.3 Flocks of Birds (Continuous Spin States)

Flocks of birds, specifically starlings, display correlations which are scale-free [157]. By treating the (normalised) velocity of each bird as a Heisenberg spin (i.e. a spin which can point in any direction on a unit sphere) Bialek *et al* [158] show that pairwise MEMs are sufficient to account for these correlations. This description is developed further in [159], and it is shown that the maximum entropy model describing the flock is tuned towards a critical point, hence explaining the previously observed scale-free correlations.

### 3.4.4 Financial Markets (2 Spin States)

Financial signals (i.e. prices) can conceptually be reduced to binary variables by considering if the price is rising (bullish) or falling (bearish). Hence, a number of authors have used pairwise MEMs to analyse market connectivity, from identifying relations between companies operating in different sectors [160] to understanding consumer purchasing behaviour [161]. Moreover, stock markets exhibit many phenomena associated to critical systems, and pairwise MEMs have allowed these to be assessed in the context of equilibrium statistical physics [162–164].

### 3.4.5 Summary

Maximum entropy modelling is used when the correlations of a complex system can be measured but microscopic interactions are difficult (or impossible) to access. The most common use of pairwise MEMs is to reconstruct interaction networks, and this can be done for binary, multi-state or continuous variables. Many complex systems, from flocks of birds to collections of neurons to stock markets, display signatures of criticality<sup>13</sup>. MEMs also provide an intuitive way to investigate and validate any such claims

<sup>13</sup> Indeed we partly chose to highlight the above systems as evidence supporting criticality has been observed in each of them.

of criticality, as the inferred models correspond exactly to equilibrium statistical physics models of the data. Our work will use pairwise maximum entropy modelling for both of these purposes; we aim to understand if the resting state of the brain is critical, and to use the inferred couplings to better characterise the functional connectivity of the brain.

## Summary of Chapter 3

- We introduced the *maximum entropy modelling* framework. Maximum entropy models are statistical models which maximise the entropy of a dataset and their samples follow the Boltzmann distribution.

### **Pairwise Maximum Entropy Models**

Measuring up-to-pairwise correlations within a dataset corresponds to selecting a maximum entropy model of up-to-pairwise interactions. This *pairwise maximum entropy model* is parameterised by external fields  $\mathbf{h}$  and pairwise couplings  $\mathbf{J}$ .

### **Pairwise Maximum Entropy Modelling**

Defines the process of inferring the parameters of a pairwise maximum entropy model from real data via a statistical inference, e.g. through maximising some likelihood function. This process is called *inverse Ising inference* when binary datasets are considered, leading to the inference of generalised Ising-like models.

### **Pseudo-likelihood Maximisation**

*Pseudo-likelihood maximisation* (PLM) is an approximate, data-driven method with which to perform inverse Ising inference. PLM is based on logistic regression, and we will exclusively use this method to perform our inferences as it has been shown to outperform alternative inverse Ising solvers.

- The criticality of pairwise maximum entropy models can be easily assessed within the framework of statistical physics as the inferred models represent equilibrium statistical physics models of the data.
- Pairwise maximum entropy modelling is often used as a tool for network reconstruction. We discussed that pairwise maximum entropy modelling may provide less dense and more truthful estimates of the network topology than correlation-based methods used in neuroimaging.
- Maximum entropy modelling can be applied to a variety of complex systems; from ensembles of neurons to financial markets. Although we focus on neuroimaging data, any methodological conclusions we draw from our investigation will therefore have further-reaching consequences for a variety of systems.



# Chapter 4

## Criticality in Neuroscience

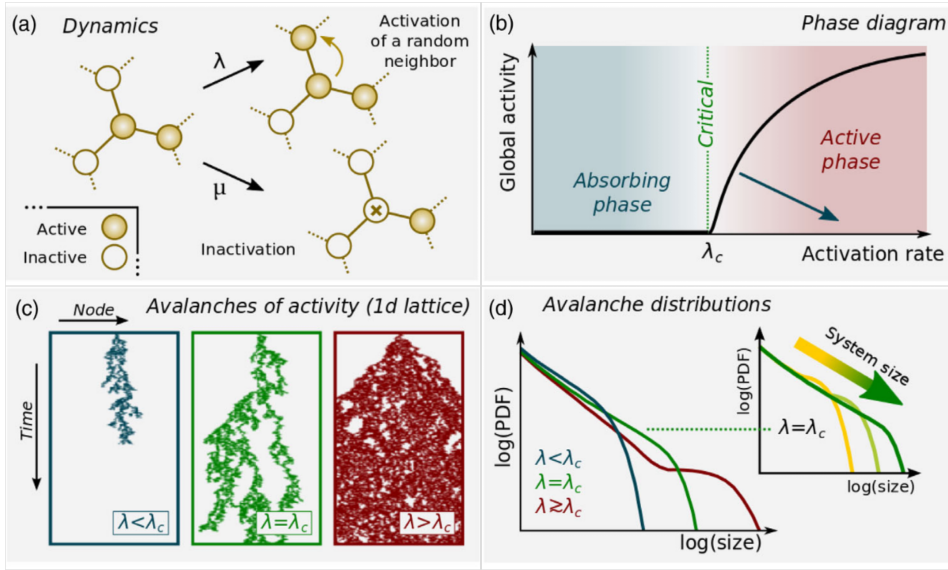
A number of complex systems appear to display phenomena indicative of criticality. In this chapter, we try to understand what might cause criticality to be so ubiquitous in nature, taking a specific focus on living complex systems. We detail evidence supporting criticality in neuroscience, as this is the primary focus of this work, followed by a general discussion of criticality in maximum entropy models. We find that the performance of maximum entropy models strongly depends on the state-point of the input data, which forms the initial strand of our research. We conclude this chapter by providing a contextualised preview of our research output.

### 4.1 Criticality as an Organising Principle

A broad range of experimental evidence seems to suggest that many living systems are tuned towards (or exist in the proximity of) a critical point [2, 3, 6, 7]. The apparent ubiquity of close-to-critical behaviour has led some to conjecture that evolutionary pressures may be responsible for fine-tuning biological systems towards the critical point, and that criticality serves as a template from which complex behaviours arise across a range of length scales [7]. Indeed a number of interesting properties are associated with the critical state, including providing optimal dynamic range and sensitivity to inputs [8], enabling long-range coordination between individual elements [157, 165], allowing a large range of dynamic responses [9, 10] and maximising computational ability through edge-of-chaos computation [166, 167], making such an organising principle theoretically appealing. In the context of neural systems this school of thought has become known as the *critical brain hypothesis*, which further states that the brain derives numerous computational advantages from operating in a near-critical regime [168, 169] such as maximising information transmission and storage, and aiding classification and non-linear input filtering [3, 13, 170–176].

#### 4.1.1 Fine Tuning

In chapter 2 we showed that the order-disorder transition of the 2D Ising model is controlled by a single parameter  $T$  and that the critical point represented a very special value, occurring at only at a single specific temperature  $T_c$ . By analogy, for the brain to truly be a critical system, it must there-



**Figure 4.1:** Overview of the contact process. Panel (a) describes the dynamical rules by which nodes activate; panel (b) shows the resulting non-equilibrium phase diagram; panel (c) shows the time evolution in the sub-critical  $\lambda < \lambda_c$  (ordered) phase, at the critical point  $\lambda = \lambda_c$  and in the super-critical  $\lambda > \lambda_c$  (disordered) phase; panel (d) shows the resulting avalanche distributions in each part of the phase diagram. Taken from Ref. [7].

fore also operate in some (infinitely) small region of a control parameter space. In other words, under the critical brain hypothesis, the brain must be fine-tuned *exactly* to the parameter values of the critical point; any deviation from these parameters and the phenomenology (and advantageous properties) of the critical state are lost. Such fine-tuning problems [177] are well known in fundamental physics<sup>14</sup> and pose philosophical problems if no reasonable explaining mechanism can be put forward. We now discuss two commonly proposed solutions in the context of neuroscience.

#### 4.1.2 Dynamic Systems and Self-Organised Criticality

The brain is an inherently dynamic system [66]. As such, neural activity is commonly modelled as a dynamic non-equilibrium process, of which a simple example is the contact process [179] - see Fig. 4.1. In the contact process, nodes  $s_i$  of a network may be active or inactive (i.e. have a binary state). Activity propagates through the network dynamically. At each time step active sites deactivate, simultaneously causing random adjacent (connected) inactive sites to activate at a rate  $\lambda$  (Fig. 4.1(a)). The system can be made to undergo an order-disorder (absorbing-active) transition by tuning  $\lambda$ , with a dynamical critical point occurring at  $\lambda_c = 1$ . We sketch out this phase diagram in Fig. 4.1(b). For  $\lambda < \lambda_c$  the system is in an absorbing phase (analogous to the ordered ferromagnetic phase) in which any initial activity decays exponentially to a globally deactivated steady state. When  $\lambda > \lambda_c$  the system is in an active phase (c.f. disordered paramagnetic phase) where any initial activity explodes to a fluctuating steady state where the density of active nodes is  $\rho = 1 - (1/\lambda)$ . At the critical point  $\lambda = \lambda_c$ , initial bursts of activity form characteristic “avalanches” (see Fig. 4.1(c)) the size and dynamics of which obey power-law distributions Fig. 4.1(d).

Real systems still need to be fine-tuned to  $\lambda_c$  to observe the dynamic critical point described above. Per Bak *et. al* [180] introduced a generic feedback mechanism through which a system can dynamically self-organise towards criticality. In such self-organized criticality (SOC) the critical point acts as an attractor which, regardless of whether initialised in an absorbing

<sup>14</sup> For instance in cosmology the initial density of matter and energy in the universe must be tuned to a critical value with an error  $\mathcal{O}(10^{-60})$  to account for the currently observed flatness of the universe [178].

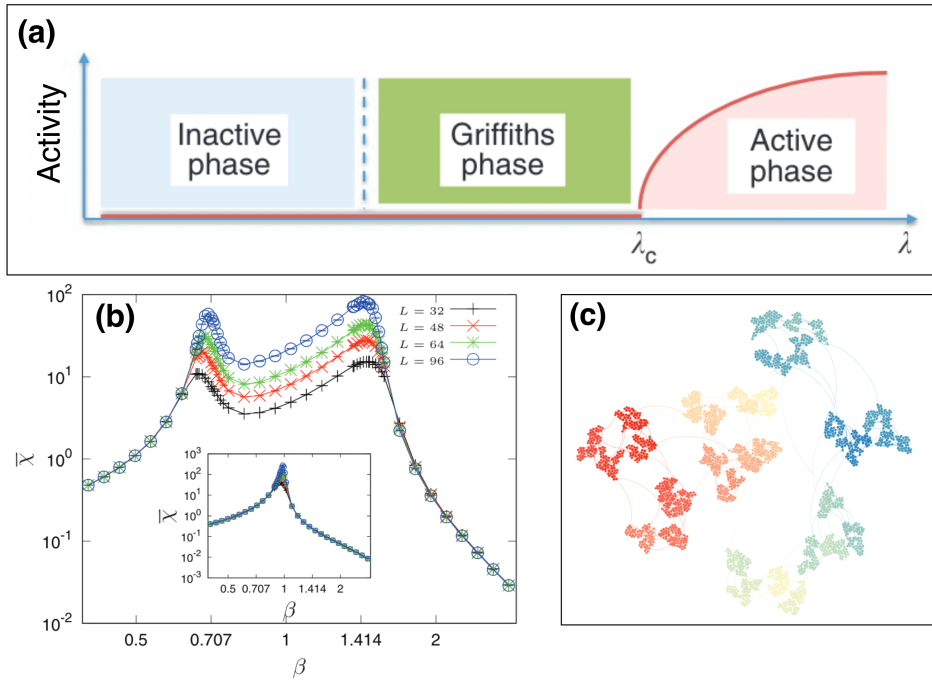
or active state, the system will evolve towards as time progresses. A varied range of physical systems, from piles of rice [181], to earthquakes [182] to neural networks [171], show scale-free power-law dynamics consistent with SOC. SOC thus seems an appealing generic organising principle through which dynamic systems are tuned towards the critical state. Note that SOC can only be used to explain fine-tuning in dynamic non-equilibrium systems. We will now detail an alternative solution to the fine-tuning problem which is also applicable in equilibrium settings.

### 4.1.3 Extended Critical Regimes

The fine-tuning problem arises because the critical point is located at a single uniquely defined set of parameters. What if instead there was an extended critical regime, throughout which scale-free and power-law scaling was observed? Luckily such a thermodynamic phase exists and is commonly termed the Griffiths phase [183, 184]. The Griffiths phase (similarly to the spin-glass phase of chapter 2) arises when quenched disorder is introduced into the Hamiltonian of the system, for instance by randomly diluting (i.e. removing) couplings in the nearest-neighbour Ising model [185]. Introducing this disorder causes local clusters to form in the bulk material, some of which will be highly inter-connected and weakly coupled to the bulk. When approaching the clean critical temperature  $T_c$  of the disordered system from above the strongly coupled clusters will organise into local ferromagnetic clusters, even though the global phase of the system remains paramagnetic. These local clusters (often called rare regions) display slow relaxation dynamics [186] and lead to a singularity in the free energy [183]. As these ordered clusters exist over an extended temperature range the corresponding Griffiths singularities also occupy an extended temperature regime, separating pure ordered and disordered phases. The same can also occur for dynamic non-equilibrium models, and we show a sketch of the associated modified phase diagram in Fig. 4.2(a).

If the disorder is uncorrelated then the effects of the Griffiths singularity are weak [185]. But introducing quenched disorder that is correlated (e.g. by introducing linear or planar defects into the lattice Ising model) can greatly enhance the effects of the rare regions. Ref. [187] for instance shows, using MC simulations, that introducing correlated disorder into the lattice Ising model replaces the usual susceptibility cusp observed at  $T_c$  with an extended  $T$  range in which the susceptibility is large, and bounded by two susceptibility peaks. We reproduce this result in Fig 4.2(b).

Note that the local clusters which lead to the Griffiths phase inherently require structured topologies with local couplings. For instance, even though there is quenched disorder, we would not expect to see any cluster formation in the fully connected SK model as the number of neighbours tends to infinity. We would therefore also not expect to observe a Griffiths phase in this model. Interestingly, simulations of the contact process on small-world networks have shown that disorder introduced by the heterogeneity of the coupling topology alone may be sufficient to generate Griffiths phases [189]. This idea is expanded further in [188] where simulated neural dynamics on hierarchical modular networks which mimic real brain networks (see Fig 4.2(c)) were shown to display power-law avalanches for a



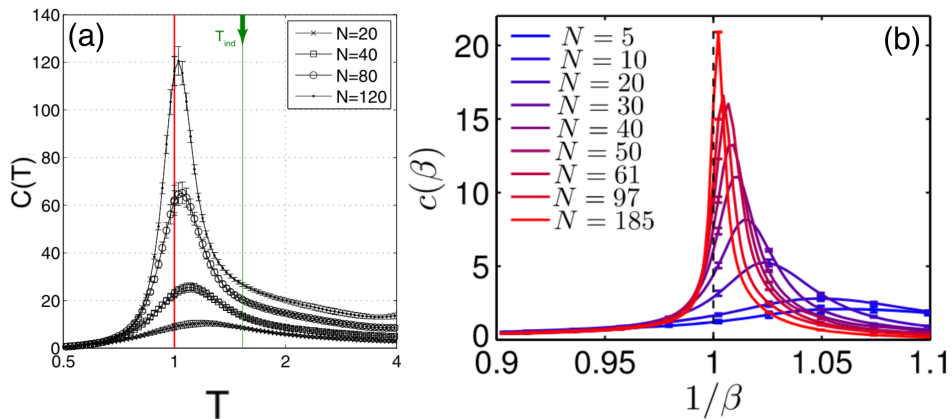
**Figure 4.2:** The Griffiths phase arises when there is disorder in the connectivity of a model. (a) Sketch of a phase diagram for a dynamic model which includes a Griffiths phase, where  $\lambda$  is a control parameter as in the contact process. (b) Shows the susceptibility  $\chi$  of an Ising model with correlated disorder as a function of the inverse temperature  $\beta$ . Inset shows susceptibility when there is uncorrelated disorder. The model with correlated disorder shows an extended critical regime, corresponding to a Griffiths phase. (c) Graphical representation of a hierarchical and modular complex network that mimics cortical brain networks. Panels (a) and (b) are taken from Ref. [188], and panel (c) is from Ref. [187].

broad range of activity parameters  $\lambda$  (i.e. a Griffiths phase). This extended critical regime is accompanied by a heightened dynamic range (analogous to the susceptibility). Modular network topology is thus a sufficient condition for the formation of the rare regions required to observe the Griffiths phase. Interestingly, evidence from fMRI studies suggests that the resting-state brain wanders around a broad regime near the critical point [165], supporting the idea of a neural Griffiths phase.

## 4.2 Evidence of Criticality in Neuroscience

Neuroscience is an expansive field, and a vast number of different experimental modalities have been developed to understand the brain across a hierarchy of length-scales [28]. Evidence supporting different types of phase transitions has been observed using many of these experimental techniques [7], and the topic is too vast to exhaustively discuss here. As the purpose of this thesis is to understand and model the brain as an equilibrium system (i.e. by construing a maximum entropy representation of the data), we will focus our discussion on evidence supporting that the brain exists near a *thermodynamic* phase transition. Although all data collected in neuroscience is inherently dynamic (i.e. follows some time-dependent update rule), maximum entropy models (MEMs) assume that each observation simply represents a sample drawn from some underlying probability distribution. We will therefore refer to this type of criticality as statistical criticality. For completeness, we also provide a surface-level overview of some of the most well-studied evidence for dynamic criticality in neuroscience.

**Figure 4.3:** Evidence for criticality in maximum entropy models of neural ensembles. Both panels show the specific heat capacity  $C$  as a function of a simulated temperature. The “natural” model inferred from the data occurs at  $T = 1$ . Both papers sub-sample their system to perform a finite size analysis, with each line corresponding to a different system size  $N$ . Panel (a) is from Ref. [190], and panel (b) is from Ref. [191].



#### 4.2.1 Statistical Criticality of Maximum Entropy Models

Following pioneering work on the construction of maximum entropy models from populations of retinal neurons [14], a number of authors have applied the maximum entropy modelling framework to neural systems. This framework discards the neural dynamics, and simply considers each observation of the system as corresponding to a sample drawn from the (equilibrium) Boltzmann distribution. Interestingly, MEMs inferred from neural populations consistently appear to sit near an order-disorder transition [6, 144, 190, 191]. Refs. [190, 191] for example show this by simulating their inferred models at a range of fictive temperatures  $T$ , which we reproduce in Fig. 4.3. When  $T = 1$  the inferred model is in its natural state, i.e. corresponds to the MEM that was extracted from the data. By tracking the specific heat, a second-order derivative of the free energy, as a function of  $T$  and performing finite size scaling they find that the MEMs inferred from neural populations sit very close to the systems critical point, i.e. that  $T_c = 1$ . While these are positive results with regard to the critical brain hypothesis, we do note that these studies were carried out for relatively small systems, with  $N \leq 185$ . From the literature we surveyed, there appears to be no direct evidence of an extended critical regime (i.e. Griffiths phase) within MEMs inferred from neural populations; for instance, the finite size extrapolations shown in Fig. 4.3 both converge towards a singular critical temperature.

Previous studies have also shown that simple Ising models tuned to their critical point can replicate some of the complex dynamics of coarse-grained experimental whole-brain datasets [137, 176, 192]. This suggests that large-scale brain dynamics could be the result of a critical equilibrium process. All these studies were comparative; they simply noted that critical Ising models displayed some of the correlations as the brain, and made no use of the actual data when building their models. Maximum entropy modelling allows this analysis to be extended, and in a recent study, MEMs built from fMRI data were found to exist near the spin-glass (SG) - paramagnetic (P) phase transition of an SK-like model [13]. There is therefore evidence supporting neural criticality from across a range of length scales, from ensembles of neurons to coarse-grained fMRI data.

## 4.2.2 Dynamical Criticality and Neural Avalanches

Perhaps the most well-studied example of neural criticality is that of “neuronal avalanches”. Beggs and Plenz [171] observed that collections of neurons (both in culture and in slices of rat cortex) activate in highly correlated ways forming neural “avalanches”. More specifically, they found that the size distribution of these avalanches follows a power-law with exponent  $3/2$  (c.f. Fig. 4.1(c) and (d)) and that this exponent agrees well with exponents observed for a mean-field critical branching process [193]. While observing power-laws does not strictly imply criticality (see [194] for a discussion), the neuronal avalanches appear to obey finite size scaling [174] and can be collapsed by a universal scaling function [195], both of which are strong indicators of a true critical point. Moreover, avalanche dynamics have been observed across a range of length scales, from neuron-resolved studies [196] to fMRI datasets [165] supporting the idea that this process may be scale-invariant. Interestingly, the critical state of the mean-field branching process may be achieved via SOC [197], and more realistic biological models of neurons have also been shown to tune to criticality via SOC [198, 199]. Neuronal avalanches may therefore be an example of a process self-organised to criticality.

## 4.3 Maximum Entropy Models and Proximity to Criticality

As noted in Section 4.2.1, MEMs inferred both from populations of neurons [6, 144, 190, 191] and coarse-grained whole-brain studies [13] appear to sit close to a thermodynamic (statistical) critical point. But these findings are in fact much more ubiquitous; MEMs of diverse datasets, from protein sequence banks [200] to flocks of birds [159] to natural images [201], all appear close to criticality. This prevalence could be interpreted as evidence supporting criticality as a generic organising principle for complex systems. But it could also be a result of some systematic bias of the maximum entropy inference, which tunes the inferred parameters towards their critical values. This question forms the foundation of our first two results chapters, and we now detail previous investigations which aimed to understand how well maximum entropy models performed on simulated data.

### 4.3.1 Inference Errors in Simulated Systems

A number of authors have investigated the performance of pairwise maximum entropy modelling on simulated binary data (i.e. data from generalised Ising models). Numerous network topologies, coupling disorders and system sizes have been investigated, and to name but a few, previous authors looked at 2D lattice Ising models with ferromagnetic couplings ( $N = 49$ ) [202], dilute and fully connected SK models ( $N = 16$  to 128) [15], ferromagnetic models on random graphs with a fixed degree, square lattices and fully connected SK models ( $N = 64$ ) [16], varied topologies for small system sizes ( $N = 16$ ) [118], and recently ferromagnetic couplings (i.e.  $J_{ij} = J$  when connected) on Erdős–Rényi networks ( $N = 50$  to 400) [203]. Regardless of



the precise topology or coupling distribution of these models, they all display an order-disorder transition at a critical point, which can be accessed by varying a parameter like the temperature  $T$  (or equivalently the inverse temperature  $\beta$ ). We showed an example of one such investigation in the previous chapter, see Fig. 3.1. Universally, all the above studies find that reconstruction errors are smallest in the vicinity of the critical point, i.e. when generating models at  $T \approx T_c$ .

These results are commonly attributed to the theoretical finding that the generalised susceptibility of the Ising model can be considered as a Fisher information matrix [204], and that the maximisation of these quantities at the critical point corresponds to a maximal density of distinguishable models. One should consider the claims of minimum errors at  $T_c$  in the above list of papers with a degree of scepticism as the system sizes probed (bar Ref. [203]) were all small and likely to be strongly affected by finite size effects. As a case study, let us consider the zero-field SK model for  $N = 64$  investigated in Ref. [16] and shown in Fig. 3.1. The authors find a minimal error at  $T = 1$ , which for  $N \rightarrow \infty$  does indeed correspond to the critical temperature of the SG-P transition. However, careful finite-size scaling studies of the SK model [205] show a peak in the specific heat (which is an empirical way of determining the location of the finite size critical temperature) at  $T_c^{\text{fs}} \approx 0.7$ . The region of minimum inference error reported in Ref. [16] therefore occurs entirely within the paramagnetic phase of their finite-size SK model. Simulations thus reveal that parameters from datasets close to but offset from the critical point are easiest to learn. But they also highlight another point; inference errors are highly dependent on the true state point (i.e. temperature) that generated the data.

## 4.4 A Taste of Things to Come

This concludes the introductory chapters of this thesis. As mentioned previously, the primary scientific goal of this work is to infer maximum entropy models of human brain data so that we may assess whether the mind is positioned close to a thermodynamic critical point. This is motivated both by theoretical work suggesting that the brain may derive a number of computational advantages from operating near criticality, and also the breadth of previous experimental evidence supporting neural criticality. We will construct *pairwise* maximum entropy models (MEMs) of resting-state functional magnetic resonance imaging (fMRI) data, which we assume to be stationary (in a dynamical sense), allowing us to model the resting mind as an equilibrium system. Inferring a pairwise MEM is equivalent to solving the inverse Ising problem, and so we infer Ising-like models with complex network structures that are encoded within the couplings between spins. A number of methods exist to solve the inverse Ising problem, and comparative studies by previous authors have found that pseudo-likelihood maximisation (PLM) provides the best overall performance. This will therefore be our method of choice and the only method we investigate.

In this chapter, we have shown that inference errors strongly depend on the state point, e.g. the temperature, and the topology of the true generative model from which the data was drawn. It is generally impossible *a*

*priori* to know the state point of any real-world dataset<sup>15</sup>. To confidently assert that any models we infer from real datasets are close to criticality we must first understand how PLM performs on data generated from known models. We thus perform a detailed analysis of the PLM method on the zero-field SK phase diagram for system sizes relevant to fMRI datasets in chapter 5. This will reveal that, while PLM is exact for infinite data, parameter estimates are strongly biased when the sample size is finite, leading to a tendency to overestimate the criticality of data drawn from paramagnetic (i.e. fluctuating) state points. Any criticality reported for MEMs inferred from small datasets may therefore be an artefact of the inference scheme, throwing into doubt some of the ubiquitous claims of near-critical maximum entropy models surveyed in section 4.3. We explain these findings by analysing how the inference error depends on sample size. This reveals that estimates of control parameters like the temperature depend linearly on the inverse sample size, similar to the first-order bias of generic maximum likelihood estimators. We also present a number of ways to correct this bias in chapter 6 and show that the PLM temperature estimate of a dataset generally provides a lower bound estimate of the true temperature.

In chapter 7, we explore the effect of the PLM bias on a small fMRI dataset on mindfulness meditation. This case study shows that not accounting for the small sample bias can lead us to mischaracterise the phase of data and to false claims of criticality. In chapter 8, we apply PLM to one of the largest openly available neuroimaging datasets from the human connectome project (HCP). We find that this dataset is large enough to disregard small sample biases and that the overall human resting state corresponds to a super-critical state point with enhanced critical fluctuations. The coupling structure of the HCP network is complex, and we find that the coupling strengths obey a truncated power law. We analyse these models and their structure and propose a self-consistent thresholding scheme with which we extract a simplified model of the brain that conserves the correlations of the input data. We further show that small couplings play a vital role in mediating the fluctuations of the brain, and conclude that representative network reconstructions must somehow account for these negative edges.

<sup>15</sup> Otherwise we would not have to perform the inference in the first place...



## Summary of Chapter 4

- Evidence from a range of complex biological systems supports the hypothesis that these operate at or near a phase transition. In the context of neuroscience, this is known as the *critical brain hypothesis*.
- Complex systems must be *fine-tuned* towards the critical point to achieve criticality. Two proposed solutions to this fine-tuning problem are:

### **Self-Organised Criticality**

Feedback mechanisms can drive dynamic systems to self-organise towards criticality. So-called neuronal avalanches agree closely with the critical exponents of self-organised criticality.

### **Extended Critical Regimes**

Disorder in the coupling structure of a network can lead to extended critical regions where observables follow power-law scaling. These regions are called *Griffiths phases*. Neural dynamics simulated on complex hierarchical and modular network topologies have been shown to form Griffiths phases, supporting the idea of a neural Griffiths phase.

- We surveyed some of the most important results supporting neural criticality. These can be split depending on whether one considers the data as a dynamic process or as statistical samples from a probability distribution.

### **Dynamical Criticality**

Refers to evidence for dynamic phase transitions, such as neural synchronisation and neuronal avalanches.

### **Statistical Criticality**

Refers to evidence for criticality in statistical models. Maximum entropy models inferred from neural ensembles appear to support statistical criticality.

- Simulations have shown that inference errors in maximum entropy modelling depend on the state point of the input data. Inference errors are the smallest in the paramagnetic phase close to but offset from the phase transitions. This means that some state points are harder to learn (i.e. require more data) than others.

## Chapter 5

# Pseudo-likelihood Maximisation in the Sherrington-Kirkpatrick Model

In the preceding chapter, we established that inference errors in inverse Ising inference strongly depend on the state point of the data. As we do not know the state point of any given neurological dataset *a priori* we wanted to understand how pseudo-likelihood maximisation (PLM) would perform on a model with a well-understood phase diagram. Similarly to previous authors, we chose the fully connected Sherrington Kirkpatrick (SK) model as our benchmark model (see chapter 2 for an overview of the SK model its and connection to neural systems). We will only investigate the PLM solution to the inverse Ising problem, as numerous previous authors have established that this provides the best parameter estimates (see chapter 3). While PLM is exact in the limit of infinite samples  $B$ , we will focus on the case where  $B$  is small and finite to establish any systematic errors that could arise in relation to this. This will allow us to understand the types of errors one might encounter in real datasets, which are always limited in size. We will generate our datasets using Monte Carlo (MC) simulations, the details of which are given below. We will find that small sample biases cause PLM to overestimate the variance of the parameter distribution, biasing models towards the critical point.

### 5.1 Problem Setup

In this chapter, we will generate data from across the SK phase diagram. We denote the input (i.e. the true) parameter set  $\theta^0 = \{h^0, J^0\}$  with the superscript 0. The maximum likelihood parameters estimated through PLM will be indicated with the superscript \*, and we will commonly refer to  $\theta^* = \{h^*, J^*\}$  as the inferred model or the inferred parameters. We perform un-regularised PLM inference on each trajectory, as the models are not sparse. This section will discuss how we generate our input models and datasets.

### 5.1.1 Model Generation

We will benchmark the PLM method on a zero-field ( $\mathbf{h}^0 = 0$ ) fully connected SK model [35] using system sizes comparable to typical coarse-grained fMRI brain region analyses [13, 23, 133],  $N = 50$  to 800. We generate our input couplings  $\mathbf{J}^0$  by drawing from a Gaussian distribution with mean  $\mu_J = \mu/TN$  and standard deviation  $\sigma_J = \sigma/TN^{1/2}$ ;

$$J_{ij} \sim \mathcal{N}(\mu_J, \sigma_J^2). \quad (5.1)$$

Note that we have absorbed the temperature into our definition of the model parameters. We do this as for real applications the “temperature” (in a statistical physics sense) of the system is undefined, and only coupling strengths can be extracted. Strongly coupled regimes correspond to low temperatures and vice versa.  $\mu$  and  $\sigma$  are intensive variables and the state of the system is characterised by the dimensionless average coupling strength  $\mu/\sigma$  and temperature  $T/\sigma$ . We fix  $\sigma = 1$  and sample from the range  $\mu \in [0, 2]$ ,  $T \in [0.5, 2]$ , where in the  $N \rightarrow \infty$  thermodynamic limit, the system explores all of its phases. These phases are a low-temperature disordered spin-glass (SG), low-temperature ordered ferromagnetic (F) and high-temperature disordered paramagnetic (P) phase; see chapter 2 for more detail. When referring to the “state point” of the system we explicitly mean a specific unique  $(\mu, T)$  pair. Generally, we will generate and simulate multiple models for each state point, with each set of couplings referred to as a *realisation* of the model.

### 5.1.2 Data Generation

For every  $(\mu, T)$  we produce input datasets via standard Metropolis-Hastings Monte Carlo (MC) sampling [97, 98]. This commonly employed sampling technique is discussed at great length elsewhere, and we point the reader to [206, 207] for extensive overviews. Briefly, MC sampling simulates a Markov chain process where the probability of updating the state (i.e. configuration)  $\mathbf{s}(t)$  at a time  $t$  to a new configuration  $\mathbf{s}(t+1)$  at time  $t+1$  depends on only on the current state of the system,  $\mathbf{s}(t)$ , and the probability distribution one wishes to sample. Here we wish to draw samples from the equilibrium (Boltzmann) distribution. As before the probability of observing a configuration in equilibrium depends on the energy associated with that configuration via

$$P(\mathbf{s}) = \frac{1}{Z} e^{-\mathcal{H}[\mathbf{s}]/T_f}, \quad (5.2)$$

where we have introduced a “fictive temperature”  $T_f$  which for the time being is set  $T_f = 1$ , but will come to play a role in the later chapters of this thesis.  $T_f$  is *not* the same as the temperature of the model  $T$ , which as stated above, has been absorbed into the definition of the parameters  $\mathbf{h}$  and  $\mathbf{J}$ . We define the energy difference between the new configuration  $\mathbf{s}(t+1)$  and the old configuration  $\mathbf{s}(t)$  as  $\Delta\mathcal{H} = \mathcal{H}[\mathbf{s}(t+1)] - \mathcal{H}[\mathbf{s}(t)]$ , and our MC algorithm updates the configuration with a probability dependent on whether or not the new configuration lowers the energy of the system. New configurations  $\mathbf{s}(t+1)$  are proposed via a simple and standard local update

rule; we choose a spin  $s_i$  at random with uniform probability  $1/N$  and flip the state of that spin,  $s_i(t+1) = -s_i(t)$ . The probability of accepting this proposal can then be shown to be

$$P(\mathbf{s}(t) \rightarrow \mathbf{s}(t+1)) = \begin{cases} 1, & \text{if } \Delta\mathcal{H} < 0, \\ e^{-\Delta\mathcal{H}/T_f}, & \text{if } \Delta\mathcal{H} \geq 0. \end{cases} \quad (5.3)$$

Note that because we perform local (i.e. small) changes to the configuration at each time step the Markov chain will be highly correlated. Moreover, the chain needs to be initialised with some configuration  $\mathbf{s}(t=0)$  which may or may not be close to the equilibrium state. As such two rules of thumb must be followed to obtain representative samples from MC simulations. Firstly, the chain must be given time to “equilibrate” or “burn-in”, so that it may evolve to a region of phase space representative of the state point being simulated. Samples generated from this period should not be used to estimate the thermodynamic properties of the system. Secondly, samples should be drawn at sufficiently large time intervals, so that the correlation between them is small and so that they may be considered independent and identically distributed (i.i.d). There are thus 3 sampling meta-parameters which need to be tuned when sampling from our MC algorithm: the equilibration time  $t_{eq}$ , the production time  $t_{pr}$  and the sampling interval  $t_s$  which sets the time intervals at which we collect samples from the chain during the production time. We will report all times in units of Monte Carlo sweeps, where one sweep represents  $N$  spin update attempts. These sampling time meta-parameters will vary throughout the results presented below, and in order to report the MC meta-parameters in a concise way we will quote them as  $p_{mc} = \{t_{eq}, t_{pr}, t_s\}$ . For example, we will quote a MC simulation which equilibrated for  $10^5$  sweeps, and from which  $B = 10^3$  samples were generated at intervals of 1 per 10 sweeps as  $p_{mc} = \{10^5, 10^4, 10\}$ .

The full dataset of  $B \times N$  observations obtained from the production period will either be referred to as a *trajectory* or as the *dataset*. Thermodynamic averages are then estimated from this trajectory as simple arithmetic means

$$\langle x \rangle \approx \frac{1}{B} \sum_{t=1}^{t=B} x(\mathbf{s}(t)), \quad (5.4)$$

where  $t$  indexes time in the dataset. We commonly run multiple independent MC chains for each state point over which we average to refine these estimates. Unless stated otherwise, each MC chain was initialised with a configuration of randomly oriented spins.

### 5.1.3 Relevant Observables

From each trajectory, we compute the following quantities to characterise the SK phase diagram and assess the performance of the PLM method. Starting from the microscopic spin averages

$$m_i = \langle s_i \rangle, \quad (5.5)$$

and correlations

$$C_{ij} = \langle s_i s_j \rangle - \langle s_i \rangle \langle s_j \rangle, \quad (5.6)$$

we compute the spin-glass order parameter (the *overlap*) as

$$q = \frac{1}{N} \sum_{i=1}^N \langle s_i \rangle^2, \quad (5.7)$$

and the global (squared) covariance as

$$C^2 = \frac{1}{N} \sum_{i,j=1}^N C_{ij}^2. \quad (5.8)$$

$C^2$  is related to the spin-glass susceptibility via  $\chi_{SG} = C^2/T^2$  [13, 89]. When approaching the SG phase from the higher temperature P phase, the susceptibility (and hence the covariance) increases rapidly, reflecting the development of spontaneous large correlations near the phase transition, i.e. the critical point. We track and measure  $C^2$  instead of  $\chi_{SG}$  as this quantity can be defined independently of  $T$ , and is therefore also measurable in real datasets where  $T$  is unknown.  $q$  and  $C^2$  will be used to characterise the phase boundaries of the finite size SK phase diagram. We further define the auto-correlation  $C_t$  within the sample as

$$C_t(\Delta t) = \frac{1}{N} \sum_{i=1}^N \langle s_i(t) s_i(t + \Delta t) \rangle, \quad (5.9)$$

where  $\Delta t$  is the time delay. We define the auto-correlation time  $\tau$  as the delay when  $C_t(\Delta t = \tau)/C_t(\Delta t = 0) = 1/e$ , and monitor  $\tau$  to ensure that subsequent MC samples are decorrelated and can be considered as i.i.d samples. Turning now to the inference process, we define the inference error  $\varepsilon$  according to

$$\varepsilon = \sqrt{\frac{\sum_{i \leq j} (\theta_{ij}^* - \theta_{ij}^0)^2}{\sum_{i \leq j} (\theta_{ij}^0)^2}}, \quad (5.10)$$

which is a robust aggregate measure of the deviations in parameter estimation previously defined in [16]. For clarity, we repeat that  $\theta$  is the symmetric matrix containing all PLM parameters, with  $\theta_{ii} = h_i$  and  $\theta_{ij} = J_{ij}$  as all  $J_{ii} = 0$  (there are no self-couplings). Note that the error is dominated by contributions from the couplings as the number of couplings,  $N_J = N(N-1)/2$ , is much larger than the number of fields,  $N_h = N$ .

Each generated model realisation will deviate by a small amount from  $T$ , and so we define the *measured* temperature of each input model realisation as  $T^0 = 1/(\sigma_J^0 N^{1/2})$ , where  $\sigma_J^0$  is the standard deviation computed from the realised couplings. We similarly define the measured inferred temperature as  $T^* = 1/(\sigma_J^* N^{1/2})$  where  $\sigma_J^*$  is the standard deviation of the corresponding inferred couplings. This allows us to define a second global metric on the inference quality, i.e. how well the inferred model reproduces the temperature. For brevity, we will refer to our total workflow (generating input models, simulating data from them and then inferring a PLM reconstruction of that data) as a generate-simulate-inference (GSI) run.

## 5.2 PLM for Example State Points

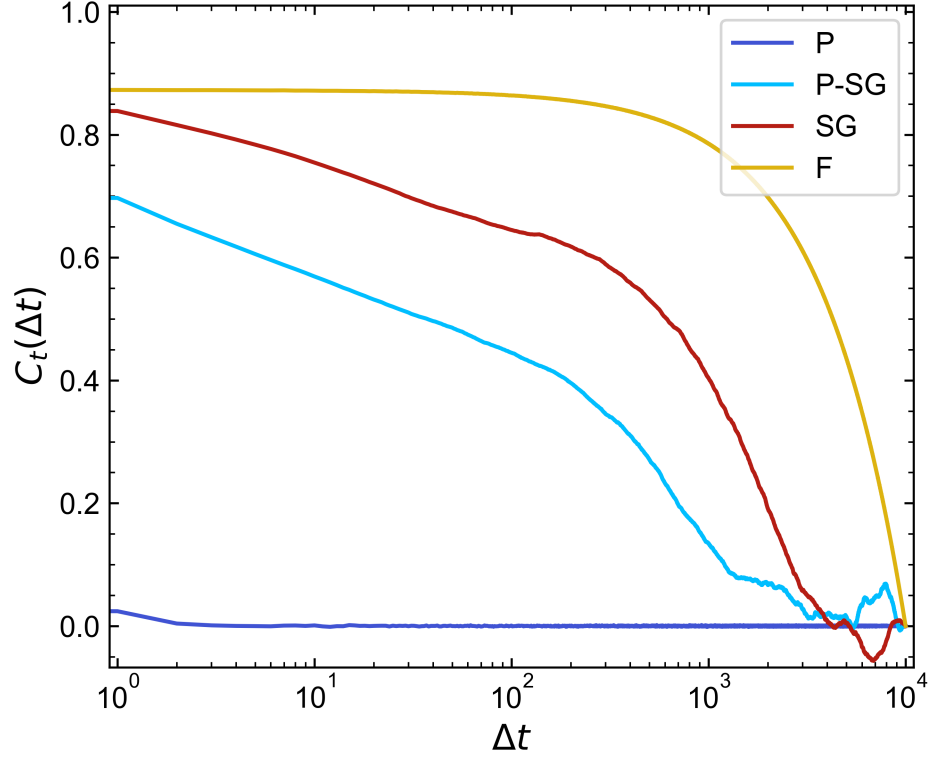
We ultimately want to understand the performance of PLM across the entire SK phase diagram. But when doing so we will necessarily have to omit some details of the inference output, and report only summary statistics such as the error  $\varepsilon$ , which give no insight into which parameters  $\theta_{ij}$  specifically are responsible for increases in the error. We thus use this section to explore the parameter-specific (i.e. microscopic) origin of the inference error in each phase. This will also allow us to establish familiarity with the matrix and distribution representations of our model parameters. We perform a single GSI run for 4 example state points from different parts of the zero-field SK phase diagram, and investigate differences between the input parameter matrix  $\theta^0$  and the inferred parameter matrix  $\theta^*$  for each. One model was generated for each of the main phases in the SK phase diagram with the labels F, P and SG denoting the ferromagnetic, paramagnetic and spin-glass phases respectively. We also generated a model from near the P-SG phase transition to demonstrate how the inference performs in the vicinity of a critical point. We present results for a small system of size  $N = 120$  so that full visualisations of the parameter matrices remain digestible. For each simulation our MC meta-parameters were  $p_{mc} = \{10^5, 10^5, 10^1\}$ . We now additionally simulate 6 independent MC chains for each model, appending the generated data together to form a final dataset consisting of  $B = 6 \times 10^4$  samples.

Phase	$\mu$	$\mu^0$	$\mu^*$	$T$	$T^0$	$T^*$	$\mu_h^*$	Fig.
P	0.80	0.8	0.81	1.50	1.50	1.49	0.00	5.2
P-SG	0.30	0.39	0.35	0.50	0.50	0.48	0.00	5.3
SG	0.00	0.07	0.03	0.35	0.35	0.31	0.00	5.4
F	2.00	1.85	0.54	0.60	0.60	0.07	-0.14	5.5

Table 5.1: Summary statistics of PLM models inferred from data generated at 4 zero-field SK model state points. Theoretical quantities have no superscripts. Empirically measured quantities from the input parameters and inferred parameters are indicated by superscripts 0 and \* respectively. Each state point is explored in more detail in the corresponding figure.

Table 5.1 summarises the ability of PLM to capture the 4 selected state points (i.e. to correctly estimate the intensive properties  $\mu$  and  $T$  and  $h$ ). From each input model realisation we estimate  $\mu$  using the relation  $\mu^0 = \mu_J^0 / (\sigma_J^0 N^{1/2})$ , and similarly each inferred PLM model using  $\mu^* = \mu_J^* / (\sigma_J^* N^{1/2})$ . The quantities  $\mu_J^0$  and  $\mu_J^*$  are the measured empirical means of the input and inferred couplings respectively. The relations with which we calculate empirical temperatures  $T^0$  and  $T^*$  are defined in section 5.1.3. We also measure the spin averaged field in the inferred model  $\mu_h^*$ , and expect  $\mu_h^* \approx 0$  for a perfect model reconstruction as we are analysing the zero-field SK model. We find that there is generally some discrepancy between the theoretical mean coupling strength  $\mu$  and the realised mean coupling strength  $\mu^0$  of the input model. The temperatures  $T$  and  $T^0$  are in much closer agreement. PLM is best at reconstructing the P state point and becomes progressively worse at capturing  $\mu$  and  $T$  as  $T$  decreases, from P

**Figure 5.1:** Semi-log plot of the auto-correlation  $C_t(\Delta t)$  vs the delay time  $\Delta t$  for the 4 example state points described in Table 5.1. Corresponding decay times are:  $\tau_P = 1$ ,  $\tau_{P-SG} = 164$ ,  $\tau_{SG} = 1333$ ,  $\tau_F = 5786$ . The slowing dynamics means that datasets of size  $B$  collected from e.g. the SG and F phases contain less information than datasets of similar  $B$  in the ergodic P phase.



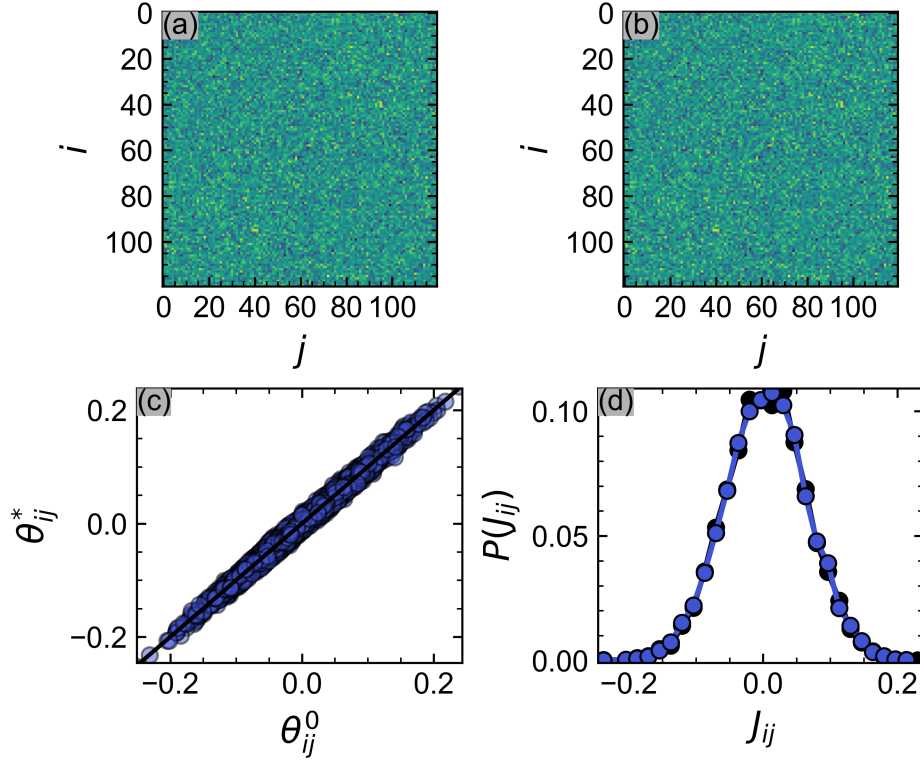
to P-SG to SG state points. In the region  $\mu < 1$  the phase is not sensitive to  $\mu$ , i.e. all  $\mu < 1$  have a SG ground state. The temperature is therefore the only relevant control parameter in this region. PLM systematically underestimates  $T$  for all  $\mu < 1$  (P, P-SG, SG) example state points. This corresponds to over-estimating the spread of the distribution, a finding that we explore at length later in this chapter. In contrast, for the ferromagnetic ( $\mu > 1$ ) state point the inferred model does not correspond to the input model at all;  $T^* \approx 0$  and a negative average external field is found.

The simplest explanation for these differences is that the auto-correlation time at each state point is different. We plot  $C_t$  in Fig. 5.1 on a semi-log plot to investigate this and find that indeed, as  $T$  decreases  $\tau$  increases substantially. In the F case in particular,  $B/\tau_F \approx 2$ , i.e. only two independent samples are contained within the dataset. There is thus not enough information in the F dataset for PLM to infer the state point. Fig. 5.1 also reveals that both P-SG and SG state points still contain dynamically fluctuating data, although the temporal correlations within these datasets are much longer. This would explain why the PLM estimates ( $\mu^*, T^*$ ) for the P-SG and SG state points are further from the input values than the P state point.

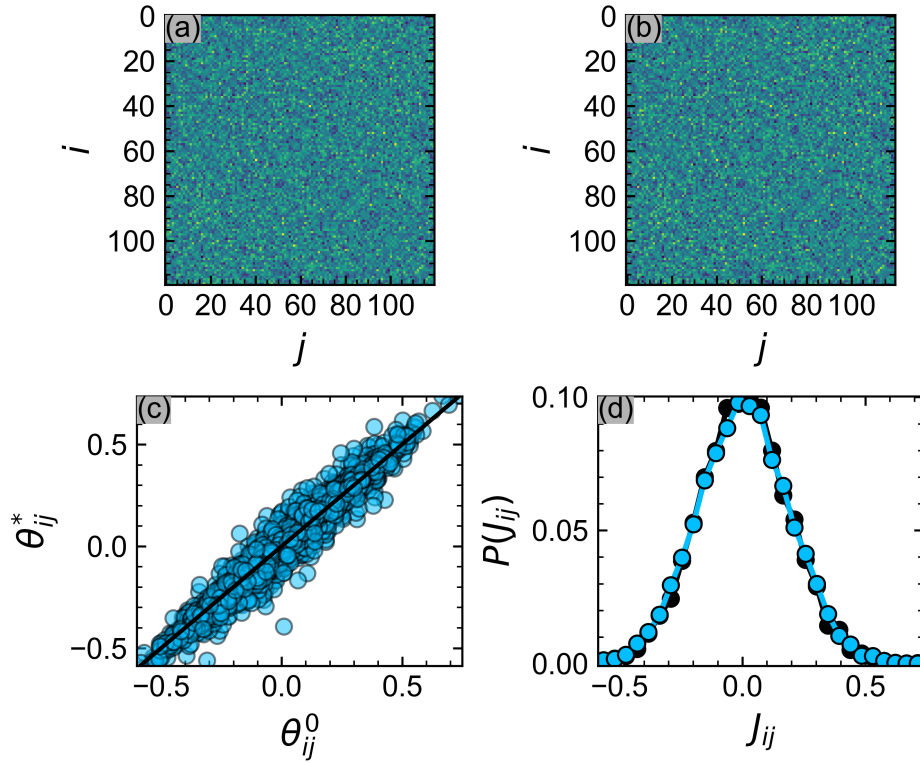
### 5.2.1 Single Realisation Investigation

To better understand how individual parameters contribute to the inference error  $\varepsilon$  we provide overviews of the difference between input and inferred models for P, P-SG, SG and F state points in Figs 5.2, 5.3, 5.4 and 5.5 respectively.





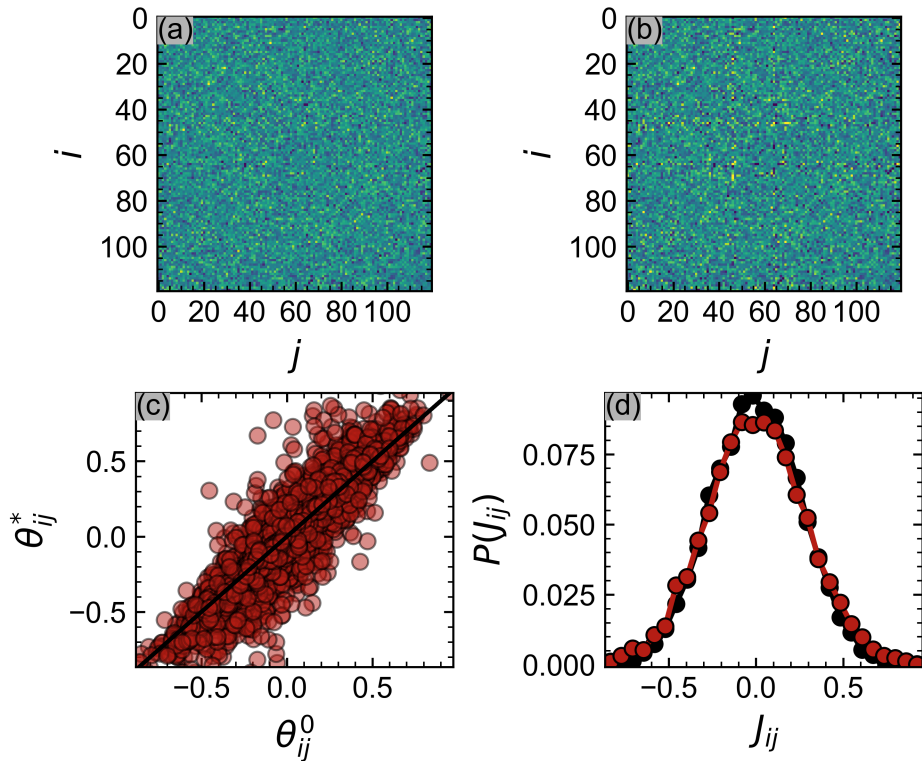
**Figure 5.2:** Overview of the performance of PLM on data from a single realisation of a paramagnetic state point, generated from  $(\mu, T) = (0.8, 1.5)$ . We show the input parameter matrix  $\theta^0$  in (a), the inferred parameter matrix  $\theta^*$  in (b), the parameter correlation in (c) and coupling distribution for input (black) and inferred (dark blue) parameters in (d). The colours shown in (a) and (b) range from  $\min(\theta^0)$  (dark blue) to  $\max(\theta^0)$  (light yellow). The number of spins is  $N = 120$ .



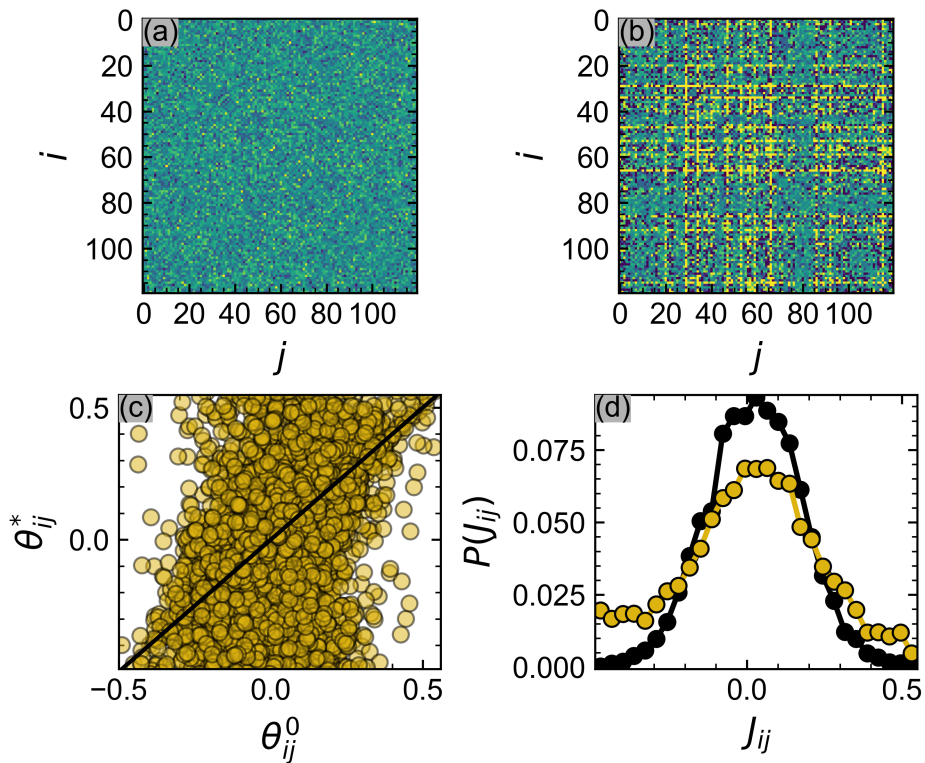
**Figure 5.3:** Overview of the performance of PLM on data from a single realisation of a state point near the paramagnetic spin-glass transition, generated from  $(\mu, T) = (0.3, 0.5)$ . We show the input parameter matrix  $\theta^0$  in (a), the inferred parameter matrix  $\theta^*$  in (b), the parameter correlation in (c) and coupling distribution for input (black) and inferred (light blue) parameters in (d). The colours shown in (a) and (b) range from  $\min(\theta^0)$  (dark blue) to  $\max(\theta^0)$  (light yellow). The number of spins is  $N = 120$ .



**Figure 5.4:** Overview of the performance of PLM on data from a single realisation of a state point in the spin-glass regime, generated from  $(\mu, T) = (0.0, 0.35)$ . We show the input parameter matrix  $\theta^0$  in (a), the inferred parameter matrix  $\theta^*$  in (b), the parameter correlation in (c) and coupling distribution for input (black) and inferred (red) parameters in (d). The colours shown in (a) and (b) range from  $\min(\theta^0)$  (dark blue) to  $\max(\theta^0)$  (light yellow). The number of spins is  $N = 120$ .



**Figure 5.5:** Overview of the performance of PLM on data from a single realisation of a ferromagnetic state point, generated from  $(\mu, T) = (2.0, 0.6)$ . We show the input parameter matrix  $\theta^0$  in (a), the inferred parameter matrix  $\theta^*$  in (b), the parameter correlation in (c) and coupling distribution for input (black) and inferred (orange) parameters in (d). The colours shown in (a) and (b) range from  $\min(\theta^0)$  (dark blue) to  $\max(\theta^0)$  (light yellow). The number of spins is  $N = 120$ .



In these figures panel (a) shows the input parameter matrix  $\theta^0$ , panel (b) the inferred parameter matrix  $\theta^*$ , panel (c) the correlation between input and inferred parameters and panel (d) the distribution of both input and output parameters. Comparing panels (a) and (b) visually shows that for the  $\mu < 1$  (i.e. P, P-SG, and SG state points) there is a good overall agreement between the parameter matrices (i.e. the structure of the input model is captured), and input and inferred models cannot be visually distinguished. The inferred F model however, displays a completely different behaviour; whole rows (equivalently columns) of parameters contain large values that are orders of magnitude greater than input the maximum coupling found in the input model ( $\mathbf{J}_{\max}^0 \approx 0.5$  vs  $\mathbf{J}_{\max}^* \approx 20$ ).

The correlation plots in panels (c) communicate the parameter errors in more detail. We see that for the P example, there is an (almost perfect) linear correspondence between  $\theta_{ij}^0$  and  $\theta_{ij}^*$ . As we decrease the temperature to the P-SG and SG state points we find increasing deviations from this linear relationship, and that the  $\theta_{ij}^*$  point cloud occupies a larger area of the plot. This corresponds to an overall larger inference error. So while the PLM still conserves the overall structure or the ranking of the couplings for these state points, the values of the parameters themselves become more error-prone. For the SG state point in particular we observe that  $\theta_{ij}^*$  point cloud appears to be rotated about the line  $\theta_{ij}^* = \theta_{ij}^0$ , i.e. negative values of  $\theta_{ij}^0$  become more negative, while positive values of  $\theta_{ij}^0$  become more positive. This corresponds to a “stretching” of the parameter distribution, and through the relation  $T^* = 1/(\sigma_j^* N^{1/2})$ , a lower temperature state point. PLM thus underestimates the temperature of data from these state points. We study this effect in further detail later in this chapter and will find that PLM always provides a biased estimate for the temperature that is systematically lower than the true temperature. The F correlation plot shows extreme behaviour, with there being little correlation between input and output couplings (i.e. PLM has failed to reconstruct the model). We set the limits of the correlation plots to the maximum and minimum values of  $\theta_{ij}^0$ . In the F case, however, there are many extremal points outside these bounds, and Fig. 5.6 shows the full correlation plot. Panels (d) further show that the inferred coupling distributions for P, P-SG and SG state points are representative of the input model: they capture the gaussianity of the input model. The inferred F model however is non-Gaussian, with long tails and large couplings, corresponding to temperatures at zero temperature.

### 5.2.2 Error Distribution and Row Dependence

We further compare the distributions of the errors in  $\mathbf{h}$  and  $\mathbf{J}$  for each example state-point in Fig. 5.7. We see the same behaviour as in panels Fig. 5.2-5.5(c): as the temperature decreases and the dynamics slow the size of the errors increases. Errors for the fields are generally small for P, P-SG and SG state points, while large field values are inferred for the F state point. Intuitively this makes sense; one way of generating a ferromagnetic state is to have a strong uniform field acting on all of the spins. Turning now to the couplings  $\mathbf{J}$ , we note that the tails of the errors become longer and larger as the coupling strength of the model increases from P to P-SG to SG to F. Although we can clearly see that the overall error for the

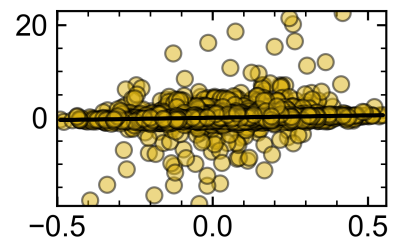
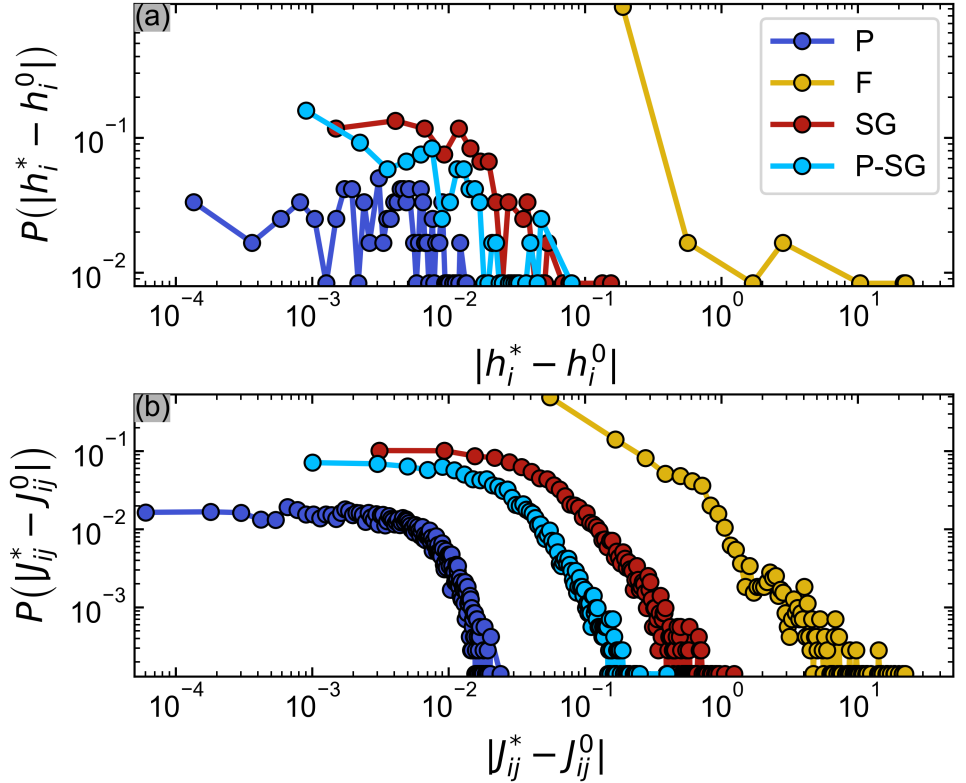


Figure 5.6: Expanded view of  $\theta^0$  vs  $\theta^*$  for the ferromagnetic inference errors shown in Fig 5.5(c). Highlights divergence of specific inferred parameters to (infinitely) large values.

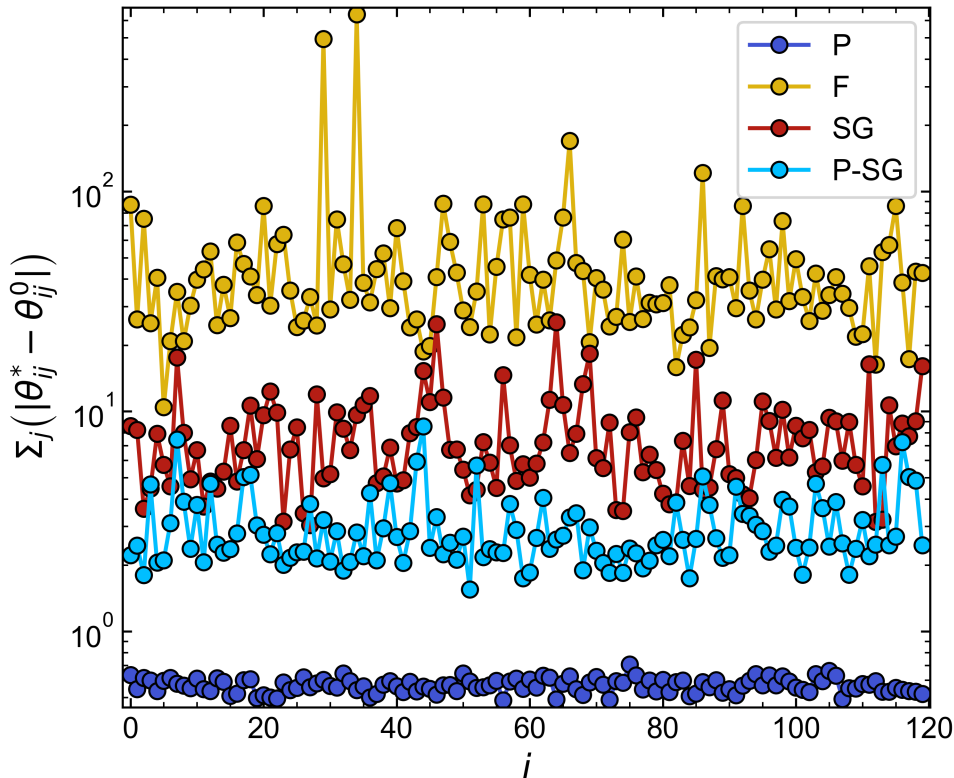
**Figure 5.7:** Probability distributions of the inference errors for the fields (a) and couplings (b) for our 4 example SK state points. We observe that the average error magnitude increases as the temperature decreases and that the tail of the error distribution becomes longer as the temperature decreases. We attribute this to the slowing down of the dynamics the system freezes on approach to the SG and F phases.



F example will be larger than the others, the distribution representation obscures the fact that certain parameter rows in particular contribute most to the error. The distribution is therefore good at representing the gradual increase in error when inference is possible, but not at identifying if the type of catastrophic failure seen in Fig. 5.5 has occurred.

We therefore also plot the row error (i.e. the sum of all errors in the  $i^{\text{th}}$  row of the parameter matrix) as a function of  $i$  in Fig. 5.8. This clearly highlights the two discussed behaviours. There is an overall average “baseline” error for each state point which captures the difficulty of learning that state point. As the coupling strength increases this baseline shifts up, and the variation of the row error increases. For the F case with the strongest coupling, we find that certain rows have errors orders of magnitude above the baseline, and note that for these PLM has completely failed and inferred arbitrarily large parameter values. For now, we simply state that these divergent parameters are the result of a process called *separation* [208].

In summary, we have used this section to demonstrate how PLM performs for 4 different example state points representative of the different regions of the SK phase diagram. This has had two purposes: a) to establish familiarity with how to represent and compare parameter inputs and outputs (parameter matrices, correlation plots, distributions etc.) and b) to investigate how individual parameters contribute to the overall inference error  $\varepsilon$ . We identify two separate sources of error when performing PLM on finite data. Generally, PLM has a tendency to overestimate the size of individual parameters  $\theta_{ij}$ , causing the inferred coupling distribution to appear more spread than the true coupling distribution. This effect worsens (the baseline error rises) as the coupling strength of the input model increases.



**Figure 5.8:** The row error (i.e. the sum of all errors in the  $i^{\text{th}}$  row of the parameter matrix) as a function of the row index  $i$ . We observe two trends. The baseline error increases as the temperature decreases, this reflects a smooth contribution to the error as a result of a decrease in sampling quality. Upon approaching the static F phase, we find specific rows with errors an order of magnitude larger than the baseline. These reflect a failure of PLM to identify the underlying model structure, c.f. Fig 5.5(a) and (b), with arbitrarily large parameters inferred for a subset of rows.

In the strongly coupled regime (as in the F example), specific parameter rows appear to diverge to values that are orders of magnitude larger than the maximal values expected from the input couplings  $\theta^0$ , representing a failure of PLM to infer the couplings entirely. These rows dominate the contribution to  $\varepsilon$ .

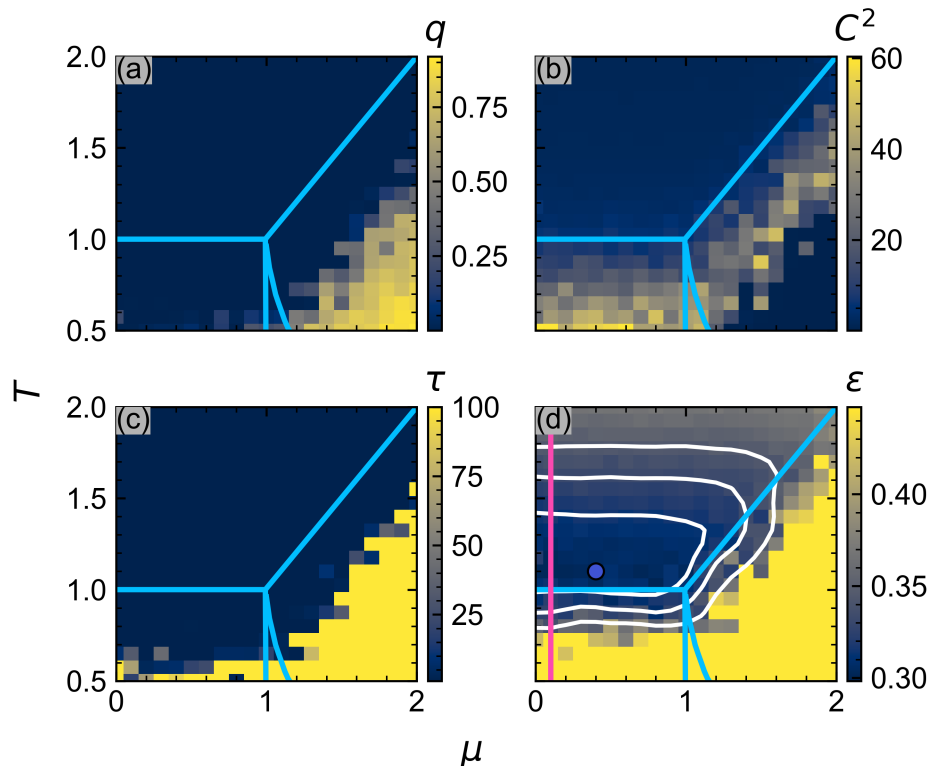
### 5.3 Overview of Inference Errors in the SK Model

From previous findings, we expect the inference to perform best near the critical point [15, 16, 118], which for finite-sized systems is surrounded by an extended regime of enhanced critical fluctuations. This region of phase space is however notoriously hard to sample due to a phenomenon termed critical slowing down [209], where the auto-correlation time  $\tau$  (similarly to other thermodynamic properties) diverges in the vicinity of the critical point. To mitigate this, we chose a long sampling frequency  $t_s = 1000$  when performing our initial survey of the phase diagram. We fix the MC meta-parameters to  $p_{mc} = \{10^6, 10^7, 10^3\}$  and collect  $B = 10^4$  samples for all state points. We begin by showing results for a single fixed  $N = 200$ , as we found that (bar finite size effects shifting and sharpening the transitions) the results of our analysis were similar for all system sizes.

#### 5.3.1 Error Phase Diagram

In Figure 5.9 we illustrate the overall phase behaviour of the SK model and of the inference error for  $N = 200$  and  $B = 10\,000$ , with the phase boundaries of the  $N = \infty$  system overlaid. Each pixel represents an average

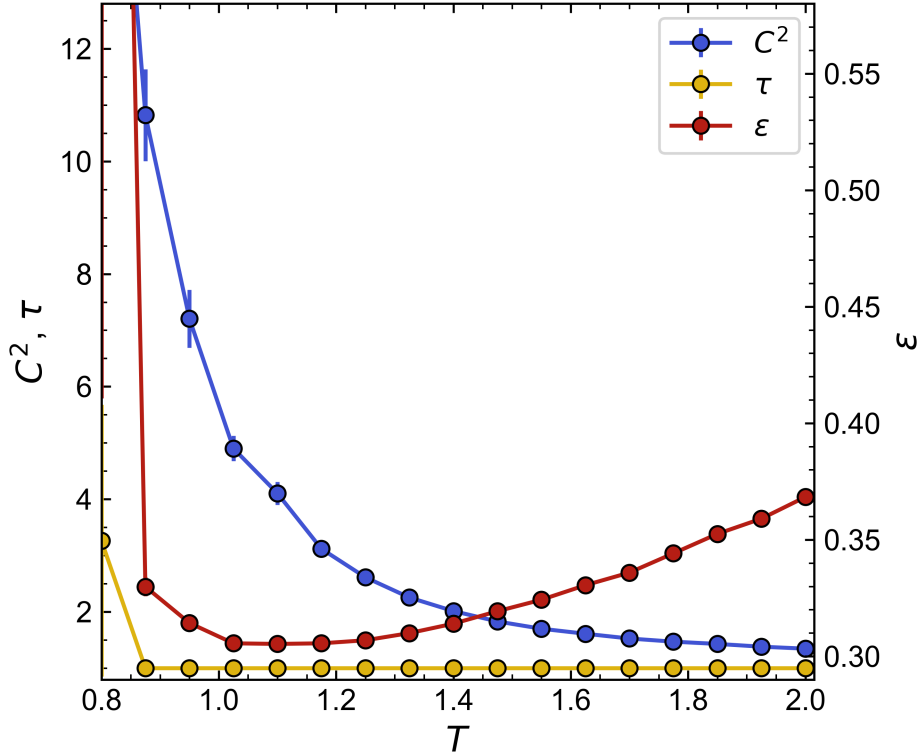
**Figure 5.9:** Overview of the order parameter (a), susceptibility (b), auto-correlation time (c) and error (d) for the  $N = 200$  zero-field SK phase diagram. The blue lines show phase boundaries between the paramagnetic (P), ferromagnetic (F) and spin-glass (SG) phases in the  $N \rightarrow \infty$  limit. We have taken these from Ref. [19]. In panel (d), white lines show contours of  $\varepsilon$  and the pink line labels the line  $\mu = 0.1$  across which a more detailed examination of the error is made in Fig. 5.10. The location of the minimum error is denoted by a dark blue circle.  $\varepsilon$  is thresholded so that  $\varepsilon_{\max} = 1.5\varepsilon_{\min}$ . Each pixel presents an average over 3 GSI runs.



of 3 input model realisations and GSI runs. In (a) we recover the known SK phase diagram, with low values for the overlap in the paramagnetic (P) and spin-glass (SG) phase compared to the ferromagnetic (F) phase. The phase transitions correspond to regimes of increased susceptibility, which peak at the phase boundaries but are blurred and shifted due to finite size effects [42] as shown in (b). Deep in the F and P phases  $C^2$  is low due to the lack of fluctuations in the F phase, and a lack of spin-spin correlations in the P phase. Panel (c) shows that  $\tau = 1$  everywhere besides deep in the F and low-temperature SG phases, reflecting that our sampling frequency  $t_s = 10^3$  is sufficiently large to draw independent and identically distributed (i.i.d) samples from most of the phase diagram. The F and SG phases are static (i.e. do not fluctuate) and therefore will always have large values of  $\tau$ . From the discussion regarding optimal samples and inference in the previous chapter, we intuitively expect PLM to perform poorly in these regions as the drawn samples contain limited information about  $P(\mathbf{s})$ .

Fig. 5.9(d) shows the performance of the PLM method as quantified by the error in (5.10). We observe a minimum in  $\varepsilon$  in the paramagnetic phase ( $\mu = 0.4, T = 1.1$ ), and a rapid increase as the two correlated F and SG phases are approached. As discussed, this is consistent with previous studies of the fully-connected SK model for  $N = 64$  [15, 16], who found the error to be minimized around  $T \sim 1$  for  $\mu = 0$ . Although the error minimum is close to the peak of the critical fluctuations, the two are not coincident - in our simulations we find e.g. the P-SG  $C^2$  peak at  $T \approx 0.6$ , while finite size studies of the SK model have shown the critical fluctuations of the specific heat of the P-SG transition to peak at  $T \approx 0.7$  for similar system sizes [205]. The  $\varepsilon$  colour-map is capped at  $1.5\varepsilon_{\min}$  but much larger errors,  $\mathcal{O}(10^3 - 10^4) \times \varepsilon_{\min}$ , are observed in the correlated F and SG phases.





**Figure 5.10:** Correlation measure  $C^2$ , auto-correlation time  $\tau$  and error  $\varepsilon$  as functions of  $T$  for fixed  $\mu = 0.1$ . Error bars are the standard errors of the observables over 21 GSI repeats. Values  $T < 0.8$  are not plotted as  $\tau$  began to substantially deviate from 1 on the approach to the SG regime.  $p_{mc} = \{10^6, 10^7, 10^3\}$ .

For now, we attribute these large errors to the  $B$  configurations becoming highly correlated ( $\tau \gg 1$ ).

### 5.3.2 Error Dependence on Temperature

The region of minimal error in Fig. 5.9(d) is characterised by iso-contours that run parallel to the phase transition lines, implying a strong dependence of the error on the *distance* from the phase boundary. This is especially clear at low  $\mu$ , where the  $\varepsilon$ -contours lie roughly along lines of constant  $T$ . To understand the relationship between the inference error and the emergence of criticality, we also study a profile of the phase diagram at fixed  $\mu = 0.1$ , away from the ferromagnetic phase. This profile characterises the SG-P transition and is a function of the control parameter  $T$  only. This choice of  $\mu$  was also partially motivated by a recent PLM analysis of fMRI data which showed the human mind to sit in the vicinity of the SG-P transition [13]. In Fig. 5.10 we compare the inference error and  $C^2$  as a function of  $T$  for this  $\mu = 0.1$  profile. Note that each point here corresponds to averages obtained by carrying out GSI runs for 21 independent input model realisations for each state point. This confirms that for  $N = 200$ ,  $\varepsilon$  has a flat minimum, centred around a temperate  $T_{\min} \approx 1.1$ . The shape of the minimum is asymmetric, with  $\varepsilon$  diverging slowly as  $T$  goes from  $T_{\min} \rightarrow \infty$  and rapidly as  $T_{\min} \rightarrow 0$ . At the minimum, the fluctuations  $C^2(T_{\min})$  are 3-4 times larger than their high-temperature limit. But as  $T$  decreases further, the fluctuations continue to increase while the error rapidly diverges. The divergence occurs without a significant increase in the auto-correlation time, indicating that it is not due to poor sampling. The minimum error is therefore within the regime of enhanced critical fluctuations and occurs

close to but offset from the phase transition. At the supposed finite size critical temperature the inference fails due to the inherently highly correlated nature of the data from this regime ( $\tau$  is large).

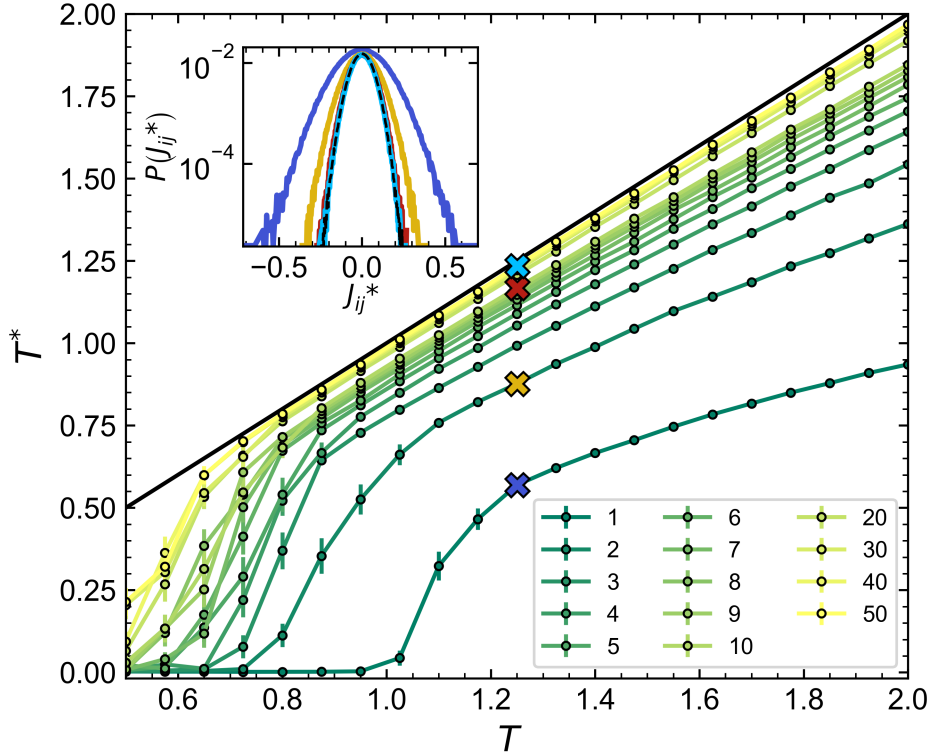
Summarising; the PLM error on the SK phase diagram is encapsulated by a linear dependence on the proximity to the phase transitions. While we have only analysed the SG-P transition, similar results for F-P transitions have previously been shown in the literature [118, 202]. We find the error to be minimal when the data is super-critical, and near to but offset from the phase transition. These results agree with previous numeric investigations [15, 16], but contradict theoretical findings [204, 210], which would expect the error to be minimized *exactly* at the critical point.

## 5.4 PLM Performance with Limited Data

In the previous sections, we established how the inference error  $\varepsilon$  depends on the state point  $(\mu, T)$  of the SK model for a fixed number of samples  $B$ . Like previous authors, we found that the state-point of the data has a large impact on the ability of PLM to reconstruct the model (see Fig. 5.10), and that the smallest reconstruction errors are observed in the P phase near the P-SG transition. An alternative way to interpret  $\varepsilon$  is to say that state points with large errors would require more samples to obtain the same  $\varepsilon$ . We investigate the dependence of the PLM reconstruction on  $B$  in this section as we do not know what phase our real-world datasets belong to. We return to a system size  $N = 200$ , and again generate data along the line of constant  $\mu = 0.1$  highlighted in Fig. 5.9. This allows us to study PLM across the P-SG phase boundary. For every  $T$  we generate 21 independent model realisations, from which we run MC simulations with  $p_{mc} = \{10^6, t_{pr}, 10^3\}$ , where  $t_{pr}$  is set to generate datasets containing sample sizes of  $B = \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 30, 40, 50\} \times 10^3$ . Plotted quantities then represent averages over these 21 GSI runs, with error bars being the standard error over the same 21 runs.

### 5.4.1 Inferring Temperature from Finite Sized Datasets

The inference error  $\varepsilon$  is the combination of the error on each individual parameter. It is dominated by the couplings  $J_{ij}$ , as there are many more couplings ( $N_J = N(N - 1)/2$ ) than fields ( $N_h = N$ ). In section 5.2 we briefly noted that P-SG and SG state points showed inferred coupling distributions that were more “spread” than the input coupling distributions. These effects were small in section 5.2 as we performed the analysis in a high sample number setting,  $B = 6 \times 10^4$  for  $N = 120$ , i.e. near the limiting case of  $B \rightarrow \infty$  where we know PLM to be an exact reconstruction method. For small (finite)  $B$  these effects become much more apparent, which we show by directly inspecting the probability distribution of  $\mathbf{J}$  for a selection of sample sizes  $B$  in the inset of Fig. 5.11. Specifically, we observe that the inferred distributions remain symmetric and appear approximately Gaussian, but that they systematically over-estimate the variance, gradually converging to that of the input distribution with increasing  $B$ . The origin of this spread can be explained in terms of the bias of maximum



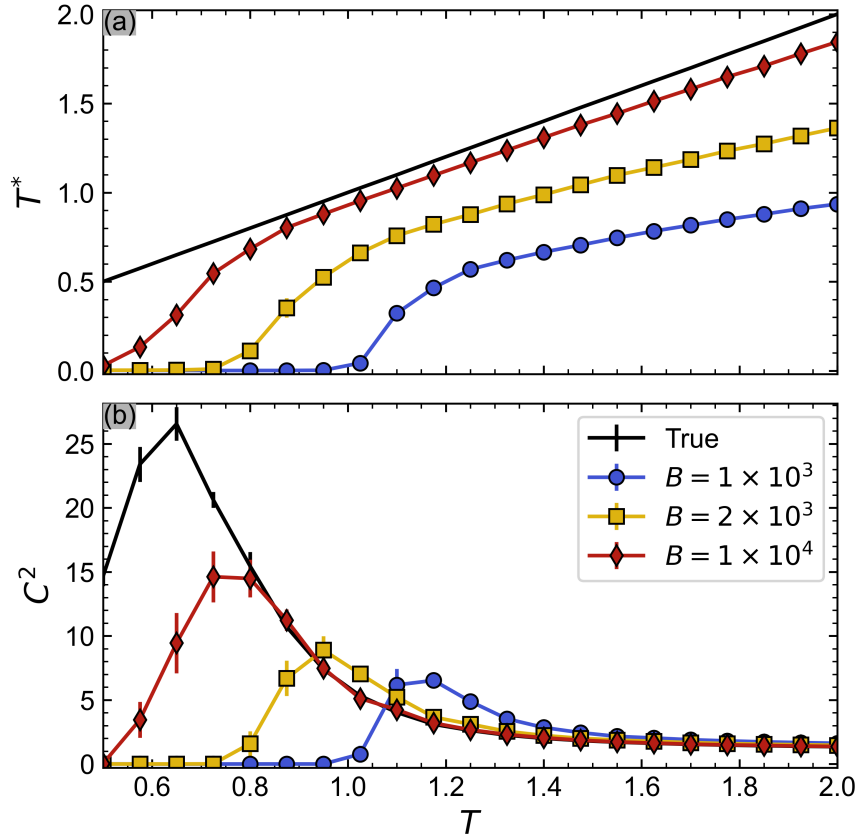
**Figure 5.11:** The main figure shows the inferred temperature  $T^*$  as a function of the input temperature  $T$  for a range of sample sizes  $B$ . The darkest green line shows results for the smallest  $B$  and the brightest yellow line corresponds to the largest  $B$ . Coloured crosses indicate temperatures corresponding to the coupling distributions in the inset.  $N = 200$ , and points and error bars are the mean and standard error over 21 GSI runs for each  $T$ . The legend shows the number of samples for each curve in units of  $10^3$ . In the inset, we show probability distributions of the inferred couplings for different  $B$  from the same state point ( $\mu = 0.1, T = 1.25$ ) near the minimum  $\varepsilon$  in Fig. 5.10. Dark-blue, orange, red and light-blue lines correspond to  $B = \{1, 2, 10, 50\} \times 10^3$  respectively. The black dashed line shows the ground truth distribution for reference.

likelihood estimators, as parameter estimates obtained from the logistic regressions for each parameter row in PLM are known to be biased away from 0, i.e. over-estimated [37]. In our case, this spreads the overall distribution of parameters. Since  $T^* = 1/(\sigma^* N^{1/2})$ , applying PLM to small datasets leads to an inaccurate estimate of the state point, biased towards a lower temperature. For disordered datasets, this corresponds to biasing the model towards the near-critical regime. The question is: how large is this effect?

In Fig. 5.11(b) we plot the dependence of the inferred temperature on the input temperature for various  $B$ . Taking the worse case,  $B = 1000$ , as an example, we find that the inference provides incorrect estimates of the state point in the entire range of temperatures considered. Even at the optimal conditions, where the inference error is minimal,  $T^* \approx 0.4T^0$ , mislabelling the state. Parameter estimates where PLM failed as in the F example (Fig. 5.5) in Fig. 5.11(b) correspond to low  $T^* \rightarrow 0$ , as the anomalously large inferred parameters cause the variance to tend to infinity. We clearly show that collecting more data allows the onset of this failure effect to be delayed, and a lower temperature state points closer to the P-SG transition to be correctly characterised. Irrespective of sample size, the inferred temperature  $T^*$  is lower than the input temperature  $T^0$ . The PLM temperature estimate is therefore systematically lower than the true temperature. As the SK model is not sensitive to  $\mu$  for  $\mu < 1$  we did not study the dependence of  $\mu^*$  on  $\mu^0$ . Table. 5.1 however, provides some small indication that we also expect PLM to under-estimate  $\mu$ .



**Figure 5.12:** Inferred temperature  $T^*$ , panel (a), and susceptibility measure  $C^2$ , panel (b), as functions of the input temperature  $T$  for three data quality conditions;  $B = 1 \times 10^3$  (circles),  $B = 2 \times 10^3$  (squares),  $B = 1 \times 10^4$  (diamonds). The black lines show the temperature and susceptibility as measured from simulations of the true model. Each data point and error bar in the black lines represents the mean and standard error calculated over  $3$  (for each  $B$  condition)  $\times 21 = 63$  independent model realisations.

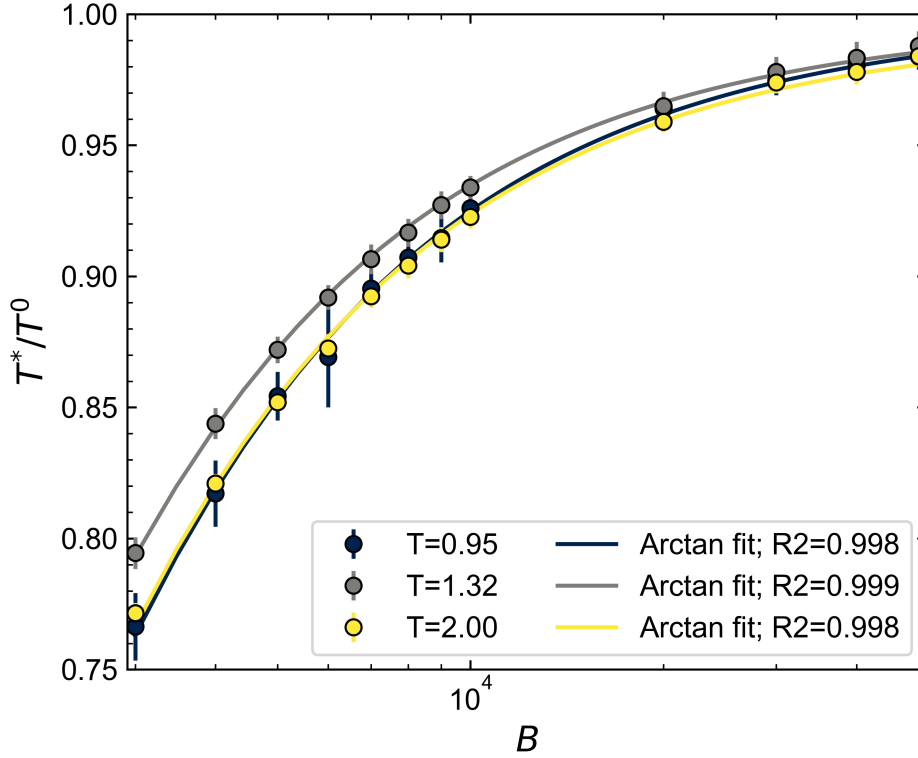


#### 5.4.2 Inferring Susceptibility from Finite Sized Datasets

The critical state is characterised by large correlated fluctuations. As such, we consider it important that the models we reconstruct via PLM accurately capture the fluctuations of the data. However, we have just shown that PLM on finite datasets will under-estimate the temperature, and expect that these miss-attributed temperatures will cause the inferred models to exhibit falsely enhanced critical fluctuations. This is because the spin-glass correlation  $C^2$  (and susceptibility  $\chi_{SG}$ ) increase rapidly as  $T$  decreases and one crosses the P-SG transition line. To investigate this, we re-simulate the models inferred via PLM for each  $T$  and produce an estimate of the correlations  $C^2$  corresponding to the inferred model. This process can be summarized as follows. For every state point:

- Produce 21 independent datasets of sample size  $B$
- Extract 21 PLM models (one per dataset)
- Run 6 MC simulations using the PLM estimate for  $10^5 \times N$  steps, sampled every  $10N$  steps ;
- Evaluate  $C^2$  over each simulation and average.

Fig. 5.12 shows the corresponding results for  $C^2$ ; the black line represents the correlations of the input models, while the coloured symbols show those corresponding to the PLM estimates for three different sample sizes. As suspected from the temperature shifts in Fig. 5.11, we observe that PLM over-estimates  $C^2$  for input models generated at high  $T$ . With decreasing



**Figure 5.13:**  $T^*/T^0$  as a function of  $B$  for three illustrative input temperatures; (dark-blue)  $T = 0.95$ , (grey)  $T = 1.4$  and (yellow)  $T = 2$ . Points and error bars for each  $B$  are means and standard errors obtained by repeating the GSI process for 21 independent input model realisations at each  $T$ . The coefficient of determination of the heuristic arctan fit (5.11) is  $R^2 > 0.980$  for all  $\tilde{B}$  plotted. See Fig. 5.15 for data at lower temperatures where the fitting quality is drastically reduced.

$T$ ,  $C^2$  reaches a peak value, which gets higher and is located closer in  $T$  to the true  $C^2$  peak with increasing  $B$ . For small sample sizes the arbitrarily strong couplings found for  $T < T_{\text{peak}}$  fix  $C^2 \rightarrow 0$ . In summary, our numerical experiments on the SK model indicate that PLM on small datasets provides couplings that under-predict  $T$  and artificially enhance  $C^2$ . PLM will therefore misattribute finite datasets stemming from the fluctuating P phase to a near-critical state point.

### 5.4.3 Formalising Temperature Dependence on Sample Size

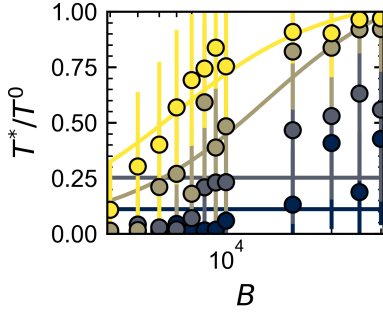
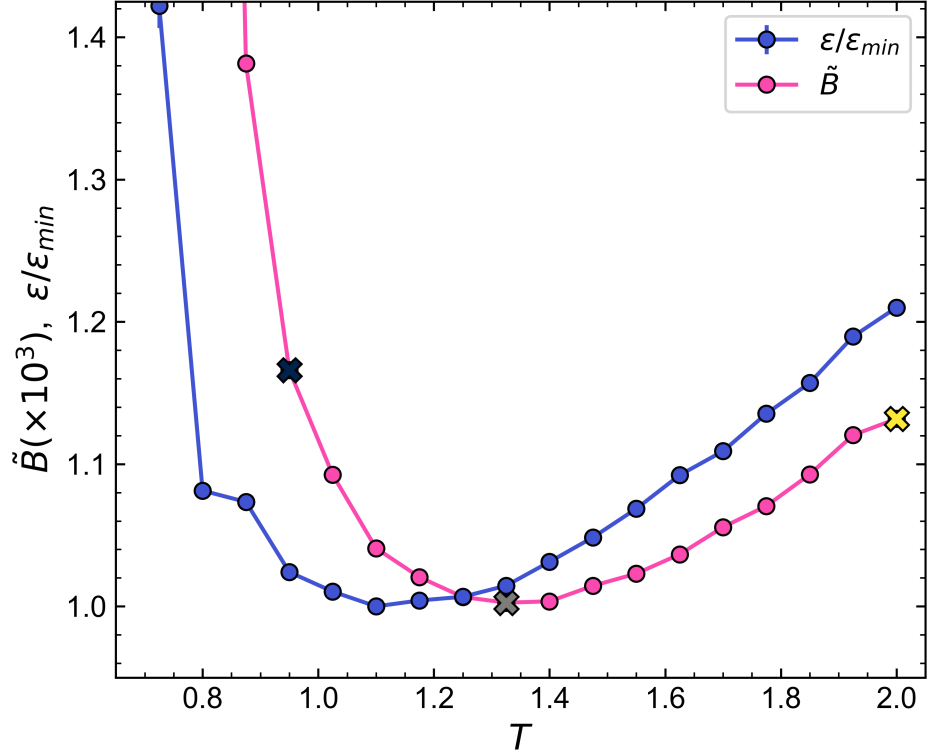
In the previous section, we qualitatively described that the inferred temperature depends on  $B$ . Here we quantitatively demonstrate that this effect is governed by a slow  $1/B$  convergence and that the temperature of the input model sets the learning difficulty of the problem. In Fig. 5.13 we plot the dependence of the ratio  $T^*/T^0$  on the sample size for different  $T$ . We find that for increasing  $B$  the curves follow a saturating behaviour, which can be fitted with high accuracy to the following heuristic model

$$T^* = (2T^{B \rightarrow \infty} / \pi) \times \arctan(B / \tilde{B}), \quad (5.11)$$

where  $T^{B \rightarrow \infty}$  and  $\tilde{B}$  are fitting parameters.  $T^{B \rightarrow \infty}$  measures the infinite sample size prediction of the temperature, while  $\tilde{B}$  quantifies the rate at which information is accumulated about the temperature, providing a measure of how “easy” it is to learn a given state point. Smaller values of  $\tilde{B}$  mean fewer samples are required to perfectly reconstruct the input temperature.

We show the dependence of  $\tilde{B}$  and  $\varepsilon$  on  $T$  in Fig. 5.14. Both share

**Figure 5.14:** The empirical scaling parameter  $\tilde{B}(T)$  and the error  $\varepsilon(T)$  when  $B = B_{\max} = 5 \times 10^4$  samples are used for PLM. Crosses show  $\tilde{B}$  for the correspondingly coloured  $T^*/T^0(B)$  saturation curves in FIG. 5.13. The last two input temperatures for which  $\tilde{B} \neq 0$  (i.e. for which the inference is possible from  $B_{\max}$ ) are at  $T = 0.725$ , with  $\tilde{B} = 6014$  and  $T = 0.8$  with  $\tilde{B} = 2714$ . We do not show these as  $\tau$  diverges from 1 for  $T < 0.8$  (see Fig. 5.10).



**Figure 5.15:** Temperature saturation plots as in Fig. 5.13, but for low-temperature input state points. Plotted temperatures are  $T = 0.5, 0.57, 0.65, 0.72$  from dark to light.

a non-monotonic behaviour indicating a correspondence between the scale  $\tilde{B}$  (quantifying the typical sample size to have small deviations in  $T^*/T^0$ ) and the average error on the couplings  $\varepsilon$ . The minimum of  $\tilde{B}$  is shifted further into the P phase, to  $T \approx 1.4$ . Note that for  $N = 200$ , the minimum number of samples is  $\tilde{B} \gtrsim 1000$ , pointing to the necessity of several thousand samples for reliable inference. For  $T < 0.8$  PLM performs so poorly that it becomes challenging to estimate  $\tilde{B}$  accurately for the range of  $B$  we explore. We show the saturating dependence of these low  $T$  state points in Fig. 5.15. For  $T = 0.5$  and  $T = 0.57$   $\tilde{B}$  is divergent; i.e. these temperatures cannot be learned even with our largest dataset  $B = 5 \times 10^4$ . This highlights a poignant issue for real datasets; the bias decay parameter  $\tilde{B}$  is not known *a priori* and can vary by orders of magnitude depending on the temperature of the input model. The limit defining a “small” dataset, therefore, depends on the underlying state point of the data. For  $x = B/\tilde{B} \geq 1$  (i.e. high quality data limit) the empirical arc-tan fit can be expanded as

$$\arctan(x) = \frac{\pi}{2} - \frac{1}{x} + \mathcal{O}(x^{-3}), \quad (5.12)$$

so that

$$T^* = T^{B \rightarrow \infty} - \frac{\tilde{B}'}{B} + \mathcal{O}(B^{-3}), \quad (5.13)$$

has a first-order linear dependence on  $1/B$ . In summary, the PLM method on the SK model displays biases that scale as the inverse of the sample size,  $B^{-1}$ . The magnitude of the bias strongly depends on the state point of the input data. This small sample bias causes the temperature to be under-

estimated, falsely enhancing the critical fluctuations exhibited by models inferred from near-critical paramagnetic data. Any PLM model inferred from fluctuating (i.e. dynamically varying) data will thus appear as closer-to-critical than it actually is.

## 5.5 Explaining the Sample Size Bias

We have identified that properties related to averages of the couplings, such as the temperature, are biased by small sample size effects. We explore the origin of this bias in this section.

### 5.5.1 Bias of Maximum Likelihood Estimates

Let us briefly diverge from the specific case of solving PLM, and consider a generic likelihood maximisation problem where the aim is to estimate some true set of parameter  $\boldsymbol{\theta}^0 = (\theta_0^0, \dots, \theta_k^0, \dots, \theta_n^0)$  from a dataset of  $B$  i.i.d observations. The maximum likelihood estimate (MLE) of these parameters is again indicated by  $\boldsymbol{\theta}^*$ , and it is well known that the bias of the MLE for each  $\theta_k^0$  may be expressed as [37, 211]

$$\begin{aligned} \theta_k^* &= \theta_k^0 + b_k(\boldsymbol{\theta}^0, B), \\ &= \theta_k^0 + \frac{b_{1,k}(\boldsymbol{\theta}^0)}{B} + \frac{b_{2,k}(\boldsymbol{\theta}^0)}{B^2} + \frac{b_{3,k}(\boldsymbol{\theta}^0)}{B^3} + \dots \end{aligned} \quad (5.14)$$

The bias of the estimate of the  $k^{\text{th}}$  parameter  $b_k$  depends both on the *complete* set of true parameters  $\boldsymbol{\theta}^0$  and the number of samples  $B$ . Clearly, as  $B \rightarrow \infty$ ,  $b_k \rightarrow 0$  and one recovers the true model perfectly. For simplicity, we will only consider the first-order bias in our discussion, so that

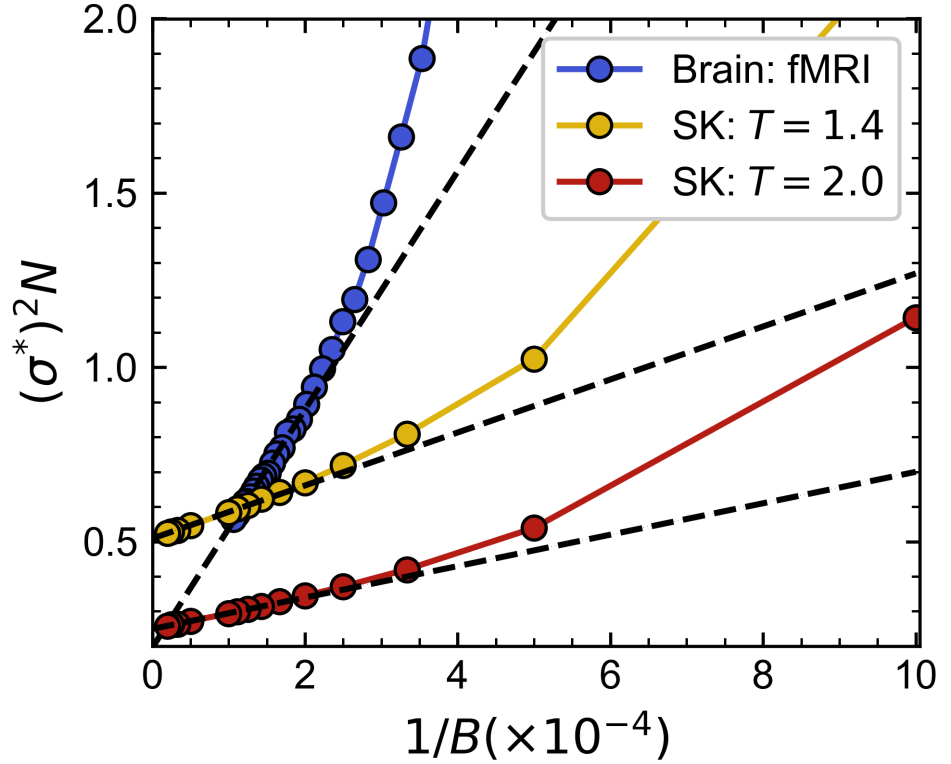
$$b_k(\boldsymbol{\theta}^0, B) \approx \frac{b_{1,k}(\boldsymbol{\theta}^0)}{B}. \quad (5.15)$$

We would therefore expect that each parameter of the logistic regressions performed by PLM would contain a bias which decays (to the leading order) as a function of  $1/B$ . Our analysis of PLM additionally revealed that average properties of the inferred parameter set, such as the standard deviation (i.e. the inverse temperature), also display a bias which vanishes linearly in  $1/B$  (see Fig. 5.16). We now investigate PLM specifically to see why this should be the case.

### 5.5.2 Bias of Average Properties in PLM

As a reminder, in PLM one maximises  $N$  pseudo-likelihoods  $\mathcal{L}_r$  to produce a set of non-symmetric inferred couplings. We will denote the PLM output for each of the  $r$  logistic regressions as  $\boldsymbol{\phi}_r^* = (h_r^*, \{J_{rj}^*\}_{j \neq r})$ . Each set  $\boldsymbol{\phi}_r^*$  fills the  $r^{\text{th}}$  row of the non-symmetric PLM parameter matrix  $\boldsymbol{\phi}^*$ . As we require our couplings to be symmetric we always perform a post-inference symmetrising step, setting  $\boldsymbol{\theta}^* = \frac{1}{2}[\boldsymbol{\phi}^* + (\boldsymbol{\phi}^*)^T]$  where  $T$  is the transpose of the matrix, to obtain our final PLM estimates of  $\mathbf{h}$  and  $\mathbf{J}$ . The final symmetric PLM estimate of each parameter is  $\theta_{ij}^* = 1/2(\phi_{ij}^* + \phi_{ji}^*)$ .

**Figure 5.16:** Convergence of the variance of the inferred couplings  $(\sigma^*)^2$  (scaled by system size) vs inverse sample number  $1/B$ . The dashed lines indicate linear fits to the asymptotic (large  $B$ ) regime of the data. Asymptotic intercepts of the two SK state points correspond to  $T^* = 1.40$  and  $T^* = 2.00$ ; PLM is able to perfectly reproduce the state point given infinite data. The gradient  $b_1$  of the asymptotic fits sets the severity of the small sample bias. This gradient depends strongly on the state point, system size and topology of the input model, e.g. we find  $b_1(\text{SK: } T=1.4) = 760$ ,  $b_1(\text{SK: } T=2.0) = 450$ , and  $b_1(\text{Brain}) = 3416$ .



We now investigate how the bias of average properties of the inferred parameter distribution, such as the variance  $\text{Var}[\boldsymbol{\theta}^*]$ , arises from the bias of each individual parameter. All  $\phi_{ij}$  have been obtained via a likelihood maximisation process. As before, we may thus describe these in terms of their first-order bias as

$$\phi_{ij}^* \approx \phi_{ij}^0 + \frac{b_{1,ij}(\phi^0)}{B}, \quad (5.16)$$

where each bias  $b_{1,ij}(\phi^0)$  depends on the full true parameter set and we discard all higher order bias terms. Note that the true values of  $\boldsymbol{\theta}^0$  and  $\boldsymbol{\phi}^0$  are equal, i.e.  $\theta_{ij}^0 \equiv \phi_{ij}^0$ . We now replace the index  $i$  with  $r$  to keep the connection to the index of the  $r^{\text{th}}$  row of the parameter matrix explicit. The average of our PLM estimates  $\boldsymbol{\theta}^*$  is:

$$\mathbb{E}[\boldsymbol{\theta}^*] = \frac{1}{N_p} \sum_{r \leq j} \theta_{rj}^*, \quad (5.17)$$

where  $N_p = N + N(N - 1)/2$  is the number of parameters in the upper triangle and diagonal of the matrix  $\boldsymbol{\theta}^*$ , which due to the symmetry of  $\boldsymbol{\theta}^*$  fully describes the model. Rewriting this in terms of  $\phi_{rj}$  gives:

$$\mathbb{E}[\boldsymbol{\theta}^*] = \frac{1}{2N_p} \sum_{r \leq j} (\phi_{rj}^* + \phi_{jr}^*). \quad (5.18)$$

We then use (5.16) to write:

$$\begin{aligned}
\mathbb{E}[\boldsymbol{\theta}^*] &= \frac{1}{2N_p} \sum_{r \leq j} \left( \phi_{rj}^0 + \frac{b_{1,rj}}{B} + \phi_{jr}^0 + \frac{b_{1,jr}}{B} \right) \\
\mathbb{E}[\boldsymbol{\theta}^*] &= \left( \frac{1}{2N_p} \sum_{r \leq j} 2\phi_{rj}^0 \right) + \left( \frac{\sum_{r \leq j} b_{1,rj} + b_{1,jr}}{2N_p B} \right) \quad (5.19) \\
\mathbb{E}[\boldsymbol{\theta}^*] &= \mathbb{E}[\boldsymbol{\theta}^0] + \frac{b_{1,\mathbb{E}[\boldsymbol{\theta}]}(\boldsymbol{\theta}^0)}{B},
\end{aligned}$$

where in (5.19) we used the fact that  $\boldsymbol{\phi}^0 \equiv \boldsymbol{\theta}^0$  and the symmetry of  $\boldsymbol{\theta}^0$ . The constant  $b_{1,\mathbb{E}[\boldsymbol{\theta}]}$  is made up of sums of the individual biases  $b_{1,ij}$ , and depends on  $\boldsymbol{\theta}^0$ . We can now clearly see that the PLM estimate of the parameter mean is also affected by a true model-dependent bias which decays as a function of  $1/B$ . We noted previously that the phase of the SK model in the parameter region we are interested in (i.e. the P-SG transition region) is not sensitive to  $\mu$ . Moreover, we have shown that we can use the variance of the coupling distribution to define the temperature of our inferred models. We therefore also want to express this in terms of the bias. The variance is defined as

$$\sigma_{\boldsymbol{\theta}}^2 = \text{Var}[\boldsymbol{\theta}^*] = \mathbb{E}[(\boldsymbol{\theta}^*)^2] - \mathbb{E}[\boldsymbol{\theta}^*]^2, \quad (5.20)$$

and once more using (5.17) we find

$$\begin{aligned}
\mathbb{E}[(\boldsymbol{\theta}^*)^2] &= \frac{1}{4N_p} \sum_{r \leq j} \left[ (\phi_{rj}^0 + \phi_{jr}^0)^2 + \right. \\
&\quad \left. \frac{2 \left( b_{1,rj}(\phi_{rj}^0 + \phi_{jr}^0) + b_{1,jr}(\phi_{jr}^0 + \phi_{rj}^0) \right)}{B} + \right. \\
&\quad \left. \mathcal{O}(B^{-2}) \right] \\
&= \mathbb{E}[(\boldsymbol{\theta}^0)^2] + \frac{b_{1,\mathbb{E}[\boldsymbol{\theta}^2]}(\boldsymbol{\theta}^0)}{B},
\end{aligned} \quad (5.21)$$

where  $b_{1,\mathbb{E}[\boldsymbol{\theta}^2]}$  is another bias constant depended on the true model. We again use the symmetry of  $\boldsymbol{\phi}^0$  and equivalence of the true distributions,  $\boldsymbol{\phi}^0 \equiv \boldsymbol{\theta}^0$ , to get the final result. Using (5.19) and (5.21) we find that to the first-order in  $1/B$

$$\text{Var}[\boldsymbol{\theta}^*] = \text{Var}[\boldsymbol{\theta}^0] + \frac{b_{1,\sigma_{\boldsymbol{\theta}}^2}(\boldsymbol{\theta}^0)}{B}, \quad (5.22)$$

where  $b_{1,\sigma_{\boldsymbol{\theta}}^2} = b_{1,\mathbb{E}[\boldsymbol{\theta}^2]} - 2\mathbb{E}[\boldsymbol{\theta}^0]b_{1,\mathbb{E}[\boldsymbol{\theta}]}$  is the constant setting the bias of the estimated variance. The above results are calculated over the entire parameter set  $\boldsymbol{\theta} = \{\mathbf{h}, \mathbf{J}\}$ , but they also hold over the individual distributions of  $\mathbf{h}$  and  $\mathbf{J}$ . We now have all the ingredients required to show how our PLM temperature estimates depend on  $B$ . We previously defined the empirically measured temperature of our inferred models via

$$T^* = 1/(\sigma_J^* N^{1/2}), \quad (5.23)$$

where  $\sigma_J^* = \sqrt{\text{Var}[\mathbf{J}^*]}$ . Using (5.22) and expanding in the limit of small bias ( $\text{Var}[\mathbf{J}^0] \gg b_{1,\sigma_J^2}/B$ ) we can express the asymptotic convergence of the temperature estimate as

$$\begin{aligned} T^* &= \frac{1}{\sigma_J^* N^{1/2}} = \frac{1}{N^{1/2}} \left( \text{Var}[\mathbf{J}^0] + \frac{b_{1,\sigma_J^2}(\boldsymbol{\theta}^0)}{B} \right)^{-1/2}, \\ &\approx T^0 - \frac{b_{1,\sigma_J^2}(\boldsymbol{\theta}^0) N (T^0)^3}{2B} + \dots, \\ &= T^0 - \frac{b_{1,T}(\boldsymbol{\theta}^0)}{B}, \end{aligned} \tag{5.24}$$

to the leading order of  $1/B$ . The individual first-order biases on each parameter thus cause average properties of the parameters like the temperature to also vary as  $1/B$ . We can now rationalise our previous empirical arc-tan fit

$$T^* = (2T_{B \rightarrow \infty} / \pi) \times \arctan(B/\tilde{B}), \tag{5.25}$$

by noting that for  $x = B/\tilde{B} \geq 1$

$$\arctan(x) = \frac{\pi}{2} - \frac{1}{x} + \mathcal{O}(x^{-3}), \tag{5.26}$$

so that

$$T^* = T_{B \rightarrow \infty} - \frac{\tilde{B}'}{B} + \mathcal{O}(B^{-3}), \tag{5.27}$$

follows the same first-order linear dependence on  $1/B$  as (5.24). Collectively, the prefactors  $b_{1,ij}(\mathbf{h}^0, \mathbf{J}^0)$  set the difficulty of learning a given state point (i.e. model) by increasing or decreasing the amount of data required to dissipate the bias. Biases to statistical averages, like the temperature or standard deviation, are proportional to sums of these parameters. Our investigation of the SK model however revealed that the bias strength (which we measured via  $\tilde{B}$ ) was a function of the state point (i.e.  $T^0$ ) and not the specific parameter realisation  $\boldsymbol{\theta}^0$ . We thus additionally show that the bias behaves as

$$b_{1,ij}(\mathbf{h}^0, \mathbf{J}^0) \approx b_{1,ij}(T^0), \tag{5.28}$$

that is, the bias is, to a good approximation, purely a function of the state point of the input model (i.e. the inverse temperature). We can now also re-contextualise previous computational [15, 16, 118, 202] and theoretical [204, 210] findings about the dependence of the inverse Ising inference error on the input state point in terms of the bias. Topology, state point and other properties of the true model set the size of the bias terms  $b_{1,ij}(\boldsymbol{\theta}^0)$  and modify the amount of data required to remove the bias. Irrespective of the values of  $b_{1,ij}$  however, the rate of convergence is always governed by a slow dependence on  $1/B$ .

### 5.5.3 Small Sample Sizes and Logistic Regression

The above discussion explains why PLM generally underestimates the temperature of the input dataset, especially when  $B$  is small. However, our analysis of the bias has done little to explain the type of catastrophic failure encountered in the strongly coupled F regime, as shown in Fig. 5.5. These failures are instead the result of another small sample size effect specific to each logistic regression performed in PLM. Logistic regression is known to be affected by an additional small sample size issue termed *separation* [208]. Separation occurs when a subset of covariates (e.g.  $\mathbf{s}_{\text{sep}} \subset \mathbf{s}_{\setminus r}$ ) in the logistic regression can perfectly predict the outcome variable ( $s_r$ ). When data is separated the MLE of the logistic regression does not exist [208], and infinitely large parameter estimates are inferred [212].

Most commonly the separation will be *quasi-complete* and only parameters associated with  $\mathbf{s}_{\text{sep}}$  will be infinite, with the remaining parameter estimates remaining relatively unaffected. In real settings, where the logistic regression is solved numerically, the precise values of the separated parameters will not be infinite and instead depend on the convergence criteria of the numerical optimization scheme [212]. Methods which implicitly remove the first-order bias term through modifying the log-likelihood function [37] have been shown to control separation [212, 213], highlighting the connection of this effect to the first-order bias. Separation is often encountered when sample sizes are (very) small due to physical limitations on the available data, for example in social sciences [212] or medical studies [214].

With regards to PLM on the SK phase diagram, we expect that (quasi-complete) separation may commonly occur when the system is governed by slow dynamics, such as in the SG and F phases. The target variable of each PLM equation is the spin  $s_i$ , and due to the slow dynamics, we expect that throughout a given finite length trajectory there will commonly be at least one spin  $s_{j \neq i}$  which perfectly correlates with  $s_i$ . Separation then causes couplings of large size to be inferred for these specific  $J_{ij}$ , with the inferred value of the parameter  $J_{ij}$  depending only on the convergence criteria chosen for the numeric optimization of the logistic score function [212]. Although the hallmark signs of separation have previously been commented on in other studies of PLM (e.g. in Ref. [15]) none - to our knowledge - have previously linked the two conceptually. To summarise, low temperature or strongly coupled state points therefore cannot be inferred with this logistic regression-based technique.

## 5.6 Inference for Varied System Sizes

So far we have largely focused our analysis on a single system size,  $N = 200$ . This is because the analysis of other system sizes lead to consistent conclusions, because simulating this intermediate system size only required intermediate computational resources, and because  $N = 200$  was similar to system sizes in real fMRI analyses [13]. However, we also looked at the performance of PLM across the zero-field SK phase diagram for a variety of other system sizes. We will present the results of these investigations in this section and will show that they largely confirm what was already stated for the  $N = 200$  case. In total, we tested PLM on systems of size



$N = \{50, 100, 200, 400, 800\}$ . We generated multiple model realisations for each state point in the phase diagram, the number of which depended on  $N$ ; 6 realisations were drawn for  $N = 50$  and 100, 3 for  $N = 200$  and 400 and 1 for  $N = 800$  due to computational limitations. For each realisation we fixed the MC meta-parameters to  $p_{mc} = \{10^6, 10^7, 10^3\}$  and collected  $B = 10^4$  samples for all state points.

### 5.6.1 Phase Diagrams

The resulting phase diagrams for  $q$ ,  $C^2$  and  $\varepsilon$  are shown in Fig. 5.17. We see that these phase diagrams look largely the same for varying  $N$  as they did for  $N = 200$ . Finite size effects do however play a role and impact the phase diagram. Focusing initially on  $q$  and  $C^2$ , we see that (as expected) finite size effects cause the transitions to shift towards the  $N \rightarrow \infty$  transition lines. This is shown both by the peak of  $C^2$  shifting towards the transition lines and by the regime of high  $q$  corresponding more closely to the F phase in the thermodynamic limit. We also observe a “sharpening” of the transitions, which we expect from finite-size scaling. The areas of parameter space with heightened  $C^2$  become smaller and concentrate around heightened peak lines of  $C^2$ . These finite-size effects are also accompanied by interesting changes to the behaviour of the error. The contours of  $\varepsilon$  remain parallel to the phase transitions for all  $N$ . We do find, however, that the error becomes smoother as  $N$  increases, and that the contours more closely align with the transition lines. This reinforces our previous hypothesis that the error is a function of distance to the phase transition only. The region of minimum error remains in the paramagnetic phase for all  $N$ . The values of the errors themselves increase with  $N$ , implying a dependence of the error on the size of the system.

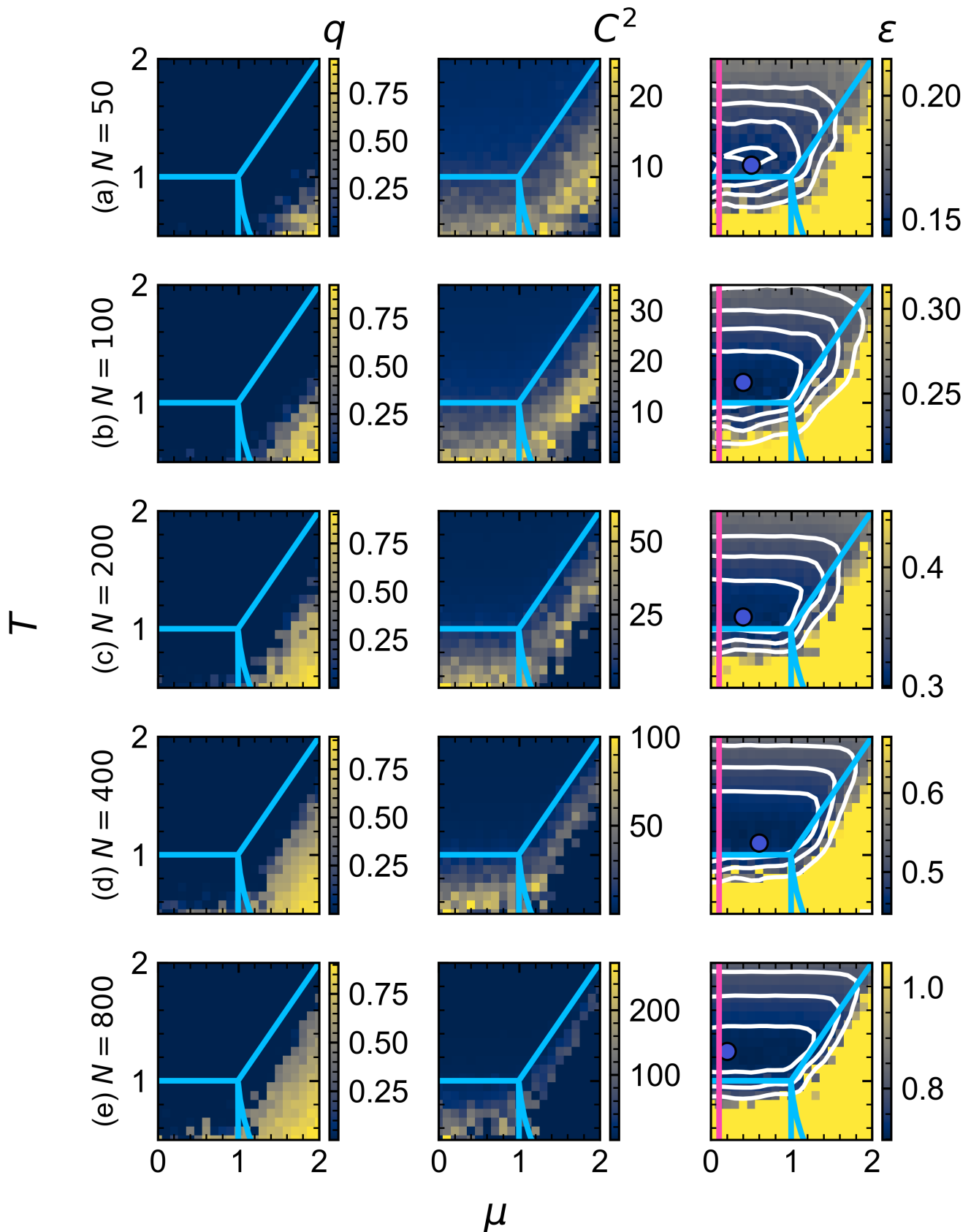
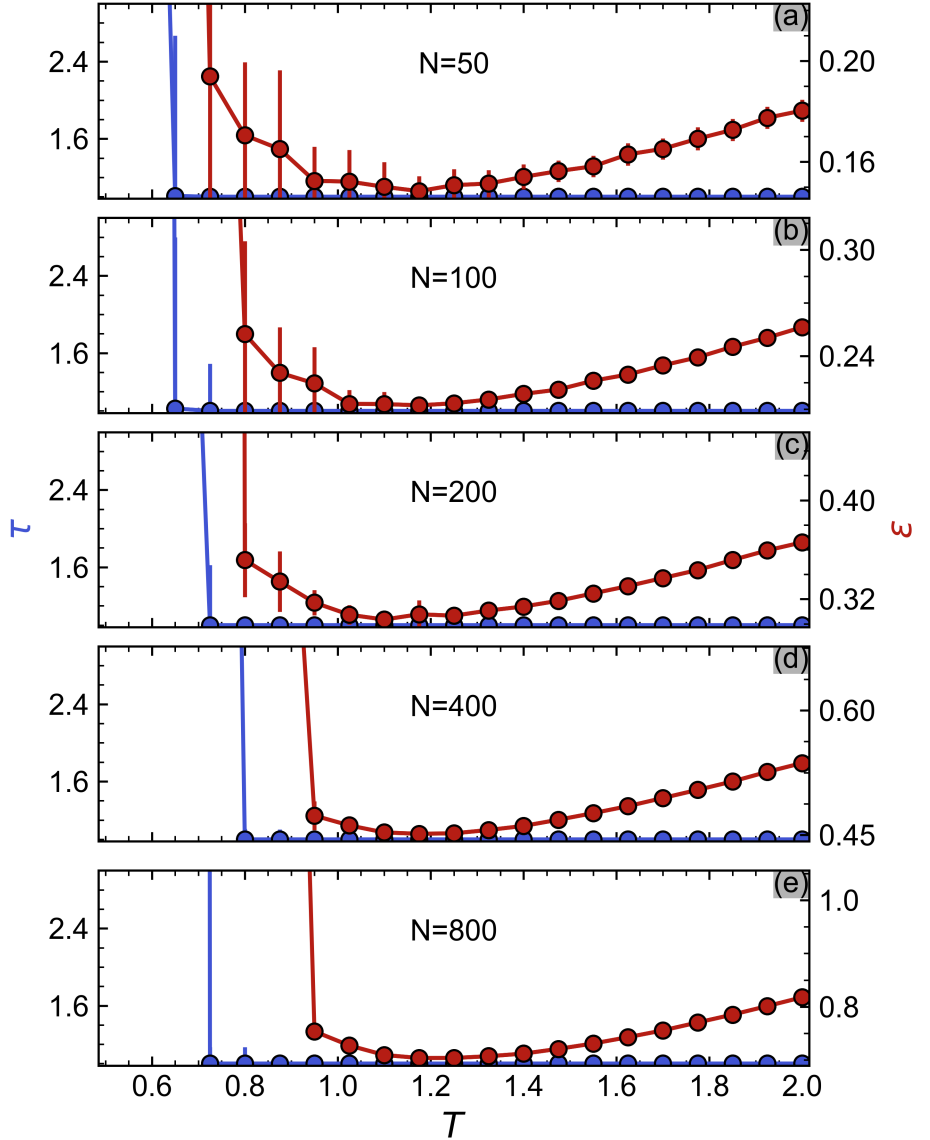


Figure 5.17: The order parameter  $q$ , susceptibility measure  $C^2$  and inference error  $\varepsilon$  across the SK phase diagram as the number of spins  $N$  varies. Each row shows the aforementioned observables for a fixed value of  $N$ . Full lines have the same meanings as in Fig. 5.9. We threshold  $\varepsilon$  so that  $\varepsilon_{\max} = 1.5\varepsilon_{\min}$ .

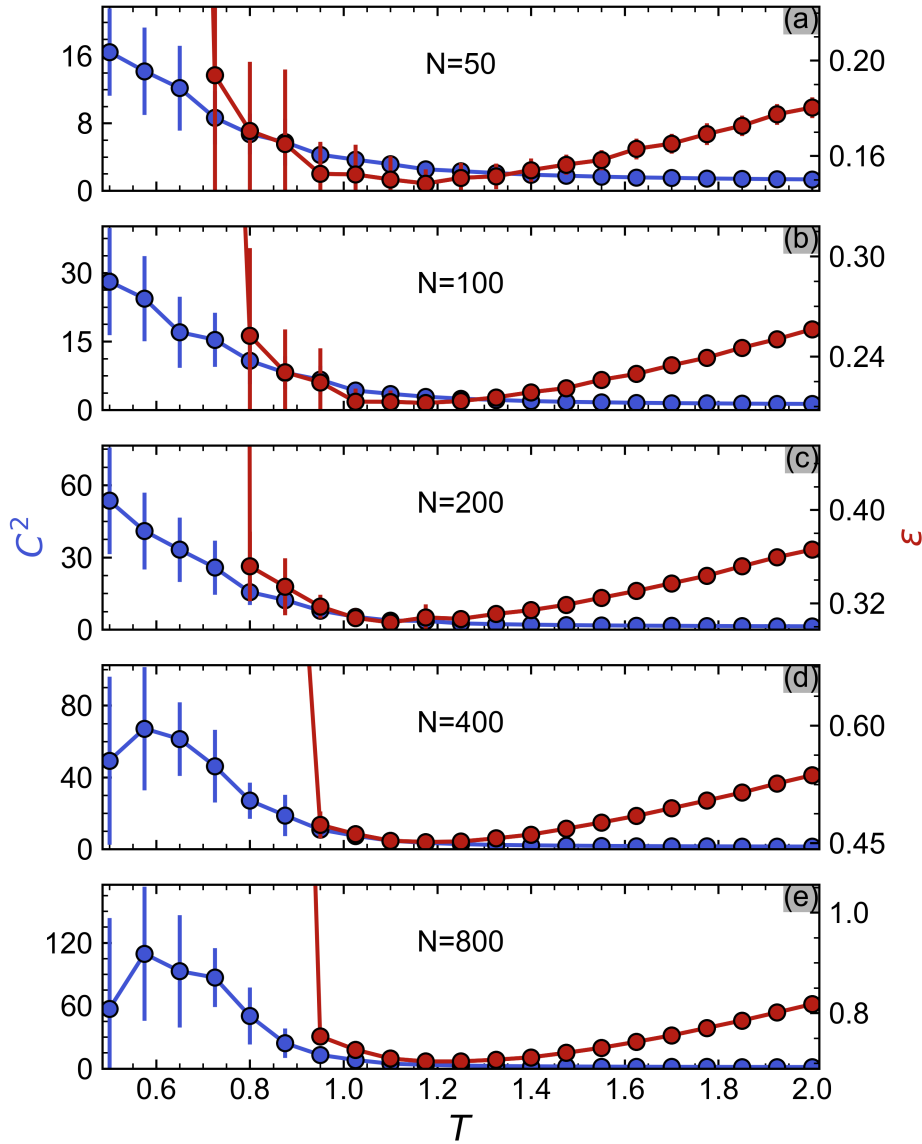
**Figure 5.18:** Average auto-correlation time  $\tau$  and inference error  $\varepsilon$  as functions of  $T$  in the  $\mu \leq 0.6$  region of the SK phase diagram for various  $N$ . We note a non-trivial dependence of  $\varepsilon$  on  $T$  for all  $N$  across a region where  $\tau = 1$ . The minimum error for all  $N$  is within the paramagnetic phase. The overall error increases as  $N$  increases. We produce these curves by averaging over all inferred models with  $\mu < 0.6$  for each  $T$ . Error bars represent standard errors over the same samples.



### 5.6.2 Error across the P-SG transition

For  $\mu < 1$  the phase of the SK model is not sensitive to  $\mu$ , and the transition from P to SG phases is captured entirely by the temperature. We perform a similar analysis as in section 5.3.2 and measure how  $\varepsilon$ ,  $\tau$  and  $C^2$  vary as functions of temperature in this region of the phase diagram for different  $N$ . We perform this analysis without generating additional data by averaging over the observables overall  $\mu \leq 0.6$  for each  $T$ , which we justify by the above in-sensitivity to  $\mu$ .

We track  $\tau$  and  $\varepsilon$  as functions of  $T$  in Fig. 5.18 to again ensure that the observed error dependence is truly a result of the PLM method and not due to a reduction in the number of independent samples contained in the dataset. As for our detailed analysis of the  $N = 200$  model with  $\mu = 0.1$ , we find that  $\tau = 1$  for the entire  $T$  range when the error is smooth. The observed minima and shape of  $\varepsilon$  are therefore consequences of PLM only. We again highlight that the value of  $\varepsilon_{\min}$  increases with  $N$ ,



**Figure 5.19:** Average susceptibility  $C^2$  and inference error  $\varepsilon$  as functions of  $T$  for  $\mu \leq 0.6$  and varying  $N$ . As  $N$  increases we see that the maximum of  $C^2$  increases and occurs at higher temperatures. Our data on the multiple system sizes are not good enough to perform a proper finite-size scaling analysis as we cannot clearly identify the temperature at which  $C^2$  is maximal for each  $N$ . This is obvious if we compare with Fig. 5.12, where we performed higher quality simulations for a single line of constant  $\mu$  in the  $N = 200$  system and found  $C^2$  to peak around  $T = 0.65$ , a feature not present here. We produce these curves by averaging over all inferred models with  $\mu < 0.6$  for each  $T$ .

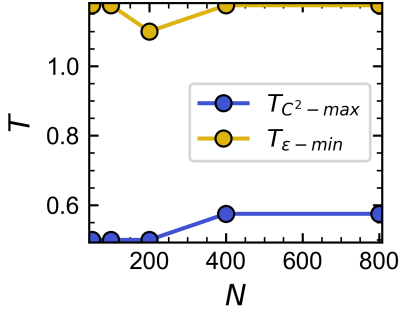


Figure 5.20: The temperatures at which  $C^2$  is maximised (blue) and  $\epsilon$  is minimised (orange) as functions of  $N$ .

and remains within the P regime. We then look at the co-dependence of  $C^2$  and  $\epsilon$  on  $T$  in Fig. 5.19. The minimum error occurs in a region with small, but heightened critical fluctuations, corresponding to a super-critical paramagnetic state point. Inference failure (which we define as the point at which  $\epsilon$  deviates from the smooth behaviour) occurs at progressively higher temperatures as  $N$  increases. This may be due to finite size effects shifting the transition temperature to higher  $T$ , or because larger systems inherently require more parameters to be inferred and therefore require more data for PLM to perform well near criticality. We do not have strong enough statistics from our measurements to identify the exact location of the  $C^2$  peak for each  $N$  to make definitive statements regarding this. We track the temperatures at which the  $C^2$  maximum and  $\epsilon$  minimum occur and plot them as functions of  $N$  in Fig. 5.20. Our data suggests that for  $N \leq 200$   $T_{C^2-max} \leq 0.5$ , and for  $N = 400, 800$   $T_{C^2-max} \approx 0.6$ . Please note that we recognise that our statistics are poor and that we treat these results with a high degree of scepticism. Indeed we analysed the  $N = 200$  model in more detail in Fig. 5.12, and found that when more data was collected the transition temperature could be identified within the range  $T \in [0.5, 2]$  which we study. While the statistics of the  $C^2$  estimated obtained in the critical regime are poor, estimates of the minimum error region are good. We find that the location of  $T_{\epsilon-min}$  does not vary significantly with  $N$ , staying constant at approximately  $T = 1.2$ . Interestingly, this independence would suggest that as the system size grows the error minimum should converge to a constant offset of  $\delta T \approx 1.2 - T_c(N \rightarrow \infty) = 1.2 - 1 = 0.2$  above the P-SG transition.

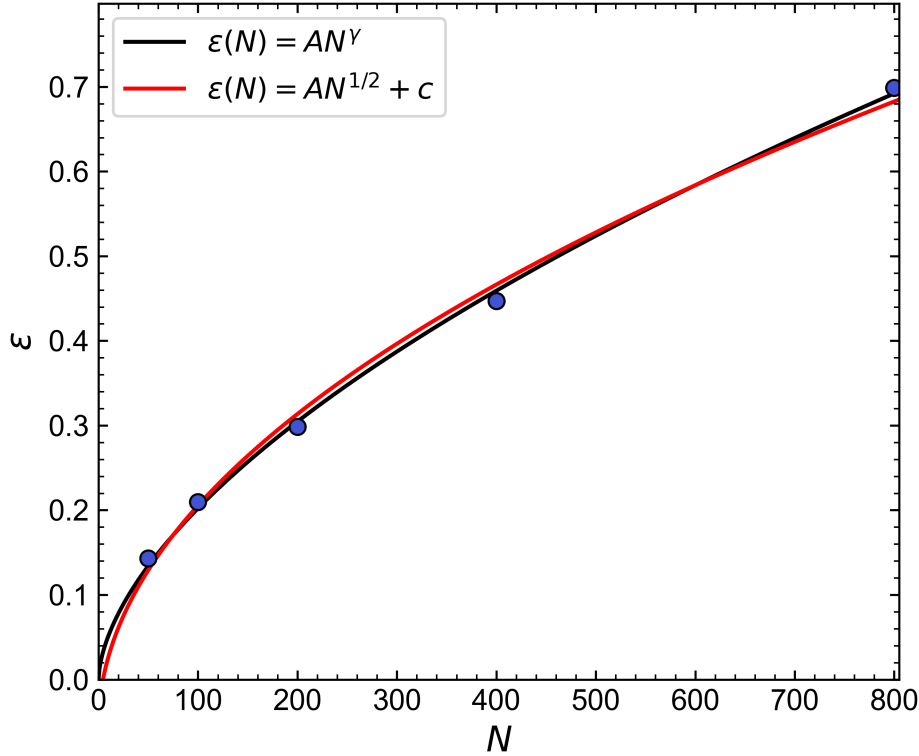
### 5.6.3 Error Dependence on System Size

So far we have only mentioned that  $\epsilon$  generally increases as  $N$  increases. We now wish to quantify this dependence. As noted, finite size effects will cause the phase corresponding to each  $(\mu, T)$  pair to change as  $N$  increases. We thus decided that the dependence of the error on  $N$  was best assessed by tracking the minimum error state point in the phase diagram for each  $N$ , indicated by the blue points in Fig. 5.17. We plot the dependence of  $\epsilon_{min}$  on  $N$  in Fig. 5.21 and find that our data can be fitted with both with a pure power-law,  $\epsilon(N) = AN^\gamma$  ( $R_{adj}^2 = 0.996$ ,  $\gamma = 0.59$ ) and a power-law with fixed  $\gamma = 0.5$  and an offset ( $R_{adj}^2 = 0.989$ ) with high accuracy, leading us to believe that  $\epsilon \sim N^{1/2}$  for fixed  $B$ .

### 5.6.4 Sub-Sampling to Find the Bias

In section 5.3, we generated a number of datasets of different lengths to understand the convergence of the temperature as a function of the sample size  $B$ . For real datasets, this process will not be possible. An alternative standard approach is to draw “sub-samples” from the full dataset of length  $B$  and perform PLM on each of these, tracking the observables of interest (e.g. the temperature) as the size of the sub-sample increases. We now extract bias parameters by sub-sampling datasets from the minimum error state points for varying  $N$  identified in Fig. 5.17 (blue points)<sup>16</sup>. This is exactly the same data we used in section 5.6.3. Each of these datasets

<sup>16</sup> As a reminder, for  $N = 50$  and 100 we analysed 6 realisations, for  $N = 200$  and 400 we analysed 3 realisations, and  $N = 800$  we analysed 1 realisation



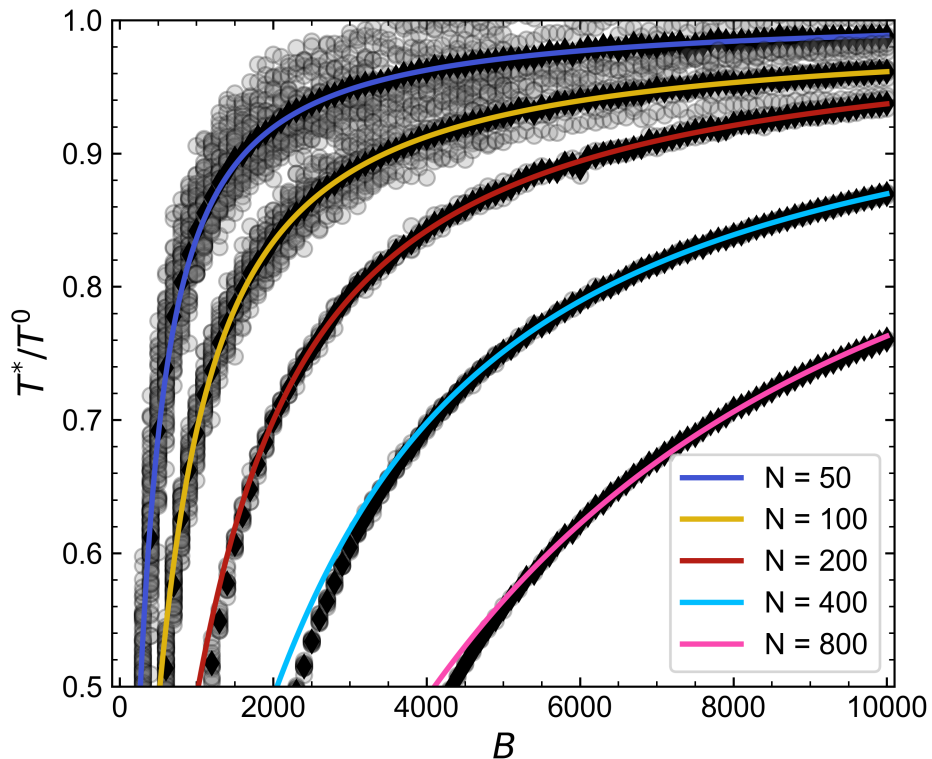
**Figure 5.21:** The dependence of the minimum error in each phase diagram (identified by the blue points in the right column of Fig. 5.17) on the system size  $N$ . We fit our empirical data with two power-law functions, the mathematical form of which is shown in the legend.

contains a fixed  $B_{\max} = 10^4$  samples. We generate each sub-sample in a greedy way, by taking  $B$  samples from  $B_{\max}$  without replacement until either  $B_{\max}/B$  sub-samples are generated or there are not enough unique samples remaining to generate another complete sub-sample of length  $B$ . We then perform PLM on each of these sub-samples, and average over the output for each  $B$ . This way we ensure that as many as possible of the  $B_{\max}$  samples are used in the analysis of each  $B$  condition.

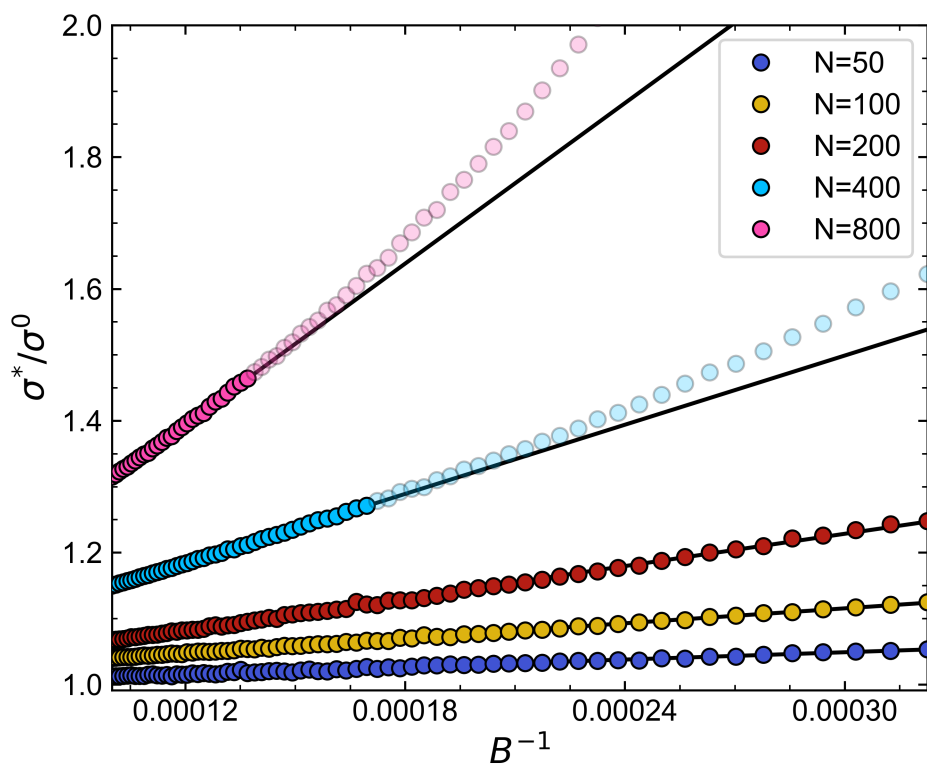
We plot  $T^*/T^0$  as functions of the sub-sample size  $B$  for each  $N$  in Fig. 5.22. Selecting only data with  $B \geq 5 \times 10^3$ , we perform an empirical arc-tan fit for each curve, as was done in section 5.4.3. For  $N = 50$  and 100, there is excellent agreement between the data and the fit even for  $B \leq 5 \times 10^3$ . But for  $N \geq 200$  we find a divergence between the fit and the data for low  $B$ , suggesting that not all sample sizes are in the asymptotic regime described by a  $1/B$  convergence. We also find that the best estimate for  $T$  rapidly decreases as  $N$  increases (as we expect from the behaviour of  $\epsilon$ ), with the optimal estimate for  $N = 800$  being  $\sim 25\%$  lower than the true temperature. System size thus clearly also impacts the size of the bias terms.

Although plotting  $T^*/T^0(B)$  is an intuitive way to visualise how our estimates improve as we accumulate more information, analysing the dependence of our observables in terms of  $1/B$  allows us to more closely establish how our estimates relate the first-order bias identified in section 5.5. With this in mind, we also show the dependence of  $\sigma^*/\sigma^0$  on  $1/B$  in Fig. 5.23. We now make use of a statistical method to identify the first-order bias regime for each  $N$  (see appendix A), showing points which are not part of the linear bias regime as transparent. We see the linearity in  $B^{-1}$  holds for all  $N \leq 200$ . We show linear fits to the linear regime as black lines and note

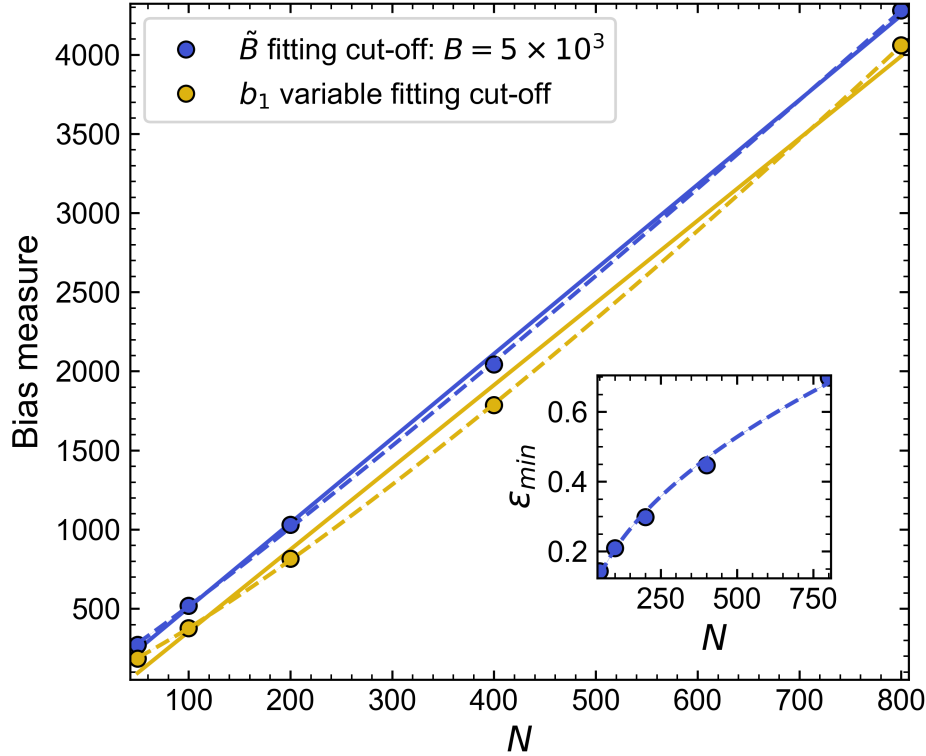
**Figure 5.22:** The saturation of the inferred temperature  $T^*/T^0$  as a function of  $B$  for the minimum error state points of varying  $N$  identified in Fig. 5.17. Transparent circles show inferred temperatures for all individual sub-samples from different realisations at each  $B$ , while black diamonds show averaged temperatures over all sub-samples for every  $B$ . We fit the average inferred temperature with our empirical arctan function to avoid biasing the fit towards low  $B$  values where we have more data points available. More data is required to remove the bias as  $N$  increases.



**Figure 5.23:** Errors in estimates of the standard deviation  $\sigma^*/\sigma^0$  of the couplings as a function of the inverse sample size  $B^{-1}$  for the minimum error state points of varying  $N$  identified in Fig. 5.17. Colours show different system sizes. Transparent points show data outside of the linear regime (see Appendix A). Black lines indicate linear fits to the linear bias regime and are used to calculate the bias strength parameter  $b_1$ .







**Figure 5.24:** We plot the two bias strength measures  $\tilde{B}$  and  $b_1$  calculated from the fits in Fig 5.22 and Fig 5.23 respectively as functions of  $N$ . These both show a linear dependence on  $N$ . Full lines show linear fits, dashed lines show power-law fits. The inset shows the dependence of the minimum error  $\epsilon_{\min}$  on  $N$  for the same state points.

a clear divergence from these for small  $B$  for  $N = 400$  and  $N = 800$ . The error of the standard deviation (i.e. temperature) estimate at  $B = B_{\max}$  approaches zero (i.e.  $\sigma^*/\sigma^0 \rightarrow 1$ ) as  $N \rightarrow 0$ . This clearly demonstrates that for the same  $B$  different systems can experience vastly different amounts of bias. Determining what constitutes a “small” sample size *a priori* is impossible. To confidently assert that a system is critical one must therefore always carry out a convergence analysis, e.g. by sub-sampling as done here, alongside the PLM analysis of the full dataset.

### 5.6.5 Bias Dependence on System Size

We have shown two valid ways of extracting the bias. By performing an empirical arc-tan fit to  $T^*(B)$  as in see Fig. 5.22 we obtain the bias measure  $\tilde{B}$ . Similarly, we can obtain the  $b_1$  as the gradient of a linear fit to the linear part of  $\sigma^*(1/B)$  (see Fig. 5.23). Intuitively we know that these two parameters should capture similar properties of the system (and so have the same dependence on  $N$ ). To confirm this, we plot both as functions of  $N$  in Fig. 5.24. We find both  $\tilde{B}$  and  $b_1$  are well described by simple linear functions of  $N$ . The values of the two bias measures are offset slightly from one another, with  $b_1$  generally being lower than  $\tilde{B}$ . Importantly, we find that the gradients of the linear fits to  $\tilde{B}(N)$  and  $b_1(N)$  are numerically very similar, taking values of 5.3 and 5.2 respectively. This confirms our hypothesis that both  $\tilde{B}$  and  $b_1$  capture the same property of the system, encoding the “difficulty” of learning a given input model. We believe that this linear dependence on  $N$  arises from the fact that although the full inverse Ising problem deals with estimating  $N + N(N - 1)/2$  parameters, each individual logistic regression equation only infers  $N$  parameters.



We previously found in Fig. 5.14 that  $\tilde{B}$  and  $\varepsilon$  showed a similar dependence on the state point. We thus considered these two variables as both measuring the same core concept; the difficulty of inferring a given state point. But we now find a discrepancy as our detailed analysis of the bias revealed that  $\tilde{B} \sim N$  (equivalently  $b_1 \sim N$ ), while our analysis of the inference errors showed  $\varepsilon \sim N^{1/2}$ , see Fig. 5.21. We consider our error measure more closely to explain this discrepancy.

### 5.6.6 Explaining the Error-Bias Discrepancy

As a reminder, each of our zero-field ( $\mathbf{h} = 0$ ) SK model realisations has a Gaussian distribution of couplings  $\mathbf{J}$  with mean

$$\mu^0 = \mu/TN, \quad (5.29)$$

and standard deviation

$$\sigma^0 = \sigma/TN^{1/2}. \quad (5.30)$$

Both  $\mu^0$  and  $\sigma^0$  therefore implicitly depend on  $N$ . We can express the error in terms of the bias as

$$\varepsilon = \frac{1}{B} \sqrt{\frac{\sum_{i \leq j} (b_1(\theta_{ij}^0))^2}{\sum_{i \leq j} (\theta_{ij}^0)^2}}. \quad (5.31)$$

As we have no fields we can express the denominator as

$$\sum_{i \leq j} (\theta_{ij}^0)^2 = N_J [(\mu^0)^2 + (\sigma^0)^2] \quad (5.32)$$

where  $N_J = N(N-1)/2$  is the number of couplings. Let us now also define a constant  $D = \sqrt{\sum_{i \leq j} (b_1(\theta_{ij}^0))^2}$  which captures the sum of the state point dependent bias of each coupling. We can therefore express the error in terms of the bias and the intensive variables  $\mu$  and  $\sigma$  as

$$\begin{aligned} \varepsilon &= \frac{D}{BN_J^{1/2}} ((\mu^0)^2 + (\sigma^0)^2)^{-1/2}, \\ \varepsilon &= \frac{D}{BN_J^{1/2}} \left( \frac{\mu^2}{T^2 N^2} + \frac{\sigma^2}{T^2 N} \right)^{-1/2}, \\ \varepsilon &= \frac{D}{BN_J^{1/2}} N^{1/2} T \left( \frac{\mu^2}{N} + \sigma^2 \right)^{-1/2}. \end{aligned} \quad (5.33)$$

We now note that for the SG-P region of the SK phase diagram  $\mu < 1$  so that  $\mu^2 \ll 1$ , and that  $N \gg 1$  for all system sizes considered. This means that in the SG-P region, the  $\mu^2/N$  term is negligible. Moreover, as  $N \gg 1$  we can approximate  $N_J = (N^2 - N)/2 \approx N^2/2$  (even for our smallest system size  $N = 50$ ,  $N/N^2 = 0.02$ ) so that the error in the SG-P transition region is given by

$$\varepsilon \approx \frac{D}{BN^{1/2}} \frac{T}{\sigma}. \quad (5.34)$$

Remembering that we found  $T(\varepsilon_{\min}) \approx \text{const}$  and that  $\sigma = 1$  throughout, we finally find that for fixed  $B$ ,  $\varepsilon \propto N^{-1/2}D$ . This explains the discrepancy between the error  $\varepsilon$  and our bias measure  $b_1$ , which we verify by comparing the re-scaled error  $\varepsilon N^{1/2}$  to  $b_1$  in Fig. 5.25. To conclude, the error measure  $\varepsilon$  does not accurately capture how the difficulty of the inverse problem depends on the system size  $N$ . Analysing the bias reveals a linear dependence on  $N$ , which is heuristically plausible as the number of parameters being estimated by each logistic regression in PLM is proportional to  $N$ .

## 5.7 Conclusion

The aim of this chapter was to explore the performance of PLM on data gathered from the much-studied SK model. We established familiarity with the types of errors encountered in PLM in section 5.2 by focusing on specific state points representative of the different phases in the SK phase diagram. These showed us that, depending on the strength of the couplings in the true model, we encounter two separate contributions to the inference error. We find that the PLM estimates  $\theta^*$  of the true parameters  $\theta^0$  are over-estimates and that this contributes to the error for all state points. For weak couplings (i.e. high temperatures) the overall structure or distribution of the input parameters can be inferred. But in the strongly coupled regimes, we find a secondary failure mode, whereby (infinitely) large parameters are inferred for certain rows (corresponding to particular spins) of the parameter matrix, causing a divergence of the error. We link this to a logistic regression-specific effect termed separation.

In section 5.3 we show that the inference error does indeed have a strong dependence on the state point of the data, verifying findings from previous authors [16]. Moreover, we show that the error is a function of distance from the critical point only as the error forms contours parallel to the phase transition lines. By analysing a line of constant  $\mu$  we show that the minimum error lies above the critical point in the paramagnetic phase. In section 5.4 we explore the origin of the error further, with a particular focus on finite (i.e. small) sample sizes. We find that PLM systematically overestimates the standard deviation of the input coupling distribution. By using the definition of our SK model parameters we can relate the standard deviation of the coupling distribution to the temperature of the model: PLM therefore systematically underestimates the temperature of the input model. We re-simulate the models and further show that the critical fluctuations in the inferred models are artificially enhanced. Models inferred from paramagnetic (i.e. dynamically varying) datasets are therefore made to appear more critical than they truly are. We further show that this effect drastically worsens when there is limited data available. The Brain is inherently a dynamic system [66], and so this poses serious questions for claims of criticality in models built using PLM from finite datasets. We formalise the dependence of the inference error on the state point by studying the convergence of the inferred temperature to the true temperature as a function

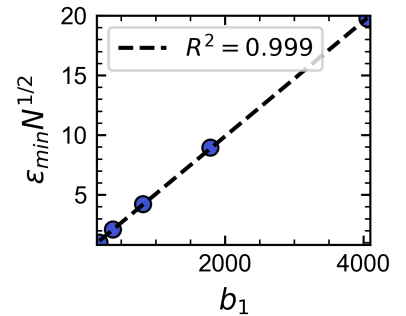


Figure 5.25: Rescaling the error by  $N^{1/2}$  shows close-to-perfect proportionality with the bias measure  $b_1$ .

of  $B$  and find that this is well described by an empirical arc-tan fit with an asymptotic linear dependence on  $1/B$ . We explain that this is a result of the standard bias of maximum likelihood estimates.

In section 5.6 we analyse PLM for a range of system sizes which are much larger than those previously reported in the literature. We find that the phase diagram of the error is largely unaffected by system size, allowing us to generalise the results from our detailed study of the  $N = 200$  phase diagram to other system sizes. We further find that the minimum error in the phase diagram depends approximately on  $N^{1/2}$ , but that the strength of the bias scales as  $\sim N$ . We conclude that the amount of data required to learn a given state point depends linearly on  $N$ .

The results presented in this chapter will be essential when interpreting inference results from real-world datasets where the underlying model is unknown. While the dependence of inverse Ising inference on the state point of the data is well understood for infinite datasets [204, 210], to the best of our knowledge, no previous authors have identified the small sample biases highlighted here. When analysing real-world datasets, the fact that the learning factor  $\tilde{B}$  depends on the unknown state point means that understanding the strength of the bias in a given dataset is vital to making trustworthy claims about the criticality of any model inferred using PLM.

## Summary of Chapter 5

- We benchmark the performance of pseudo-likelihood maximisation (PLM) on data simulated from the zero-field Sherrington-Kirkpatrick (SK) phase diagram. We varied the system size  $N$ , the state point  $(\mu, T)$  and the number of input samples  $B$  to understand how inference errors depended on these factors. We draw the following headline conclusions:

### **Errors are minimal in the paramagnetic phase**

Exploring the SK phase diagram for a range of  $(\mu, T)$  showed that errors were minimal in the paramagnetic phase above the spin-glass (SG) - paramagnetic (P) phase transition line.

### **Errors are functions of distance to the critical point**

We found error contours to line parallel to the phase transition lines. This effect becomes clearer as  $N$  increases and the phase transitions themselves sharpen. This shows that the error is a function of *distance* to the phase transitions. The behaviour across the SG-P transition line was explored in more depth by fixing  $\mu = 0.1$  and varying  $T$  only.

### **PLM underestimates the temperature**

We found that PLM systematically overestimates the standard deviation of the model. This corresponds to underestimating the temperature of an SK-like model. This effect worsens as  $B$  decreases and as the static SG and ferromagnetic (F) phases are approached. The temperature bias can be severe enough to cause the phase of the data to be misclassified.

### **PLM overestimates the susceptibility of dynamic data**

As a result of the temperature bias, PLM models inferred from dynamically fluctuating (i.e. paramagnetic) data exhibit falsely enhanced critical fluctuations. *Applying PLM to real dynamically fluctuating data thus biases the inferred model towards criticality.*

### **PLM fails at low temperatures due to separation**

We show that low-temperature failures of PLM are the result of a logistic-regression-specific effect termed separation.

- We link the temperature bias to the standard first-order bias of maximum likelihood estimates and introduce an empirical way to measure the strength of this bias.
- We find the amount of data required to solve the inverse Ising problem using PLM increases linearly with  $N$ .

## Chapter 6

# Correcting the Bias of Pseudo-Likelihood Estimates

In the previous chapter, we found that pseudo-likelihood maximisation (PLM) produces biased estimates of the temperature, which depended linearly on the inverse sample size. We linked this to the standard first-order bias of maximum likelihood estimators. We now explore ways to remove this bias, with the aim of improving the accuracy of our state-point estimation. We introduce a novel corrective method which removes the bias by enforcing self-consistency between the critical fluctuations of the input data and the inferred model.

### 6.1 Motivation and Success Metrics

Using the SK model as a benchmark, we established that PLM underestimates the temperature of the dataset, with the size of this effect depending both on the sample size  $B$  and the state-point ( $T^0$ ) of the input dataset. This effect arises as a result of generic biases in maximum likelihood methods which decays as a function of  $1/B$  in the large  $B$  limit. In the vicinity of the phase boundaries, we found that these biases may be sufficiently large to alter the labelling of the inferred phase. We have also identified that the failure of PLM in strongly coupled regimes can be related back to a logistic regression-specific effect that is also the result of this first-order  $1/B$  bias. This naturally leads to the question: can we remove the first-order MLE bias and produce better estimates of the state point from the same data? We explore this proposal in this section.

A range of bias correction methods exist in the literature [211], which can largely be grouped into explicit methods which correct for the bias *after* inferring the parameters, e.g. jackknife resampling [215], and implicit methods which correct the bias *during* inference via a modification (penalization) of the likelihood function [37, 212, 213, 216, 217]. Explicit methods are generally simpler to implement but will inherit any instabilities of the original MLE estimator. An example of such an instability is the infinite PLM parameter estimates we observed in the strongly coupled regime due to complete or quasi-complete separation [208, 212].

The focus of our research is to reliably estimate the unknown state-point of a given dataset. As such, we will assess the success of each method on

the following metrics: a) how well do they estimate the temperature  $T$ , i.e. can we classify the state-point and b) how well do they estimate the critical fluctuations  $C^2$ , i.e. do they capture the essential correlations of the data. We will assess the temperature of each corrected model through our standard relation with the standard deviation,  $T^* = 1/(\sigma_j^* N^{1/2})$ , and calculate  $C^2$  by performing Monte Carlo simulations of each corrected model.

## 6.2 Explicit Corrections

Explicit methods are corrective, i.e. we first perform unmodified PLM to estimate  $\theta^*$  and then shift the parameters in some way to obtain the bias-corrected parameter estimates. We explore three explicit bias correction methods, including standard jackknife re-sampling, a sub-sampling method based on the empirical arctan fit we explore in the last chapter, and a self-consistency enforcing correction based on our observation that the PLM estimates over-estimate the critical fluctuations of the data.

### 6.2.1 Correction 1: Sub-Sampling Temperature Estimate

As explored in section 5.4, we observed that the effect of the bias is mainly limited to a broadening of the parameter distribution, reflected by our underestimates of  $T^*$  and overestimates of  $\sigma^*$ . We analysed the convergence of the (inverse) temperature in  $B$  using

$$T^*(B) = (2T^{B \rightarrow \infty} / \pi) \times \arctan\left(B/\tilde{B}\right), \quad (6.1)$$

and  $1/B$  using

$$\sigma^*(1/B) = \sigma^{B \rightarrow \infty} + \frac{b_1}{B}, \quad (6.2)$$

to obtain estimates for the bias associated with a given state point. This process conveniently also produces estimates for the temperature as  $B \rightarrow \infty$ , through the fitting parameter  $T^{B \rightarrow \infty}$  or the intercept  $\sigma^{B \rightarrow \infty}$  (remembering that  $T^* = 1/\sigma^* N^{1/2}$ ), providing us with an estimate of the unbiased temperature of our dataset. We can re-scale our PLM parameters to construct a model which corresponds to this temperature by setting

$$\theta^{ss} = \frac{\theta^*}{T_f} \quad (6.3)$$

where we refer to  $T_f$  as the fictive temperature, which in this case is equal to  $T_f = T^{B \rightarrow \infty} / T^*$ . Note that when  $T^{B \rightarrow \infty} \geq T^*$ ,  $T_f \geq 1$  so that the correction corresponds to “shrinking” the parameter distribution. As we know that PLM over-estimates the spread, this intuitively shows that we will correct towards  $T^0$ . We refer to this correction as the sub-sampling (SS) correction. We note that the Jackknife correction shifts each parameter  $\theta_{ij}$  individually, but that the temperature correction we propose here multiplies all couplings by a common constant  $T_f$ . This sub-sampling correction thus maintains the ranking of the PLM parameter estimates, only re-scaling their values. The sub-sampling correction is less computationally intense than performing Jackknife as one does not need to sub-sample for

all  $B$  to establish  $T^{B \rightarrow \infty}$ . Indeed we showed in the previous chapter that 15 data points were enough to produce accurate fits to the data.

## 6.2.2 Correction 2: Enforcing Self-Consistent Correlations

In the previous chapter, we demonstrated that the MLE bias of the PLM parameters is captured by a global property of the parameter distribution (the temperature) and that the inferred PLM models over-estimate  $C^2$ . We thus propose to correct the bias by requiring that the models we infer display  $C^2$  as close as possible to that estimated from the input data. In other words, we aim to infer models whose fluctuations are self-consistent with those of the input dataset. To do so, we perform a second optimisation *after* estimating the PLM parameters where we minimize the objective function

$$\mathcal{L}'(T_f) = [C_{\text{input}}^2 - C_{\text{MC}}^2(T_f)]^2, \quad (6.4)$$

where  $C_{\text{input}}^2$  is  $C^2$  measured from the input dataset and  $C_{\text{MC}}^2(T_f)$  is calculated from MC simulations of a re-scaled PLM model with parameters  $\theta^{T_f} = \theta^*/T_f$ . Note again the introduction of the re-scaling parameter  $T_f > 0$ , which acts as a *fictive temperature* which can shift the state-point of the inferred model through phase-space. We denote the optimal value of  $T_f$  by  $T_f^{\text{sc}}$ , with the corresponding corrected parameter estimates being  $\theta^{\text{sc}}$ , and refer to this correction as the self-consistency (SC) correction.  $C^2$  has a strong dependence on  $B$ , so we match the amount of data in the input and in the MC simulations  $B_{\text{input}} = B_{\text{MC}}$  (see Fig 6.1). In practice we calculate  $C_{\text{MC}}^2$  for 6 independent MC simulations of length  $B_{\text{data}}$  for every  $T_f$ , and then feed the average over these 6 runs into (6.4). We chose this number to match the number of cores on the analysis computer, allowing the simulations to be performed in parallel. We minimize the SC objective function (6.4) using SciPy’s `scipy.optimize.minimize` routine [218] with a simple Nelder–Mead [219] minimization, which we found to perform well for this one-dimensional optimization problem.

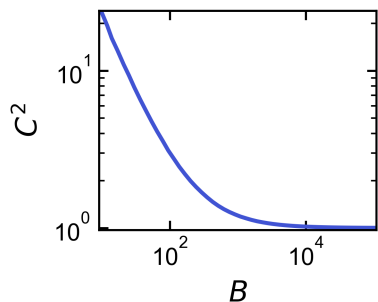


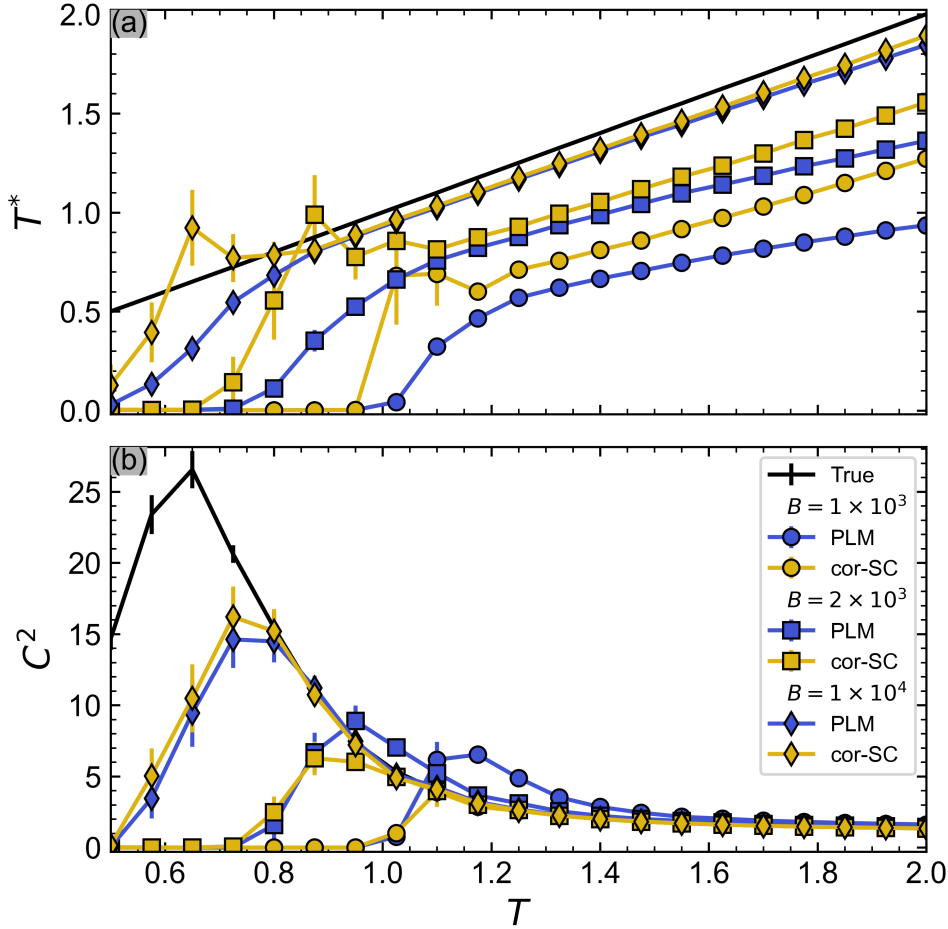
Figure 6.1: Dependence of the critical fluctuation measure  $C^2$  on the sample size  $B$  for randomly generated samples.

## 6.2.3 Correction 3: Jackknife Resampling

A standard approach to remove the first-order bias of maximum likelihood estimates is so-called Jackknife re-sampling. In a similar manner to our sub-sampling analysis, the Jackknife correction requires PLM to be performed on sub-samples of the full dataset  $B$ . Specifically, we will choose sub-samples according to a “leave-one-out” scheme, so that each sub-sample  $B^{\setminus k}$  contains all except for the  $k^{\text{th}}$  sample of the data. We then perform PLM on all  $B$  Jackknife sub-samples (each of which is of length  $B - 1$ ) and indicate the PLM each for each of these sub-samples by  $\theta_{ij}^{*\setminus k}$ . The Jackknife correction is then given by

$$\theta_{ij}^{\text{jack}} = B\theta_{ij}^* - \frac{B-1}{B} \sum_{k=1}^B \theta_{ij}^{*\setminus k}, \quad (6.5)$$

and is known to remove the bias to  $O(B^{-2})$  [215]. Jackknife re-sampling is computationally costly, as it requires performing the inference  $B$  times



**Figure 6.2:** The inferred temperature  $T^*$  (panel a) and the susceptibility measure  $C^2$  (panel b) at different input temperatures  $T$  for three data quality conditions;  $B = 1 \times 10^3$  (circles),  $B = 2 \times 10^3$  (squares),  $B = 1 \times 10^4$  (diamonds). Blue lines show observables for the PLM inference, and orange lines show observables after performing the self-consistency (SC) correction. Black lines are the temperature and susceptibility of the input models at each state point.  $N = 200$ ,  $\mu = 0.1$ , with all plotted points corresponding to means and standard errors over 21 generate-simulate-inference (GSI) runs.

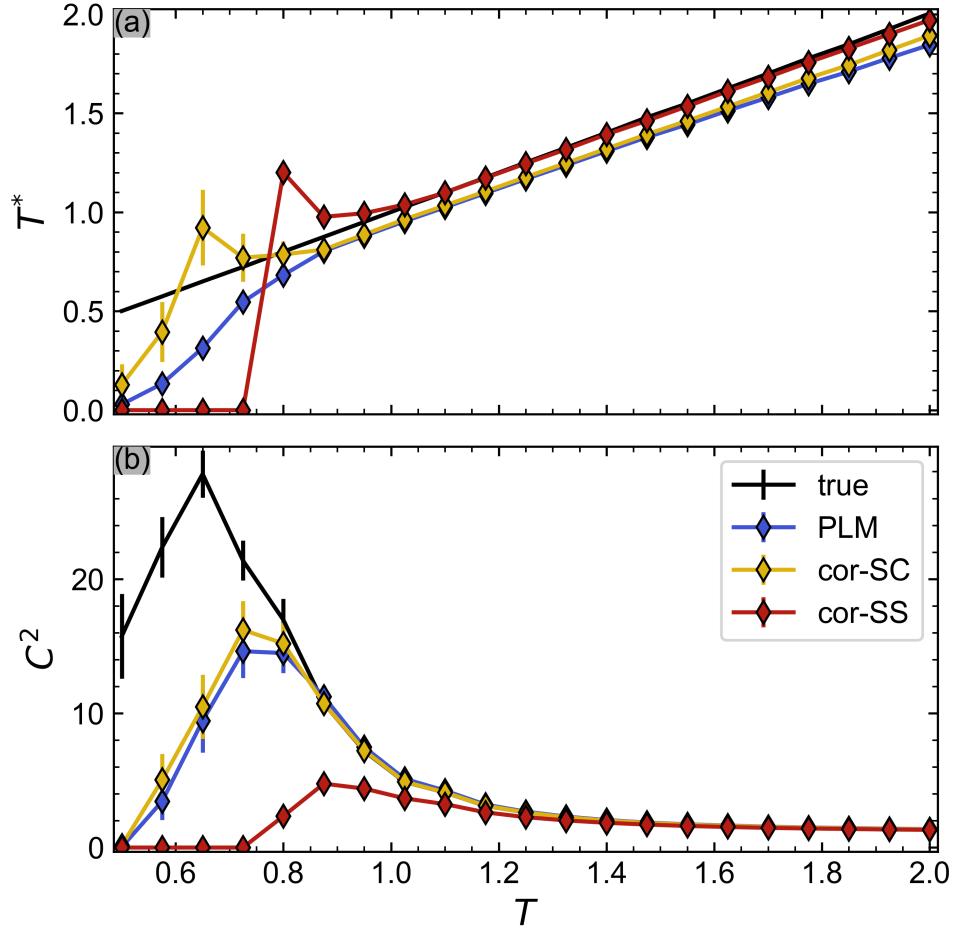
for samples of almost identical length ( $B - 1$ ), and therefore scales approximately as  $\mathcal{O}(B^2)$ . We can see how costly this becomes by looking at the data used in section 5.2 to introduce the inference of the SK model. We kept the system sizes relatively small,  $N = 120$ , and generated reasonably large  $B = 6 \times 10^4$  samples for each system. Inferring each of these models took approximately 5 seconds. Performing Jackknife for this dataset would therefore take around  $5 \times 6 \times 10^4 = 300,000$  seconds, or 83 hours. The inference time per model would increase further with increasing  $N$ . The Jackknife correction is therefore limited to small  $B$  and  $N$ . One interesting feature of jackknife resampling is that each parameter is corrected individually, something we will find relevant later. Note that jackknife resampling is largely included as a reference method through which to understand the performance of our temperature-correcting methods.

#### 6.2.4 Comparing Sub-Sampling and Self-Consistency Corrections

We will now compare the outputs from each of our explicit corrections. We do this as we ultimately wish to pick the best-performing explicit method, and then compare this to the implicit correction we introduce in the next section. As discussed, we assess each method on its ability to correctly estimate  $T$  and  $C^2$ . We will again look at the  $N = 200$   $\mu = 0.1$  SK



**Figure 6.3:** Comparison of the inferred temperature  $T^*$  (panel a) and the susceptibility measure  $C^2$  (panel b) at different input temperatures  $T$  for the self-consistency (SC) and the sub-sampling (SS) bias corrections we propose. Data are plotted for input datasets of size  $B = 1 \times 10^4$ . Blue lines show observables for the PLM inference, orange lines show observables after performing the SC correction and red lines those for the SS correction. Black lines are the temperature and susceptibility of the input models at each state point.  $N = 200$ ,  $\mu = 0.1$ , with all plotted points again corresponding to means and standard errors over 21 GSI runs.



model studied in Fig. 5.10, for which we generated 21 model realisations across a range of temperatures  $T \in [0.5, 2.0]$ . We do so to understand how each correction performs for different state points. To calculate  $C^2$  we will re-simulate the corrected models  $\theta^{\text{ss}}$  and  $\theta^{\text{sc}}$ , collecting  $B = 10^4$  samples for each state-point. The MC parameters for these simulations were  $p_{mc} = (10^5, 10^5, 10)$ , and we generated data from 6 independent MC chains for each realisation, averaging over chains and then realisations to obtain our final estimates of  $C^2$ .

We begin with the self-consistency correction, for which results are shown for three data conditions,  $B = 10^3$  (poor)  $B = 2 \times 10^3$  (intermediate), and  $B = 10^4$  (good), in Fig. 6.2. The correction significantly improves the reconstructed temperature and, by design, perfectly matches the fluctuations of  $C^2$  when separation does not occur. The improvement to the  $T^*$  prediction is particularly pronounced for small datasets and at high  $T$ . At low  $T$ , where separation occurs, the corrective optimization fails to converge and  $C^2_{\text{input}} \neq C^2(T_f^{\text{sc}})$ . This highlights a pitfall of explicit methods; they inherit any instabilities of the original MLE, such as those that lead to separation in logistic regression [211, 212]. We note that the improvement itself can also be used as a score on the reliability of the PLM estimates and as an indication of the necessity of more data, with  $T_f \rightarrow 1$  as  $B \rightarrow \infty$ .

The sub-sampling correction cannot be performed for poor ( $B = 10^3$ ) and intermediate ( $B = 2 \times 10^3$ ) sized datasets, as we require samples from

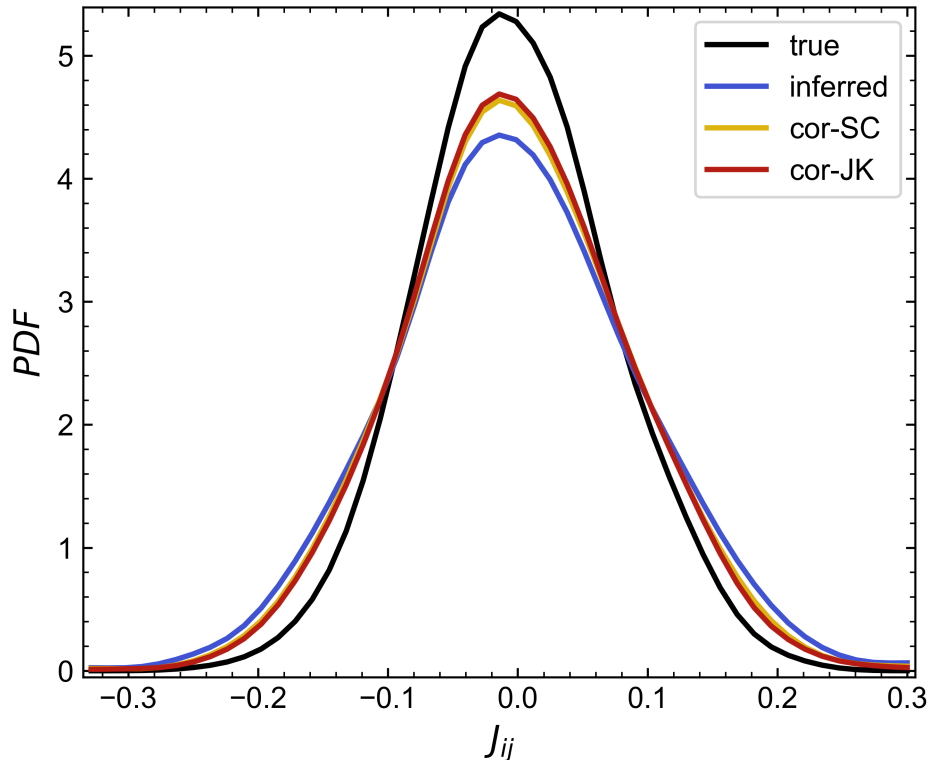
the asymptotic linear  $1/B$  regime to estimate  $T^{B \rightarrow \infty}$  accurately. We can therefore only apply it to sufficiently large datasets, in which we expect raw PLM to have given us a reasonable estimate of the state point. This method might therefore be viewed as somewhat redundant. However, as shown by our previous results, one should always perform a sub-sampling analysis of the data when claiming criticality, and so usually obtains an estimate for  $T^{B \rightarrow \infty}$  “for-free”. We compare PLM, the SC correction and the SS correction to the true values of  $T$  and  $C^2$  for the good data condition ( $B = 10^4$ ) in Fig. 6.3. We find that the SS estimates of  $T$  provide a marked improvement over both PLM and the SC correction, especially at high  $T$ . At low  $T$  we are unable to accurately estimate  $T^{B \rightarrow \infty}$  (see Fig. 5.15) from a dataset of this size, leading to the overshooting observed for  $T = 0.8$  to  $T = 1.0$ . Interestingly, although the temperatures are reproduced perfectly for  $T > 1$ , we find that the corresponding values of  $C^2$  are under-predicted. This is unsurprising, as increasing the temperature past  $T^{\text{sc}}$  should naturally lead to decreased fluctuations. The SS correction has shown us that a simple multiplicative correction (the fictitious temperature  $T_f$ ) cannot fully account for the PLM bias and that there is a trade-off between capturing the temperature and the correlations of the model. We will explore this final point further when discussing the Jackknife correction, as this does indeed provide individualistic corrections to each  $\theta_{ij}$ .

We now make some generic comments regarding the applicability and suitability of these corrections. In particular, we note that the SS correction can only be applied when  $B$  is already large, i.e. when there is less need for a correction to begin with. The SC correction on the other hand was shown to provide significantly better estimates of the temperature, especially so when  $B$  was very small ( $B = 10^3 \approx \tilde{B}$ ). It is therefore highly applicable in poor data quality settings, where corrective methods are most needed. Further, the computational complexity of the SC correction is set by the length of the MC simulation at each optimisation step (i.e. in estimating  $C_{MC}^2$ ). As we set  $B_{MC} = B_{\text{input}}$ , we thus also see that simulation times are kept small when  $B$  is small. SC corrections are thus best when  $B$  is small, while the SS correction can still provide moderate improvements when  $B$  is large. For the specific purpose of inferring criticality, however, we argue that the SC correction is much more useful, as it produces models which exhibit correlations that correspond to those in the data.

### 6.2.5 Relating the Self-Consistency Correction to the First-Order Bias

Both SC and SS corrections shift the state point by multiplying the PLM couplings by a constant,  $1/T_f$ . These do not strictly remove the bias on each parameter, but instead remove the bias of the temperature, a global property. We will now discuss the jackknife correction, which explicitly removes the  $\mathcal{O}(B^{-1})$  bias from each PLM estimate  $\theta_{ij}^*$ . We note that the computational intensity of the Jackknife method means that it cannot be used (or at least is inconvenient to use) for the system sizes and sample numbers considered in Fig. 6.2 and Fig. 6.3. We will instead look at a small system,  $N = 50$ , for which PLM requires fewer data (remembering that  $b_1 \sim N$ ), and explore how the SC correction relates to the individual shifts in  $\theta_{ij}^*$ .

**Figure 6.4:** Comparing coupling probability density functions for data generated from a single model realisation at  $\mu = 0$  and  $T = 2.0$ . The black curve shows the true (i.e. input) coupling distribution from which we produced  $B = 750$  samples via MC simulation. The system size was small ( $N = 50$ ) so that the jackknife correction could be performed. The blue line shows the PLM estimate inferred from this dataset, the orange line is the self-consistency (SC) corrected parameter distribution and the red line is the jackknife (JK) corrected parameter distribution. All distributions have been smoothed with a Gaussian kernel to ease visualisation.

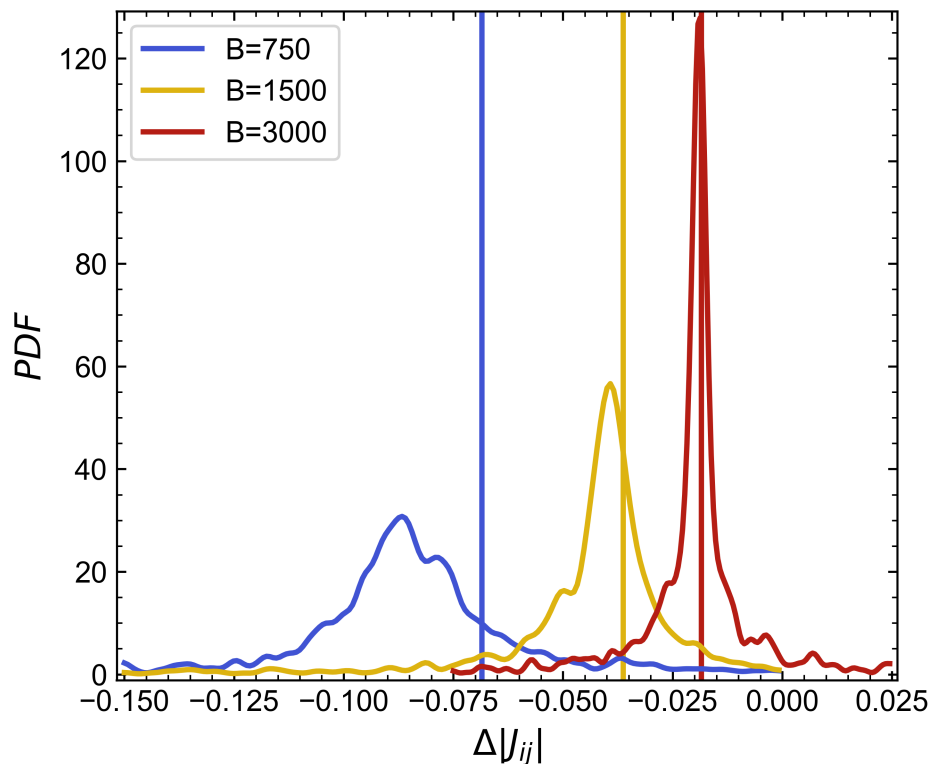


that the Jackknife correction provides. We will choose a high temperature  $T = 2.0$  state-point as we know both PLM and the SC correction will perform well here. We will generate  $B = 750$ ,  $B = 1500$  and  $B = 3000$  samples from this state-point (representing poor, average and good data) and analyse the performance of the corrections for each of these. We equilibrate each system for  $t_{eq} = 10^5$  MC cycles and sample with  $t_s = 10$ . The primary aim of this investigation is to understand how the true bias corrections produced by Jackknife re-sampling align with the global bias correction performed in the SC correction.

We show the coupling distribution for the input model, the PLM model and the SC and Jackknife corrections in Fig. 6.4 for  $B = 750$ . We choose the smallest  $B$  as the effects of the bias are most pronounced. As before, we find that PLM overestimates the spread of the couplings. Both SC and jackknife corrections narrow the coupling distribution towards the true distribution (i.e. increase the temperature), and we see little variation between the two. This would imply that the bias on each parameter  $b_{1,ij}$  is well described by the simple global re-scaling we perform in the SC correction. To investigate this further we assess how each correction shifts the individual couplings  $J_{ij}$  by defining

$$\Delta|J_{ij}| = \frac{|J_{ij}^{\text{cor}}| - |J_{ij}^*|}{|J_{ij}^*|}, \quad (6.6)$$

where  $J_{ij}^{\text{cor}}$  can stand for either Jackknife or SC correction. For the SC correction  $\Delta|J_{ij}| = \frac{1}{T_f^{\text{sc}}} - 1 = \text{const}$  for all  $J_{ij}$ . But Jackknife re-sampling will lead to a distribution of shifts. We plot  $\Delta|J_{ij}|$  in Fig. 6.5 for our 3 data quality conditions. Vertical lines indicate the constant  $\Delta|J_{ij}|$  shifts

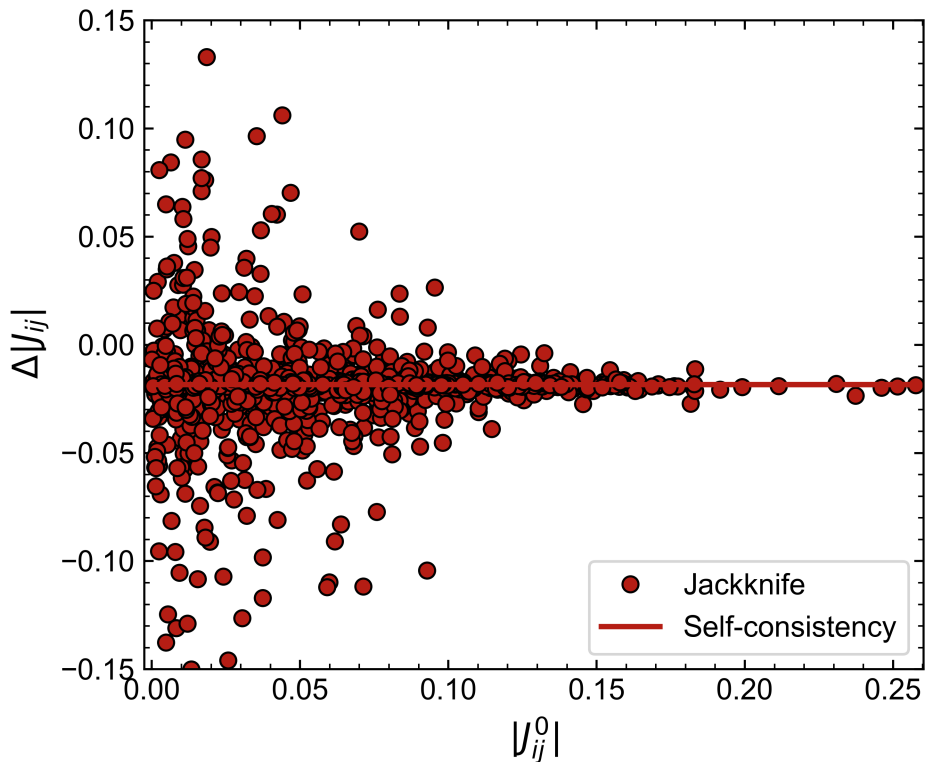


**Figure 6.5:** Tracking how individual parameter shifts induced by the jackknife correction compare to the constant parameter re-scaling of the self-consistency (SC) correction as we increase the number of samples in the input dataset. Colour represents different sample sizes  $B$ . Probability density functions are shown for the jackknife correction. The constant re-scaling of the SC correction is shown by corresponding vertical lines. The Jackknife corrections are sharply peaked distributions, that align with the position of the SC correction as  $B$  increases.

that result from the SC correction. We note that all of these are negative, as the SC correction reduces the size of every coupling, and tend to 0 as  $B \rightarrow \infty$ . For the Jackknife correction,  $\Delta|J_{ij}|$  follow distributions which are also peaked at negative values of  $\Delta|J_{ij}|$ . The overall effect of Jackknife re-sampling is thus also to reduce the parameter size. As  $B$  increases these peaks narrow and increase in height. More interestingly, the positions of the jackknife peaks and the SC correction appear to coincide as  $B$  increases. This suggests that in the limit of large  $B$ ,  $J_{ij}^{\text{jack}} = \delta(J_{ij}^{\text{jack}} - J_{ij}^{\text{sc}})$ . This shows that the self-consistency correction we propose is approximately equal to explicitly removing the first-order bias.

To establish this more clearly, we plot the parameter shifts as a function of  $|J_{ij}^0|$  in Fig. 6.6 for  $B = 3000$  where the best agreement between Jackknife and SC corrections is seen. We observe that shifts away from  $\Delta|J_{ij}|^{\text{sc}}$  largely affect small couplings. Indeed shifts for the largest parameters coincide closely with the line  $\Delta|J_{ij}|^{\text{sc}}$ . The tails of the  $|J_{ij}|^{\text{jack}}$  distribution therefore correspond to small input couplings  $|J_{ij}^0|$ , i.e. to weak couplings. Large percentage shifts in these parameters have little effect on the model, and the overall effect of the jackknife correction is well described by a constant shift (reduction) of the most important parameters (the strong couplings). We thus claim that our proposed self-consistency correction is approximately equivalent to removing the first-order bias of the inference. Interestingly, this means that constraining the fluctuations of the PLM model to those of the input data removes the first-order bias. Having established this, we will now compare our self-consistency correction to a standard method with which to implicitly remove the first-order bias via a penalisation to the log-likelihood.

**Figure 6.6:** The parameter shift  $\Delta|J_{ij}|$  for each correction as a function of the size of the input parameter  $|J_{ij}^0|$  for  $B = 3000$ . On average, parameters which are shifted by the largest amount in the jackknife correction correspond to small (unimportant) input couplings. Jackknife corrections to the largest parameters align well with the global parameter shift implemented by the self-consistency correction.



### 6.3 Implicit Correction

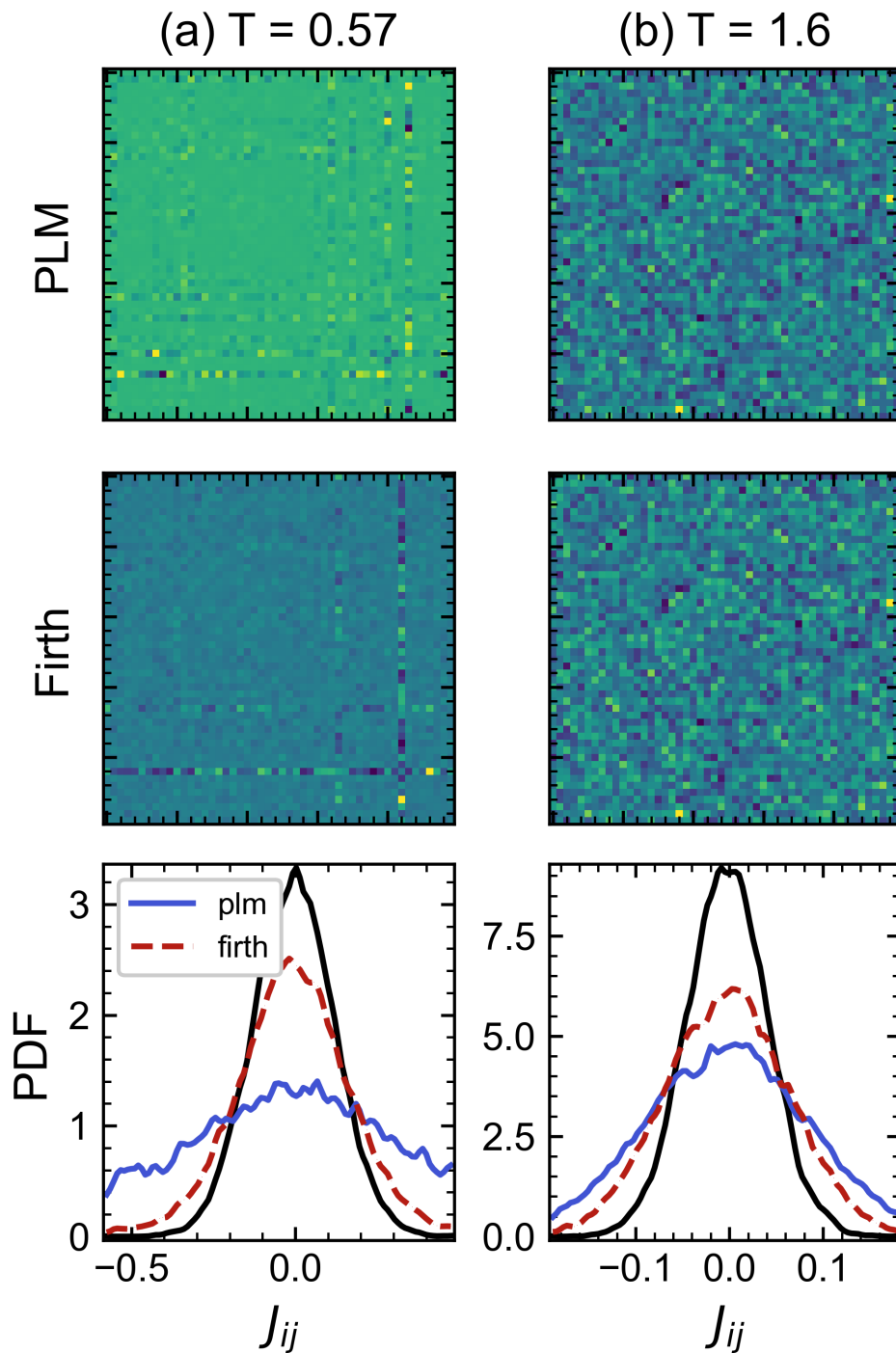
Implicit bias corrections remove the first-order bias *during* the inference [211], for instance by penalizing the log-likelihood [37, 212, 213, 216, 217]. We previously highlighted a first-order bias effect termed separation [208] as the reason behind the failure of PLM at low temperatures. Implicitly removing this effect may therefore allow us to infer models at lower temperatures than previously possible. We will only explore a single implicit bias correction, which we describe below.

#### 6.3.1 Firth’s Penalized Logistic Regression

Firth [37] introduced an implicit correction through which to remove the first-order bias of maximum likelihood estimates. This is achieved by adding a penalty term to the log-likelihood function and then maximising this penalized likelihood. In the context of PLM, Firth’s correction corresponds to maximising the penalized log-likelihood  $\mathcal{L}'_r$ :

$$\mathcal{L}'_r(h_r, \mathbf{J}_r | \{\mathbf{s}\}_B) = \mathcal{L}_r(h_r, \mathbf{J}_r | \{\mathbf{s}\}_B) + 0.5 \ln |F(h_r, \mathbf{J}_r)|, \quad (6.7)$$

where  $|F(h_r, \mathbf{J}_r)|$  is the determinant of the Fisher information matrix for each row of parameters  $r$ . When applied to logistic regression,  $F(h_r, \mathbf{J}_r)$  takes a particularly simple form (see [212, 213]), allowing the penalised log-likelihood to be maximised using standard methods. We implement this computationally by modifying the Python code available at <https://github.com/jzluo/firthlogist>. The corrective term of eq. 6.7 tends to 0 as  $B \rightarrow \infty$ , returning the un-penalised likelihood. For small  $B$  the penalty compensates the  $\mathcal{O}(B^{-1})$  bias and is known to control separation in logistic



**Figure 6.7:** Parameter matrices  $\theta$  for a subset of 50 spins for un-penalized PLM (top) and Firth's penalized PLM (middle) at two different temperatures. (Bottom) probability density functions (PDFs) of the corresponding inferred parameters along with the true parameter PDF (black line). Firth's correction controls the inference at low  $T$ , leading to finite parameters and a PDF that more closely matches the true input model. At high  $T$ , Firth's correction shifts the inferred temperature (by reducing the spread of the parameter PDF).  $B = 10^3$ ,  $N = 200$  and  $\mu = 0.1$ .

regression [212, 213]. We thus expect that this penalized regression might allow us to obtain reasonable parameter estimates even for data generated from temperatures where separation occurs when un-penalized PLM is used.

In Fig. 6.7 we demonstrate the effect of the penalty on the inferred parameters at a low  $T = 0.57$  and a high  $T = 1.6$  state point. At low  $T$  separation causes the un-penalized PLM parameters associated with specific  $s_i$  to diverge, with  $\max(\theta_{ij}^{\text{PLM}}) \approx 400$ , leading to a non-Gaussian highly spread PDF. In our language, models with these very strong couplings correspond to zero temperature. Firth’s penalty controls this effect, and although we still see large parameter estimates for the same  $s_i$ , these are orders of magnitude smaller with  $\max(\theta_{ij}^{\text{Firth}}) \approx 4$ . Firth’s correction reduces the spread of the inferred distribution and largely captures the Gaussian nature of the input coupling distribution, even at low  $T$ . It, therefore, provides better  $T^*$  estimates than un-penalized PLM.

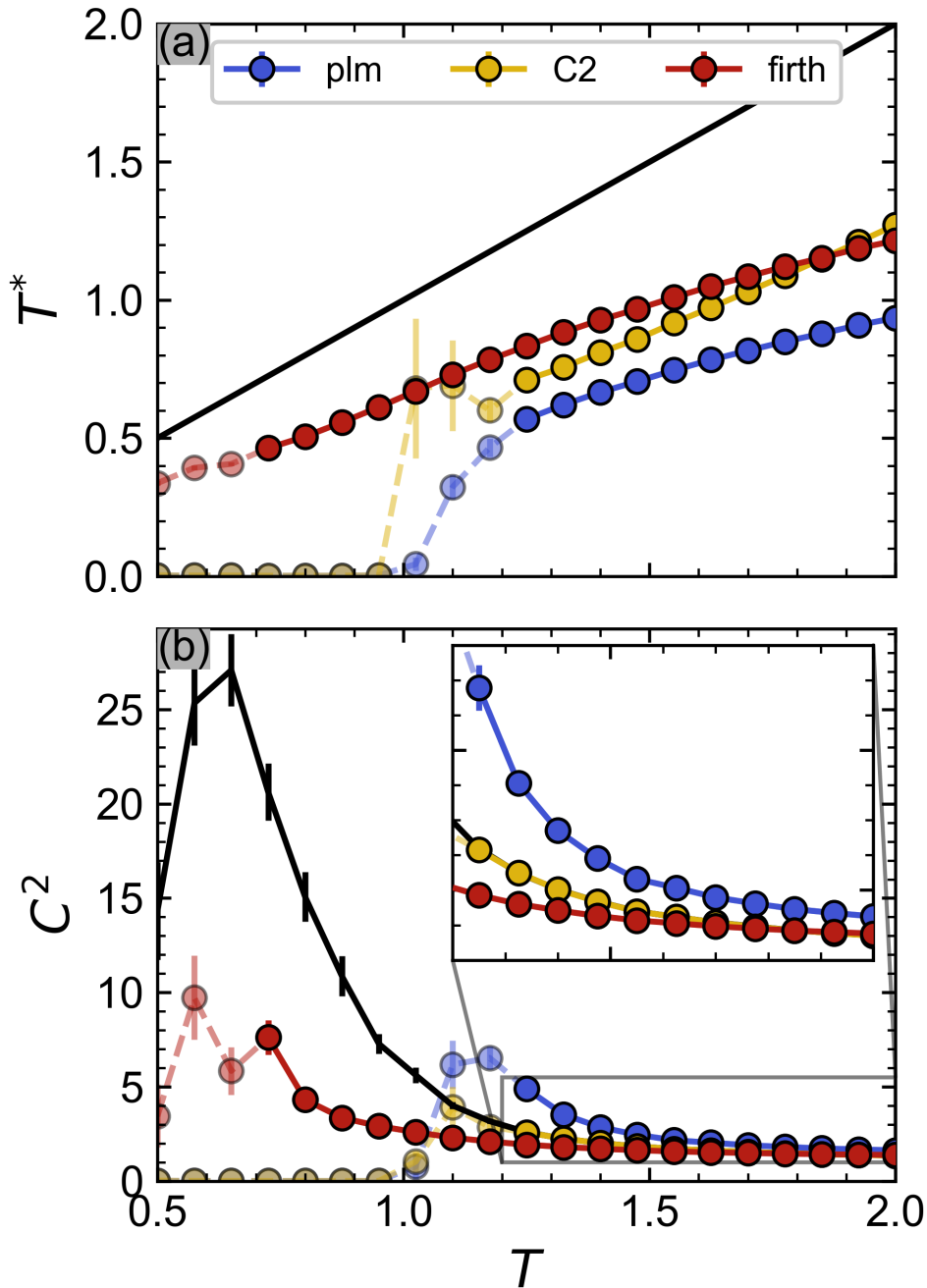
## 6.4 Comparing the Self-Consistency Correction to Firth’s Penalized Regression

In Fig. 6.8 we compare the performance of the self-consistency (SC) correction and Firth’s (firth) correction for the smallest dataset,  $B = 1000$ , where bias effects are most extreme. We again generate data from 21 independent model realisations for each  $T$  and then apply each inference method to each dataset separately. We naively assess if separation has occurred by checking if  $|\theta^*|_{\max} > \mu_\theta^* + 10\sigma_\theta^*$ , where  $|\theta^*|$  is the absolute value of the inferred parameters from each inference scheme. Any input  $T$  where a *single* inferred model satisfied the separation condition (i.e. had “anomalously” large parameters) is indicated by the transparent points in Fig. 6.8. We note that this is a conservative definition; if even a single inference fails we characterise the whole state point as separation prone. We see that (according to our definition) the onset of separation is delayed to much lower  $T$  when using Firth’s penalty, and that Firth’s penalised logistic regression predicts non-zero  $T^*$  for all  $T$ . At high  $T$  both corrections perform similarly well in improving the estimated  $T^*$ , although the dependence  $T^*(T)$  appears to scale more favourably using the SC correction as  $T$  increases.

Fig. 6.8(b) shows how well each method reproduces the correlations of the input data. We observe an interesting trade-off; although Firth’s correction provides better estimates for  $T$ ,  $C^2$  of the corresponding models is systematically under-predicted. Firth’s correction thus fails to capture the correlations of the data. We note that this contrasts with the un-penalised logistic regression results for PLM, which instead over-estimate  $C^2$ . The main advantage of Firth’s correction over the SC correction, therefore, appears that lower temperature state-points can be estimated.

This naturally raises a question: can we apply the SC correction to the models inferred using Firth’s penalization and improve the estimate of both  $T$  and  $C^2$ ? The answer is negative: unsurprisingly, applying the  $C^2$  correction to the penalized parameter estimates increases  $C^2$  at the cost of lowering  $T^*$ , and ultimately leads to the same estimates as simply applying the SC correction to the un-penalized PLM model. We summarise these findings in Table 6.1. There thus appears to be an inherent trade-off





**Figure 6.8:** Comparison of the corrective methods' ability to capture the relevant observables  $T^*$  and  $C^2$  for a small sample size  $B = 10^3$ . Self-consistency correction is labelled by C2 in the legend. Transparent points indicate  $T$  where separation occurred. Firth's correction gives reasonable estimates for  $T^*$  at much lower  $T$  than PLM, highlighting the ability of this method to control separation. In contrast to normal PLM, Firth's correction under-estimates  $C^2$  for all  $T$ . At high  $T$  Firth's correction and the self-consistency correction perform similarly. Each point represents an average over 21 independent datasets at that  $T$ .



**Table 6.1:** Percentage errors of the inferred estimates for  $T$  and  $C^2$  for a single model realisation at  $T = 1.025$ ,  $\mu = 0.1$ , with  $B = 10^4$ , using a range of inference schemes. Firth’s correction provides the best estimate of the temperature but the worst estimate of the critical fluctuations. We demonstrate an implicit trade-off between correctly inferring  $T$  and  $C^2$ . Applying the C2 correction either to the PLM model or to the Firth corrected model produces the same  $T^*$  and  $C^2$  pair. The temperature is underestimated, irrespective of the inference scheme used.

Method	$T$ % error	$C^2$ % error
PLM	-6.9	3.4(8)
PLM $\rightarrow$ C2	-6.0	0.2(9)
Firth	-4.0	-7.7(5)
Firth $\rightarrow$ C2	-6.1	-0.1(8)

between capturing the temperature and the correlations of a dataset. When deciding which method is “best” one must therefore decide which property is most important to encode correctly.

#### 6.4.1 Implications for inference around criticality

So far we have shown that small-sample biases influence the determination of the state of Ising models inferred using PLM. The problem is that what constitutes a “small” sample size itself depends on the state point (and topology) of the true model that generated the data. Studies claiming criticality in Ising models inferred using PLM thus need to control for bias, for example through sub-sampling their data and performing a similar analysis as in Fig. 5.23. They should also consider that the PLM model they infer from any dataset which is dynamic (i.e. fluctuates) will be biased *towards* the critical point and exhibits enhanced critical fluctuations when simulated. In such cases, the SC correction may be applied to re-scale the couplings and match the empirical correlations. This will also shift the inferred temperature towards the true temperature. We note, however, that the SC corrected temperature remains systematically smaller than the true temperature, and should only be considered as a lower-bound estimate of the true temperature. It may be appropriate to use Firth’s implicitly corrected penalized logistic regression if separation is found to occur. This correction will provide reasonable parameter estimates even for low  $T$  state points, but it should be noted the fluctuations displayed by these models are not representative of the data, which is important when considering near-critical phenomena.

### 6.5 Conclusion

In this chapter, we described data-driven approaches to mitigate the effect of the first-order PLM bias. The bias may be removed explicitly (i.e. corrected after inference) or implicitly (corrected during the inference). We explore three explicit methods and one implicit method. We propose a novel explicit bias correction which we term the *self-consistency* (SC) correction,

which adjusts the temperature of the inferred model so that the susceptibility measure  $C^2$  of the inferred model matches  $C^2$  of the input dataset. We show that the parameter rescaling of the SC correction is approximately the same as performing a first-order bias correction by contrasting our method with results from the standard jackknife bias correction procedure. We also investigate an implicit bias correction in the form of Firth’s penalized logistic regression. We find that Firth’s correction produces better temperature estimates and worse susceptibility estimates than the self-consistency correction. This highlights an inherent trade-off between capturing the temperature and the susceptibility of the model when the amount of data is finite. Firth’s correction allows models to be inferred from datasets generated at low temperatures, where standard PLM fails due to separation. Our self-consistency correction provides large improvements to the temperature estimates when the sample size is small and allows us to place a lower bound on the true temperature of a dataset. This will enable us to appraise whether models inferred from unknown datasets that appear close to critical truly represent a near-critical state. The next chapter seeks to explore this in a case study of typical neuroimaging data.

## Summary of Chapter 6

- We explored corrections to remove the first-order bias of pseudo-likelihood maximisation (PLM) parameter estimates, having previously identified that these cause PLM to underpredict the temperature. We assess each corrective method based on its ability to reproduce the temperature  $T$  and susceptibility  $C^2$  of the input model.
- Bias corrections may be split into two categories:

### **Explicit Corrections**

Explicit corrections are applied *after* an inference has been performed. They modify the existing parameter estimates.

### **Implicit Corrections**

Implicit corrections are applied *during* the inference process. They modify the likelihood functions, e.g. by introducing a penalty term, and produce unbiased parameter estimates.

- We introduce two novel explicit bias corrections and compare these to the standard *jackknife resampling* method. Our corrections are:

### **Sub-sampling to improve the temperature estimate**

We can sub-sample data to obtain an unbiased estimate of the temperature. This correction under-predicts the susceptibility  $C^2$ . This is only possible when large quantities of data are already available.

### **Enforcing self-consistency to improve the temperature estimate**

We require that the susceptibility of our inferred model is equal to the susceptibility of the input dataset. We enforce this self-consistency condition via a second optimisation. This shifts the temperature of the inferred model closer to the true temperature and perfectly reproduces the susceptibility so long as separation does not occur.

- We also implement an implicit bias correction, in the form of *Firth's penalized logistic regression*. Firth's correction provides better temperature estimates than the self-consistency correction but underpredicts the susceptibility. Firth's correction enables the approximate inference of state points where separation occurs for unmodified PLM.
- The self-consistency correction provides the best improvement to the temperature and susceptibility estimates when the sample size is small and the data is paramagnetic.

## Chapter 7

# Mindfulness Meditation: a Small Sample Size Study

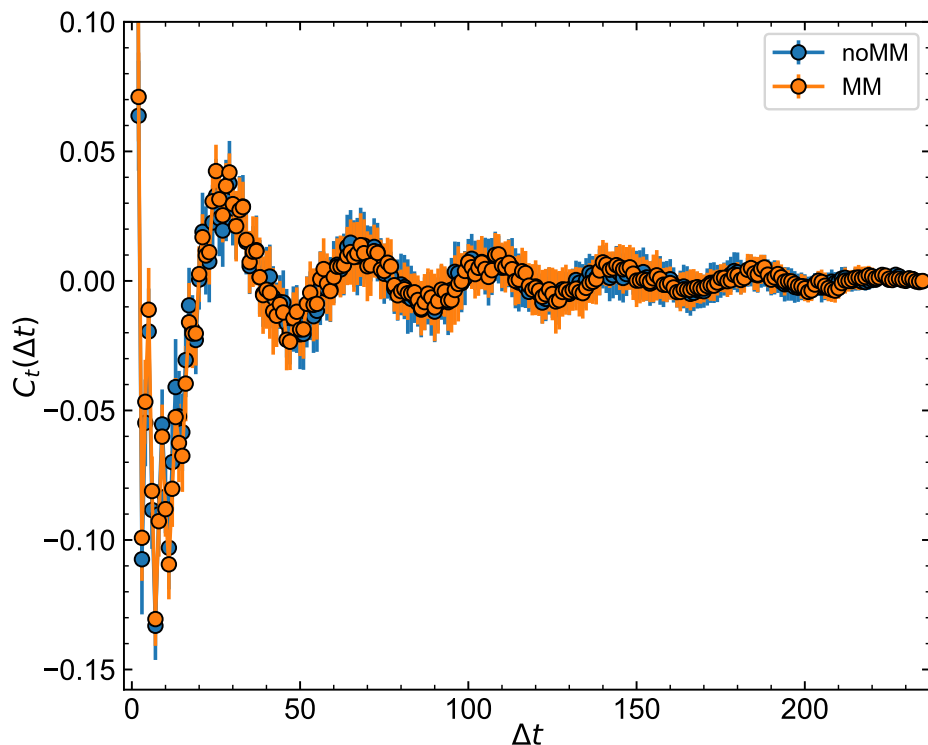
In this chapter, we explore the repercussions of the small sample bias of pseudo-likelihood maximisation in interpreting inference results from real data. We consider a functional magnetic resonance imaging (fMRI) dataset of brain activity related to meditation as a case study. Although operating at criticality is a tempting hypothesis (see chapter 4), our results lead us to caution against claims of criticality in PLM models inferred from typical fMRI data. We show that small sample size biases strongly affect these inferences, and tune models inferred from dynamical (i.e. fluctuating) data to a closer-to-critical state.

### 7.1 The Mindfulness Meditation Dataset

Mindfulness meditation (MM) is a meditation practice where one seeks to actively concentrate on the present moment and is thought to be a low-cost clinical intervention which may help with the treatment of several mental health disorders [220]. fMRI may be used to assess whether practising MM actively changes the functional connectivity (FC) of the brain [133]. We have been provided with the dataset from Ref. [133] and will use this as a case study to investigate how strongly the small sample size bias of PLM affects inferences from typical fMRI datasets.

The key biological question underpinning the study of Ref. [133] was to understand if mindfulness meditation changes the resting-state functional connectivity of an individual. As such, imaging sessions were carried out on separate days under two different conditions: those where the participant practised mindfulness meditation (MM) before undergoing imaging, and those where they did not (noMM). During each session,  $B = 236$  samples were collected from  $N = 399$  grouped regions of interest (ROIs) within the brain. We will consider each ROI as a spin  $s_i$  and perform inverse Ising inference using PLM. The trajectories for each ROI,  $s_i(t)$ , are continuous. We thus binarised the data by removing the average from the signal and setting  $s_i(t) < 0 = -1$  and any  $s_i(t) \geq 0 = +1$ . ROIs 0 to 359 originate from the cerebral cortex (i.e. the outer layer of the brain) while ROIs 360 to 399 are from limbic and cerebellar regions of the brain. In total  $B_{\text{noMM}} = 40 \times 236 = 9440$  samples were collected for the noMM condition

**Figure 7.1:** The normalised auto-correlation  $C_t$  as a function of the delay time  $\Delta t$  for both noMM and MM conditions after binarizing the data. Although there are clearly non-trivial oscillatory dynamics present, we note that the size of these correlations is small (c.f. y-axis limits). We will consider each time point as an independent and identically distributed (i.i.d) sample for the purposes of our statistical analysis.

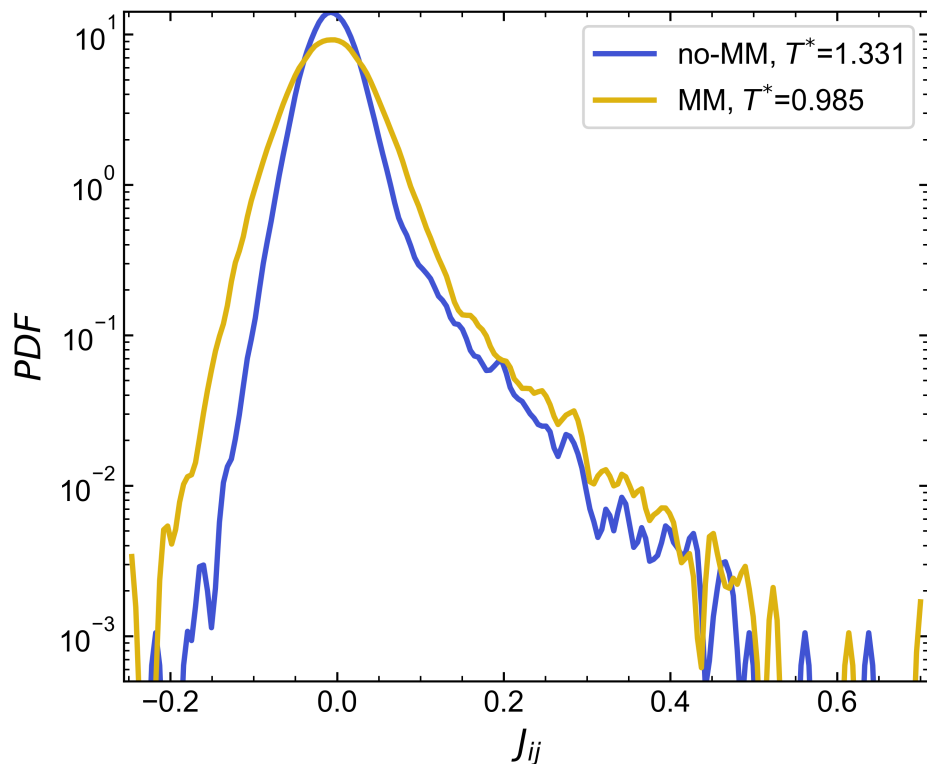


and  $B_{\text{MM}} = 18 \times 236 = 4248$  for the MM condition.

To link the biological question of this study with criticality, we investigate whether PLM inference leads to the identification of a close-to-critical state and if this state is affected by the biological condition (i.e. practising of mindfulness meditation). Moreover, we study if the inference biases identified in the last chapter significantly impact our conclusions. This investigation shows that the small sample bias can severely skew the conclusions of PLM analyses towards supporting criticality. Before moving on to the inference results, we briefly plot the auto-correlation functions for both conditions in Fig. 7.1. We do this as we require the data to be “uncorrelated” for the independent and identically distributed (i.i.d) assumption of PLM to hold. We see that the auto-correlation  $C_t$  rapidly decays to 0 with the delay time  $\Delta t$ , so that  $\tau < 1$  for this dataset. We, therefore, consider our trajectories as “decorrelated” samples from this perspective. This does not mean there are no features in the dynamics: indeed we see small oscillations in  $C_t$ , hinting that (of course) dynamics in the brain are governed by much more complex processes than the simple equilibrium fluctuations we modelled when generating Monte Carlo samples in the last two chapters.

## 7.2 Small Sample Bias in Real Data

From the perspective of criticality, we have two key questions for our analysis. Firstly, does mindfulness meditation alter the state-point of the brain? Secondly, is the resting state brain close to criticality (i.e. does criticality facilitate normal function)? We now detail our attempt to address these.



**Figure 7.2:** Probability density functions (PDFs) for the inferred couplings  $J_{ij}$  of the noMM (blue) and MM (orange) conditions. The MM condition is more spread (i.e. has a larger standard deviation). We use the standard deviation to calculate an inherent temperature  $T^*$  for the model based on our previous investigation of PLM in the Sherrington-Kirkpatrick (SK) model. Although noMM and MM conditions initially appear to correspond to different state points, we later show that this is purely the result of the small-sample bias of PLM.

### 7.2.1 Impact on State-Point Classification

The distributions of the PLM parameters obtained from the *full* datasets are shown in Fig. 7.2. First, as opposed to the SK model, the distributions are skewed, with a long tail at positive values of the couplings. Second, the MM condition corresponds to a larger variance in the couplings than the noMM one. The immediate consequence is that the mapped temperatures of the two full datasets are different,  $T_{\text{full-MM}}^* = 0.98$  and  $T_{\text{full-noMM}}^* = 1.33$ .

However, the two sample sizes are  $B_{\text{noMM}} \approx 2B_{\text{MM}}$ . Can this lead to a significant statistical effect in the estimation of the corresponding state points? Fig. 7.3 provides an affirmative and quantitative answer: as we sub-sample (ss) the noMM data we find that the PLM temperature estimate decreases with reducing  $B$ , and when  $B_{\text{ss-noMM}} = B_{\text{MM}}$ , meets the same estimated temperature of the MM data. Hence, for the data considered here, there is no significant difference between the MM and the noMM data when the bias is taken into account. We further find that below some critical value,  $B_c \approx 2000$ , separation occurs leading to the failure of the inference.

To refine the estimate of the noMM state point, we can apply the self-consistency (SC) correction (yellow circles) and fit the empirical saturation function to the PLM temperature estimates. Fitting data with  $B \geq 2500$ , we find  $T^{B \rightarrow \infty} = 1.72$  and  $\tilde{B} = 3465$ . The available datasets are with  $B_{\text{MM}} \approx 1.2\tilde{B}$  and  $B_{\text{noMM}} \approx 2.7\tilde{B}$ , indicating that the small sample size bias strongly impacts our conclusions here. We validate the applicability of our empirical fit in Fig. 7.4, where the familiar linear dependence of the inverse temperature on  $B^{-1}$  confirms that our analysis of the first-order bias is appropriate.

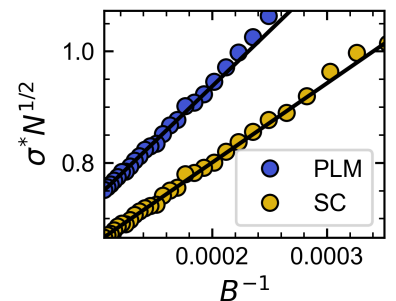
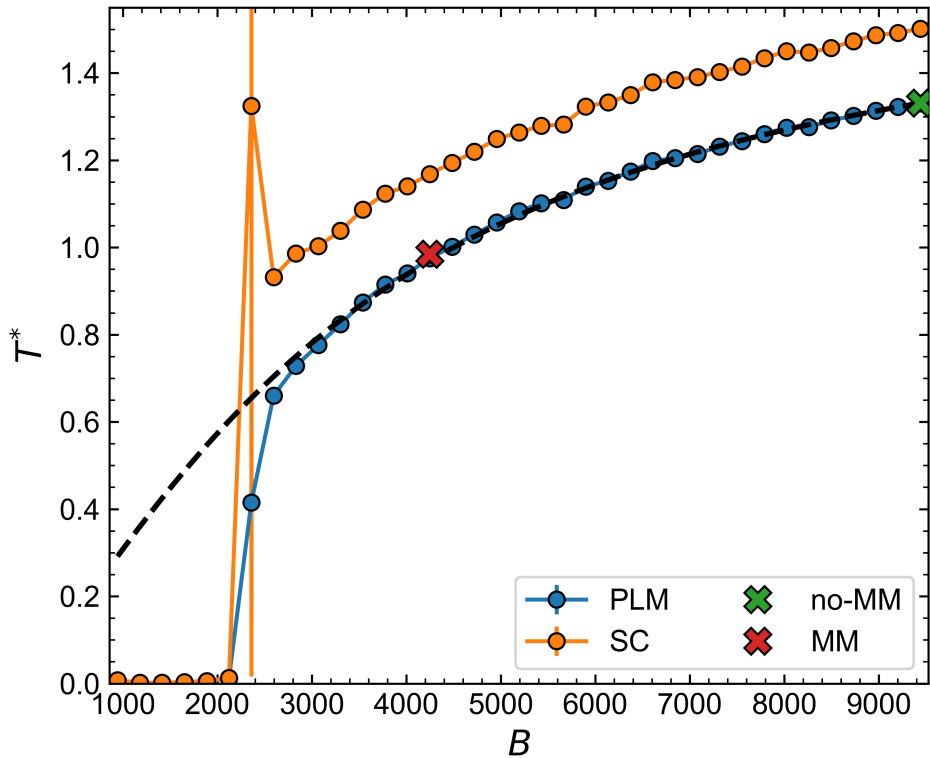


Figure 7.4: The re-scaled standard deviation for the noMM condition vs  $1/B$ .

**Figure 7.3:** Sub-sampling analysis of the larger noMM dataset. PLM temperature estimates for each number of sub-samples are shown in blue, and the corresponding temperature of the self-consistency (SC) corrected models are shown in orange. The dashed black line shows a fitted empirical arc-tan model to the PLM temperature data. Coloured crosses show the temperatures and sample numbers corresponding to the distributions in Fig. 7.2.

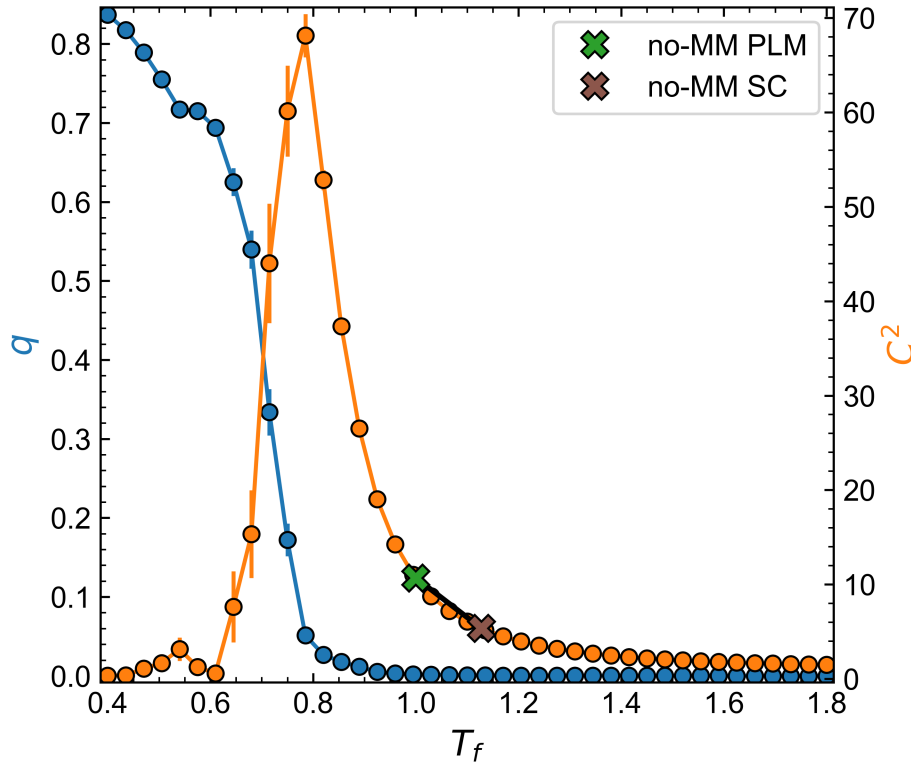


### 7.2.2 Impact on Claims of Criticality

To contextualise the meaning of the inferred temperatures in Fig. 7.3 we again introduce the fictive temperature  $T_f$  and perform MC simulations of  $\frac{1}{T_f} \theta_{\text{noMM}}^*$  for a range of  $T_f$ , see Fig. 7.5. A peak of  $C^2$  at  $T_f = 1$  would mean that the PLM model is situated exactly at the critical point. We instead find the peak at  $T_f = T_c = 0.78$  and so  $\theta_{\text{noMM}}^*$  is a paramagnetic state-point above the transition, albeit still with substantial critical fluctuations  $C^2(T_f = 1) \approx 0.15C_{\text{max}}^2$ . The self-consistency correction shifts the model further from  $T_c$ . Hence, the MM and noMM conditions appear to be at best-near critical if not paramagnetic.

### 7.2.3 Ubiquitous Criticality is a Small Sample Size Effect

Summarising these results, initial analysis of the two conditions would lead to the conclusions that a) practising mindfulness meditation changes the state-point of the brain, and b) that the noMM condition represents a near-critical paramagnetic state-point. Carefully accounting for the bias instead reveals that both datasets more likely originate from the same state-point, and the SC corrected temperature estimate shows that, as a lower bound, the true state-point of the data lies far from the transition in the paramagnetic phase. We, therefore, find no evidence to suggest that the resting state fluctuations in this imaging study correspond to a critical brain state. The results of our analysis pose serious challenges to the numerous claims of criticality in inferred Ising models highlighted in chapter 4. We have shown that small sample size effects severely bias models inferred from dynamic data towards criticality. As the brain is inherently dynamic [66], pairwise



**Figure 7.5:** Spin-glass order parameter  $q$  and susceptibility  $C^2$  as functions of the fictive temperature  $T_f$  for the model inferred from the noMM condition,  $\theta_{\text{noMM}}^*$ . The F order parameter  $m$  was also measured and found to be 0 throughout. Thus, this sweep characterises a transition from a low-temperature spin-glass to a high-temperature paramagnetic phase. Points and error bars correspond to means and standard errors of 60 independent MC simulations with  $B = 10^4$  samples;  $p_{mc} = (10^5, 10^5, 10)$ . Green and red crosses correspond to the PLM and self-consistency corrected models of the noMM condition respectively.

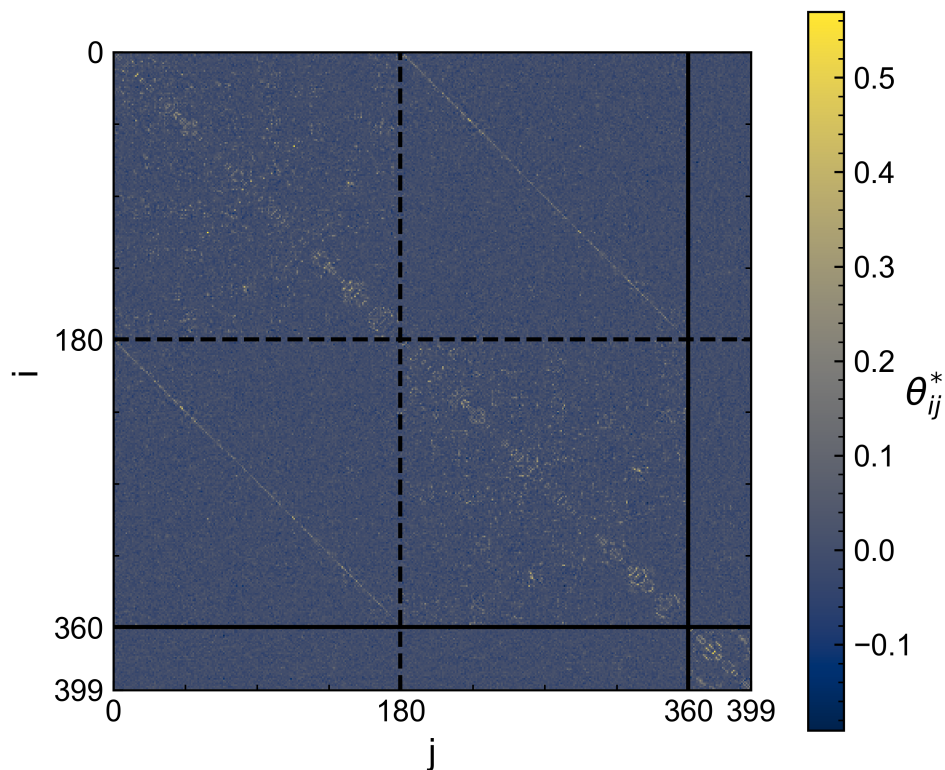
maximum entropy models inferred from brain datasets are thus always biased towards criticality. We have also shown that the severity of this bias depends on factors such as state point, coupling structure and system size (c.f.  $\tilde{B} \approx 1000$  in the SK model vs here where  $\tilde{B} \approx 3465$ ) and cannot be predicted *a priori*. The amount of data required to accurately infer the state-point of a dataset is thus also unknown, and can only be established through studying the convergence of relevant observables as functions of  $B$ . Conclusions of authors who have not properly accounted for the bias will therefore be heavily skewed towards criticality and should be regarded with a high degree of scepticism.

### 7.3 Structure and Statistical Mechanics of the noMM Resting State Network

In this section, we will investigate the behaviour of the noMM PLM model as a function of  $T_f$  in more detail. We show  $\theta_{\text{noMM}}^*$  in full in Fig. 7.6. Although we have established that  $\theta_{\text{noMM}}^*$  most likely does not fully represent the ground truth model, we can still gain interesting insights from this analysis. For instance, we see that the left (L,  $i = 0$  to  $i = 180$ ) and right (R,  $i = 180$  to  $i = 359$ ) hemispheres of the cerebral cortex are highly symmetric (as indicated by large coupling values in inter-hemisphere diagonal  $\theta_{i,j=i+180}^*$ ). This symmetry is an expected and known result as previous resting-state fMRI studies have also found highly symmetric connectivity networks across the brain hemispheres [221, 222]. We find that the L and R hemispheres are not perfectly mirrored, in line with findings from other authors who state that small asymmetries exist and support a



**Figure 7.6:** The inferred parameter matrix  $\theta_{\text{noMM}}^*$  from the noMM condition. Dashed black lines label boundaries of the left ( $i = 0$  to  $i = 179$ ) and right ( $i = 180$  to  $i = 359$ ) cortical hemispheres. The full black line delineates the cortical from the limbic and cerebellar ROIs. We observe a strong symmetry between L and R cortical hemispheres (indicated by large values in the diagonal of the inter-hemisphere squares of the matrix), and that there are few strong connections between the cortical hemispheres. We find almost no interactions between the cortex ( $i = 0$  to  $i = 359$ ) and the limbic and cerebellar ROIs ( $i \geq 360$ ), implying these operate largely independently.



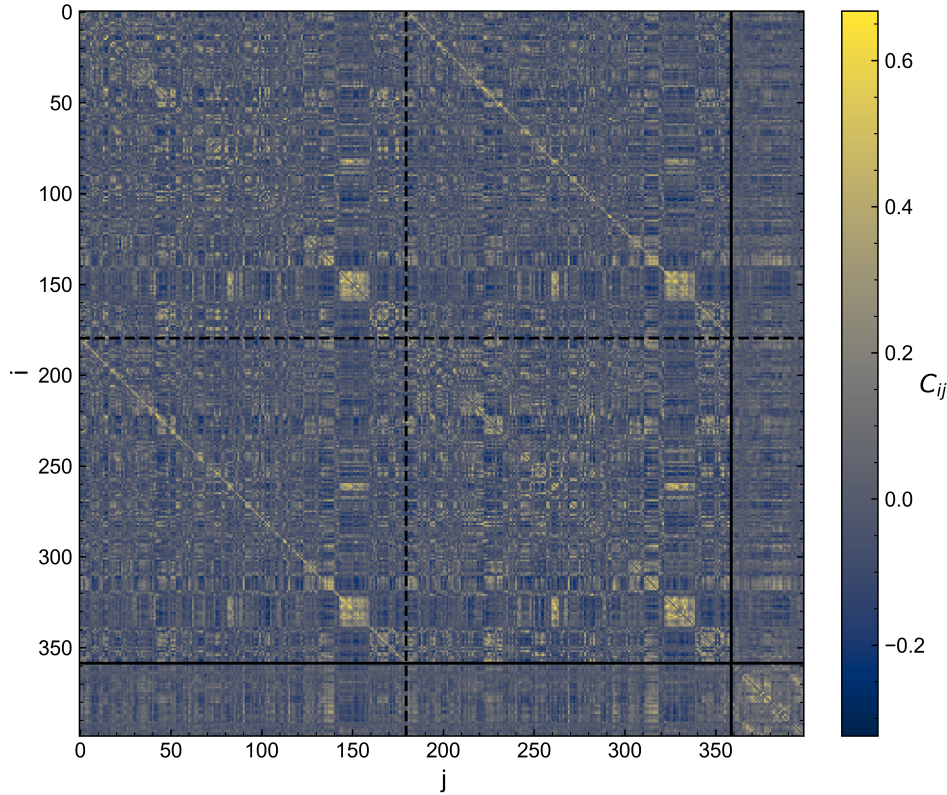
variety of behaviours [223]. The limbic and cerebellar ROIS ( $i \geq 360$ ) appear to operate largely independently of the cortex; indeed we see no strong couplings between these systems. Although PLM has not identified the size or distribution of the couplings correctly, it has still given us insights into the organisation and connectivity of the network with features that appear consistent with previous findings. The bias seems less important when PLM is used purely as a network reconstruction method. We do note, however, that PLM biases parameters to be larger, so that models inferred from small datasets will naturally appear less sparse.

### 7.3.1 Couplings and Correlations

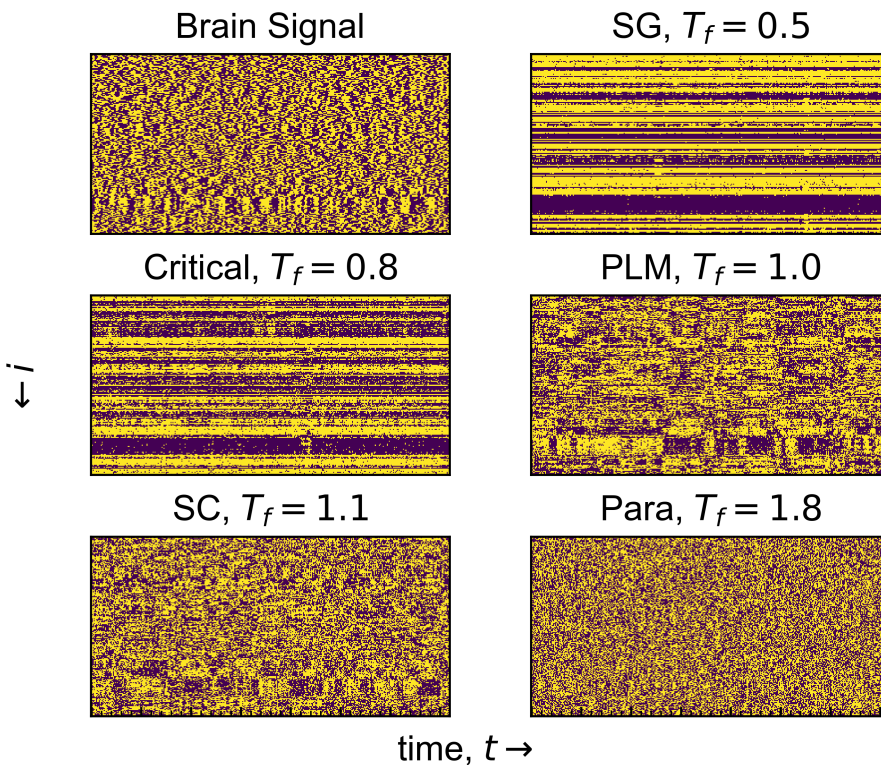
In section 3.3 we devoted some time to explaining how correlations  $C_{ij}$  and couplings  $J_{ij}$  are not equivalent, and that simple network topologies can lead to complex correlations [137]. Functional connectivity networks established from correlations<sup>17</sup> are unlikely to be representative of the true connections in the model. To demonstrate this point we show the correlation structure  $C_{ij}$  of the noMM condition in Fig. 7.6. Comparing Fig. 7.6 and Fig. 7.7 reveals that, although features such as the inter-hemisphere diagonal are present in both,  $C_{ij}$  is much more densely populated than  $J_{ij}$ . Thresholding both these to extract adjacency matrices will therefore produce networks with vastly differing topologies. Small-world properties (i.e. clustering and short topological distances between nodes) observed for networks constructed from correlations may therefore be the result of a less-connected coupling structure.

To explore these ideas further, we investigate how correlation networks associated with our inferred model  $\theta_{\text{noMM}}^*$  change as we vary  $T_f$ . We will

<sup>17</sup> Doing so is still commonplace in the neuro-imaging literature, indeed this was also done in the original analysis from which we sourced the mindfulness data analysed in this chapter [133].

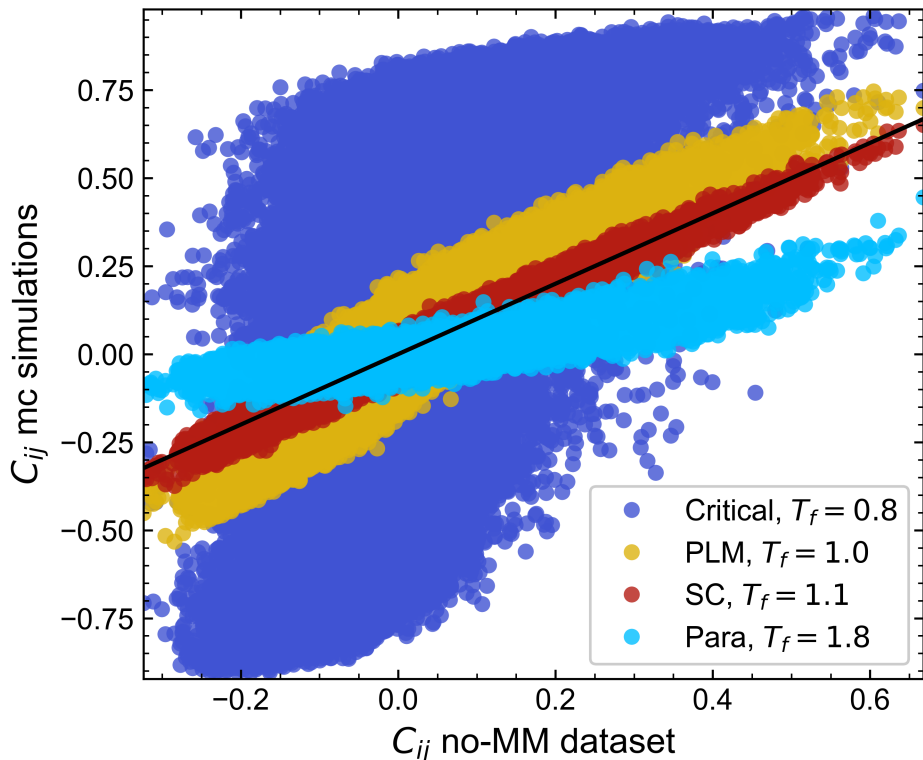


**Figure 7.7:** The covariance matrix  $\mathbf{C}$  (which here is equivalent to Pearson's correlation coefficients) for the noMM condition. Black lines represent the same structural separations described in Fig 7.6. The correlations display a similar structure to the couplings (c.f. Fig 7.6) and we observe that highly coupled ROIs are also highly correlated. The correlation matrix is much more densely populated than the corresponding inferred coupling matrix, suggesting that networks built from correlations may over-represent the true connectivity.



**Figure 7.8:** Segments of spin trajectories from the empirical brain signal and from simulations of the noMM PLM model for 5 different fictive temperatures. We only plot ROIs in the left hemisphere ( $i = 0$  to  $i = 179$ ). The ROI index runs vertically, while time is shown horizontally. We observe that the simulated dynamics of our model most closely align with those of the empirical brain signal at the fictive temperature of the self-consistency corrected model,  $T_f = 1.1$ . We identified spin-glass (SG), critical and paramagnetic state points using Fig. 7.5.

**Figure 7.9:** Dependence of covariances  $C_{ij}$  measured from MC simulations of the noMM PLM model at various fictive temperatures  $T_f$  compared to those observed in the binarized input noMM fMRI dataset. Different colours correspond to different  $T_f$  as shown in the legend. The black line shows perfect agreement. As  $T_f$  approaches  $T_c = 0.8$ , the correlations increase drastically, which is reflected by  $C^2$  being maximised at this temperature. We observe that our proposed self-consistency correction (SC) produces the model which most closely matches the covariances observed in the data (i.e. is closest to the black line). By matching the global average metric  $C^2$ , our SC correction therefore also improves agreement with individual microscopic covariances  $C_{ij}$ .

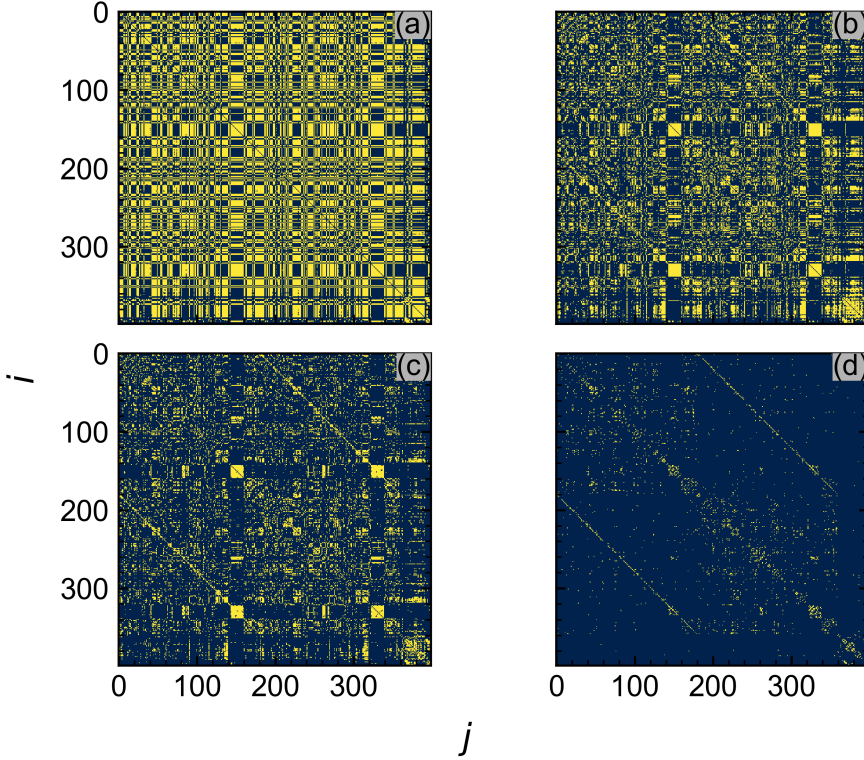


look at 5 temperatures in particular:  $T_f = 0.5$  where the model yields a SG,  $T_f = 0.8$  the critical point,  $T_f = 1.0$  the natural PLM model,  $T_f = 1.1$  the SC corrected PLM model and  $T_f = 1.8$  a P state-point. We identified the SG, P and critical state points from Fig. 7.5. As a starting point, we plot examples of trajectories from each of these state points along with the empirical signal from one of the imaging days in Fig. 7.8. Although the dynamics of the brain (empirical) and the simulated state points (Monte Carlo) differ substantially, we see the strongest visual similarity between the brain signal and inferred model at the SC temperature. This further supports that the SC correction is more representative of the data than the biased estimate obtained from PLM. We note that as  $T_f$  decreases the dynamics slow, and the correlations between spins increase.  $C_{ij}$  will therefore also vary as  $T_f$  varies.

To examine this behaviour quantitatively, we plot a scatter of the empirical covariances  $C_{ij}$  observed in the data vs the simulated covariances for our selected temperatures (excluding the spin-glass) in Fig. 7.9. We exclude the SG from our analysis as we know the data to be dynamic. As expected, the (absolute) values of  $C_{ij}$  increase as the critical point is approached. This behaviour is captured by our global correlation metric  $C^2$ . The spread of  $C_{ij}$  also increases (a larger area is occupied in our scatter). We also find the covariances of the SC corrected model are most closely aligned with those of the data. Constraining the global property  $C^2$  thus produces individual adjustments which align the microscopic covariances with those observed in the data. This provides further support for the use of the self-consistency correction; it produces the model most consistent with the  $C_{ij}$ , which is the over-arching aim of pairwise maximum entropy modelling.

We now consider how networks constructed by applying a constant





**Figure 7.10:** Adjacency matrices constructed from the correlations in Fig. 7.9 using a constant threshold of  $\delta = 0.1$ . We show the adjacency of the critical  $T_f = 0.8$  state-point in (a) of the un-modified PLM model  $T_f = 1.0$  in (b), of the self-consistency corrected model  $T_f = 1.1$  in (c) and of the paramagnetic state-point  $T_f = 1.8$  in (d). We see that varying the temperature drastically alters the implied network topology when this is constructed from correlations. Crucially, however, all these correlations were generated by simulating the PLM model  $\theta^*$  shown in Fig. 7.6.

threshold to  $C_{ij}$  change with  $T_f$ , as is common practice in the neuro-imaging community. We note that it is usually the linear correlation coefficient

$$R_{ij} = \frac{C_{ij}}{\sigma_i \sigma_j} = \frac{\langle s_i s_j \rangle - \langle s_i \rangle \langle s_j \rangle}{\sigma_i \sigma_j}, \quad (7.1)$$

that is thresholded. But as our spins only take values of  $s_i = \pm 1$ , we find that  $R_{ij}$  and  $C_{ij}$  are numerically equivalent. We extract an adjacency matrix  $\mathbf{A}$  by applying a constant threshold  $\delta = 0.1$  to each  $C_{ij}$ , setting  $A_{ij} = 1$  if  $C_{ij} \geq \delta$  and to 0 otherwise (i.e. we construct a network of positive correlations). The resulting adjacency matrices are shown in Fig. 7.10. Bright pixels indicate connections. We clearly see that the correlation network becomes more interconnected as we approach the critical point. At the critical point, panel (a),  $A_{ij}$  is highly connected<sup>18</sup>, and the sparse coupling structure we reported in Fig. 7.6 is obscured entirely. We see that both panels (b) (the PLM model) and (c) (the SC correction model) represent topologies that are more connected than  $J_{ij}^*$ . In the paramagnetic regime, panel (d),  $A_{ij}$  mirrors the strongest couplings found in Fig. 7.6. In summary, re-scaling the values of the inferred couplings by a constant  $1/T_f$  leads to vastly differing adjacency matrices when these are constructed by thresholding correlations, even though the underlying coupling structure stays constant. Defining functional connectivity through correlations means we are likely to overstate the inter-connectedness of the ROIs when there are non-trivial correlations in the data (i.e. when we are not deeply in the P phase). We thus show that networks with short topological path lengths (i.e. those that exhibit small-world properties) often observed in neuroimaging analyses [25, 67] can be the result of much simpler underlying coupling structures [137]. This further highlights the necessity of inverse methods in

<sup>18</sup> Indeed had we chosen to connect nodes based on absolute correlations  $|C_{ij}| \geq \delta$ , as done in [133], we would find an almost fully connected network.

establishing true functional connectivity.

## 7.4 Conclusions

In this chapter, we have illustrated the importance of small sample size biases when applying PLM to real neuroimaging data. We did this by performing a case study of a typical fMRI dataset. Analysing the first-order bias through sub-sampling showed that this dataset contained at most  $B_{\text{noMM}} \approx 2.7\tilde{B}$  samples, and will be strongly affected by small sample size effects. Not accounting for the bias severely impacts our conclusions, and can lead to false claims of criticality as the bias tunes the PLM model towards the critical point. Applying our purposed self-consistency correction allows us to counteract this and establish a lower bound on the true temperature of the data. Although PLM is exact in the limit of large sample size, we thus show that this “large sample size” condition is often unlikely to be achieved in real settings. Typical fMRI studies collect between 10-60 minutes of data per imaging session and sample every 1-2.5 seconds [13, 133, 224, 225]. We can therefore expect around  $B = 240 - 3600$  samples per imaging session. Applying PLM to single imaging session/participant studies poses a challenge, particularly if one wishes to consider more fine-grained representations of the brain (i.e. more ROIs) as we have previously shown that  $\tilde{B} \propto N$ . Conclusions drawn from models inferred from individuals, or comparisons across individuals, should therefore be regarded with a high degree of scrutiny. This is particularly important when one attempts to explore whether different physiological conditions (e.g. sleep vs wakefulness [199]) represent different state-points, as we have previously shown that  $\tilde{B}$  is also a function of the state-point of the data and so collecting the same number of samples  $B$  from each condition does not guarantee that the effect of the bias has been mitigated equally.

We have demonstrated that, even in the under-sampled regime, PLM still provides estimates of network structure which are consistent with previous authors’ findings. More importantly, we showed that networks constructed from correlations (as is common practice in the neuroimaging community) provide less reliable information regarding the underlying topology than inverse methods which infer couplings such as PLM. We claim that inverse methods are a necessity when trying to distinguish between true connections and coincidental correlations in functional connectivity analyses. This is especially so in the proximity of the critical point where large correlations can arise from simple model topologies. We show that the SC-corrected PLM model displays the best agreement with the empirical covariances, providing further evidence for the validity of this re-scaling approach when  $B$  is small. We expect PLM to excel when trying to establish generic features of the human state (i.e. properties of the average brain), which are studied by analysing large multi-participant aggregate datasets such as the 1000 Functional Connectomes Project\* or the Human Connectome Project†. This will be the aim of our next and final analysis chapter.

---

\*See [http://fcon\\_1000.projects.nitrc.org/](http://fcon_1000.projects.nitrc.org/) for the 1000 Functional Connectomes Project

†See <https://www.humanconnectome.org/study/hcp-young-adult> for the Human

## Summary of Chapter 7

- We applied pseudo-likelihood maximisation (PLM) to a typical resting-state functional magnetic resonance imaging (fMRI) dataset, with the aim of understanding how the inference bias impacts our conclusions.
- The dataset contains two biological conditions, which are mindfulness meditation (MM) vs the (control) no mindfulness meditation (noMM). We assess whether these two conditions constitute different state points and whether the inferred model is close to a critical point.

### **Disregarding the Bias**

If we disregard the bias, we conclude that mindfulness meditation measurably alters the state point of the human brain. The MM model has a lower temperature and it appears that practising mindfulness meditation tunes the brain towards a critical point.

### **Accounting for the Bias**

If we properly account for the bias, we conclude that the temperature difference between noMM and MM conditions is purely an artefact of the small sample size bias and that both models correspond to the same state point. Applying our self-consistency correction further shows that this state point, as a lower bound, is paramagnetic.

- We thus show that the small sample size bias can drastically alter conclusions in fMRI datasets with typical sample numbers. We argue that any PLM analysis of such data claiming criticality must therefore always perform an analysis of the bias (e.g. through sub-sampling as we do) and that analyses that fail to do so should be regarded with a high degree of scepticism.
- We find that the self-consistency correction improves our estimates of the individual covariances  $C_{ij}$ . This means that the self-consistency correction provides a better solution to the inverse Ising problem than uncorrected PLM, further supporting the use of this correction.
- We showed that a constant inferred coupling structure can lead to a range of correlation networks, and advocate for the more widespread use of inverse methods for network analyses of neuroimaging data.

## Chapter 8

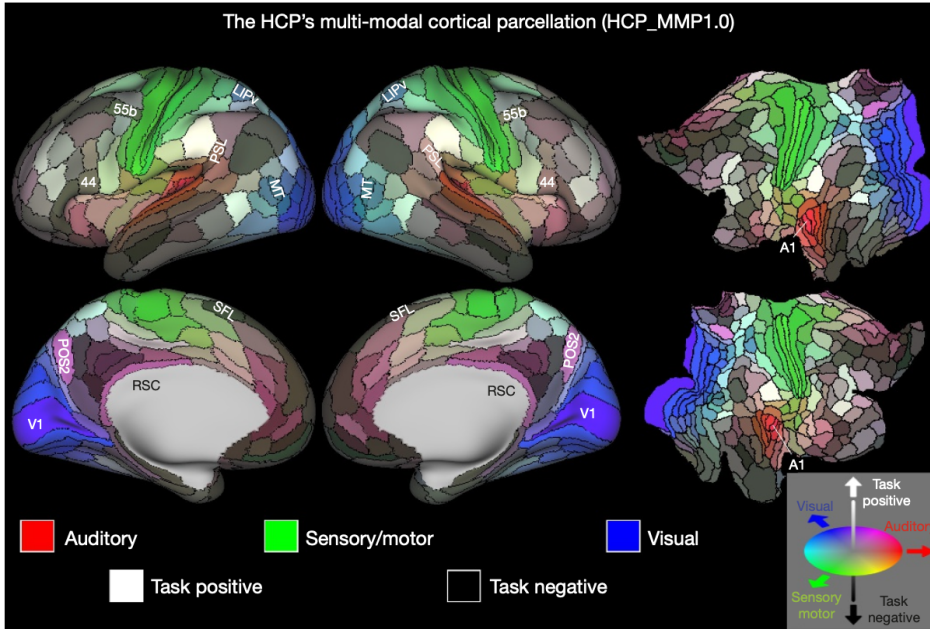
# The Human Connectome Project: a Large Sample Size Study

We have now established that statistical biases affect the quality of pseudo-likelihood maximisation (PLM) inferences (see chapters 5 - 7). Do existing large datasets overcome these limitations? To assess this, we apply our analysis to one of the largest openly available fMRI datasets, the Human Connectome Project (HCP) young adult dataset. As we will see, this contains a sufficiently large number of samples to largely disregard the effects of the bias highlighted in the previous chapters. This allows us to investigate whether the resting state of the brain is a critical state, knowing with confidence that our conclusions will not be the result of the small sample bias. We will find that the PLM model of the HCP dataset corresponds to a near-critical paramagnetic state point. This is in line with the bias-adjusted conclusions we drew from the small sample size dataset analysed in the previous chapter. We will demonstrate that the functional connectivity inferred from PLM is much sparser than that produced by commonly applied correlation-based methods. We argue that the PLM model is a better representation of the functional connectivity of the brain and advocate for the more widespread adoption of inverse methods such as PLM. We further analyse the effect that thresholding (i.e. setting some couplings to zero) has on the statistical mechanics of our models and detail a self-consistent method to threshold the couplings while conserving the overall correlations of the input data. We find that this allows us to safely remove 15% of couplings without impacting the system's statistics. This avoids issues associated with the arbitrariness of thresholding when attempting to construct network topologies from weighted quantities such as couplings or correlations.

### 8.1 The HCP Resting-State Dataset

The Human Connectome Project (HCP)<sup>19</sup> provides a state-of-the-art openly available aggregate fMRI dataset. With that said, extensive pre-processing is required to extract the region-specific time series that we ultimately use as our input datasets for PLM from the raw volumetric BOLD signal. **I do not**

<sup>19</sup> [www.humanconnectome.org/study/hcp-young-adult](http://www.humanconnectome.org/study/hcp-young-adult)



**Figure 8.1:** Regions of interest in the multi-modal parcellation of the human cerebral cortex defined in Ref [23]. There are 180 ROIs in each cortical hemisphere, which are replicated across the hemispheres for a total of 360 ROIs, e.g. both left and right hemispheres contain a V1 region of interest. Both inflated (left and middle columns) and flattened representations (right column) of the cortical surface are shown. Colours show the extent to which each region is associated with a specific function or task. Adapted from Ref. [23].

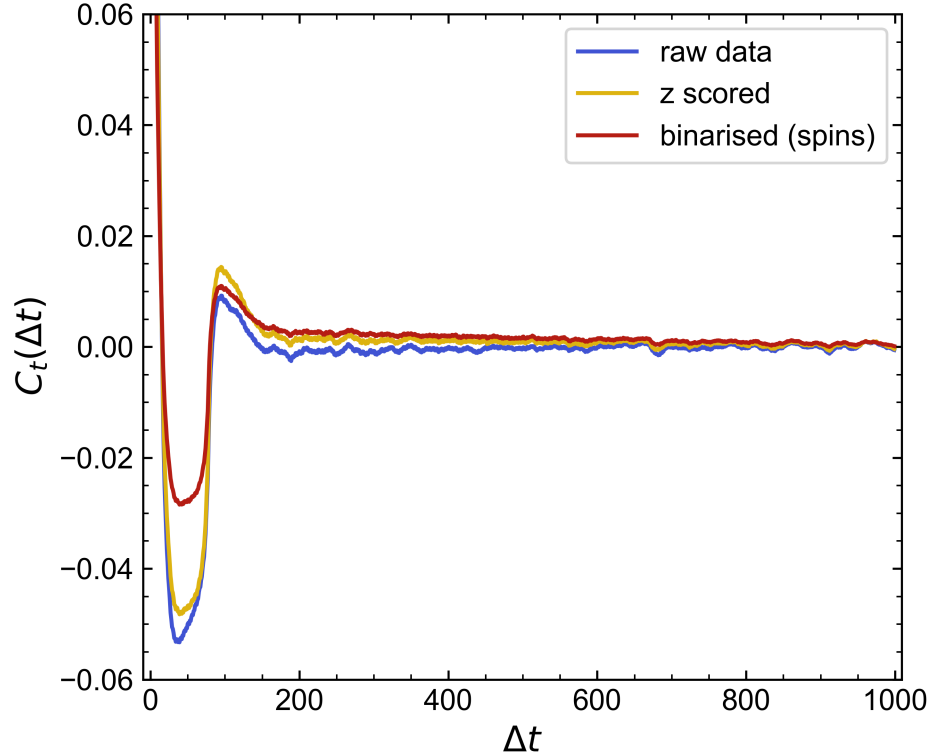
have the expertise to perform this pre-processing and explicitly acknowledge that this was done by our collaborator Dr. Hiroaki Hamada at Araya Inc. The HCP dataset contains data both from resting-state studies and from those where a specific task was performed. Here, we will exclusively analyse the resting-state data. A PLM model inferred from this dataset, therefore, represents the brain in an equilibrium-like stationary condition.

In total, we will analyse data from 161 individuals, with  $B = 4724$  samples collected from each of these. The data we receive has been parcellated (i.e. segmented into regions of interest (ROIs)) according to the Atlas defined in [23]. We show these regions in Fig. 8.1, and note that they do not represent volumes of fixed size, but rather specific areas of the brain known to be associated with specific functions. In total, there are  $N = 360$  ROIs, each of which we label as a spin  $s_i$ . ROIs  $i = 0$  to  $i = 179$  correspond to the left (L) hemisphere of the cerebral cortex, while ROIs  $i = 180$  to  $i = 359$  have come from the right (R) hemisphere. Note that the ROI definitions are symmetric, i.e. ROI  $i = 0$  corresponds to the Visual 1 (V1) region of the left hemisphere, while ROI  $i = 180$  is the V1 region in the right hemisphere, so that the same 180 regions of interest are defined within each hemisphere. We will perform PLM on the total aggregate dataset, for which there are  $B = 161 \times 4724 = 760,564$  samples, and expect that this will ensure that the bias of our inference remains minimal.

The values of each time series for each ROI are continuous. As in chapter 7, we therefore again binarize the data by first removing the average and normalising by the standard deviation (z-scoring) and then setting  $s_i(t) < 0 = -1$  and  $s_i(t) \geq 0 = +1$ . We plot the autocorrelation of the data in Fig. 8.2. We once more see that there are non-trivial dynamics, but that the autocorrelation time remains small ( $\tau = 1$ ), which for our purposes means we can consider each sample as independent and identically distributed (i.i.d). The similarity of  $C_t$  for the raw data, the z-scored



**Figure 8.2:** The normalised auto-correlation function  $C_t$  of each configuration vs the delay time  $\Delta t$  for the HCP dataset. We show the dynamic correlations of the raw data, the z-scored (i.e. normalised) data and the binarised spin data. We observe that the binarisation process does not have a significant impact on  $C_t$ . As in the last chapter, the overall size of the auto-correlation is small (see the y-scale) and we thus assume that we can take each time point as representing an independent and identically distributed (i.i.d.) random sample of the system. The plotted lines show the average over all participants and only the first 1000 delays are shown.



data and the binarised data shows that the binarisation process has not substantially affected the dynamics of the system.

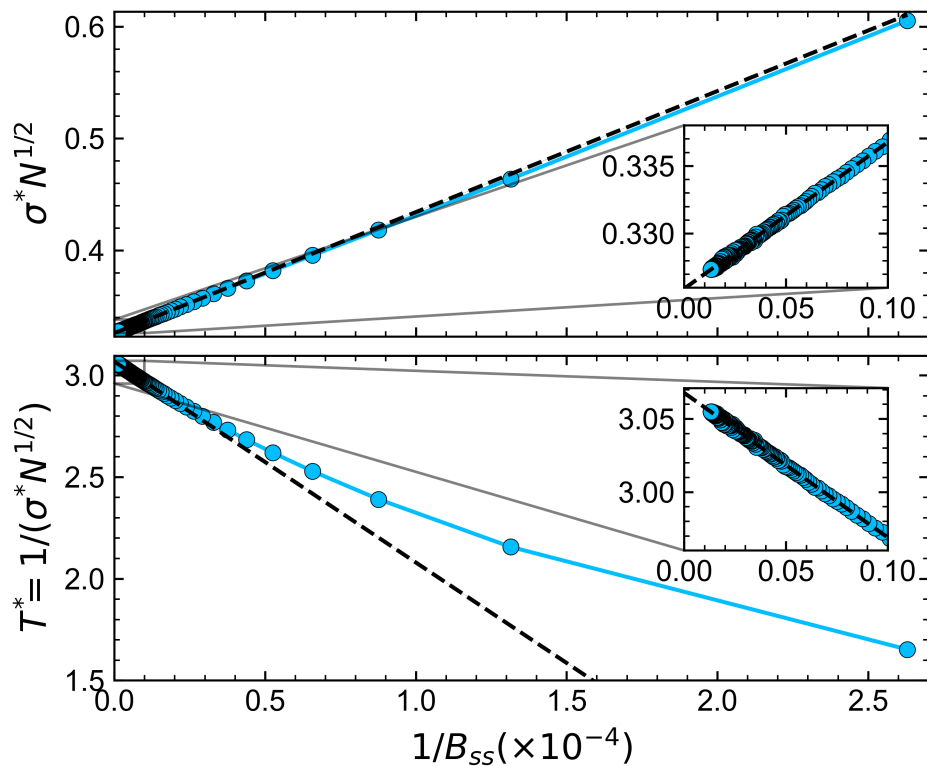
## 8.2 The Resting-State Coupling Network

### 8.2.1 Sub-Sampling Analysis

We now perform PLM on the aggregate HCP resting-state dataset containing  $B = 760,564$  samples. So as to not fall into the pitfalls highlighted throughout the previous chapters, we begin by sub-sampling our data and performing PLM on each sub-sample. We again sub-sample in a greedy way, drawing as many samples of size  $B_{ss}$  as possible from  $B$  without repetition, and then averaging over the PLM results from each of these sub-samples for each  $B_{ss}$ . This was to ensure that as much of the information contained within  $B$  as possible is present for each  $B_{ss}$ . The output of the sub-sampling analysis can be seen in Fig. 8.3, where we plot the saturation of the re-scaled standard deviation of the couplings  $\sigma^* N^{1/2}$  and the temperature  $T^*$  as functions of  $1/B_{ss}$ . We clearly see from the inset (showing data corresponding to  $B_{ss} \geq 10^5$ ) that the dependence of the (inverse) temperature follows  $1/B$ . The aggregate HCP dataset, therefore, sits in the asymptotic first-order bias regime. We now fit the saturation of the standard deviation of our inferred models with

$$\sigma^* = \sigma^0 + b_1/B_{ss}, \quad (8.1)$$

for  $B_{ss} \geq 10^5$ , and extract the asymptotic (unbiased) estimate  $\sigma^0$  and the first order bias parameter  $b_1$  from this fit. We find that when  $B = B_{\max}$ , there is only a 0.43% discrepancy between  $\sigma^0$  and  $\sigma^*$  when considering



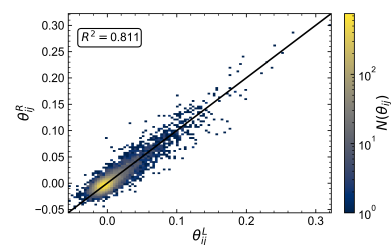
**Figure 8.3:** The top panel shows the rescaled standard deviation of the inferred parameters  $\sigma^* N^{1/2}$  as a function of the inverse sub-sample size  $1/B_{ss}$ . The associated temperature of the model for each  $B_{ss}$  is shown in the bottom panel. The inverse sub-sample size is shown in units of  $10^{-4}$ . Insets provide detailed views of the large sample size limit ( $B_{ss} \geq 10^5$ ) from which we estimate the bias parameter  $b_1$ . From our fit, we find  $\sigma^0 N^{1/2} = 0.32592 \pm 0.00001$  ( $T^0 = 3.068$ ) and  $b_1 = (0.1082 = 0.0003) \times 10^4$ .

the full HCP sample. We conclude that the PLM model inferred from the aggregate HCP dataset is free from significant bias effects. We will therefore accept the estimated PLM model as representative of the ground truth model, and perform no correction. We will then assess if this model represents a critical state point to address our original hypothesis.

Before continuing with our primary analysis, we briefly use this opportunity to again highlight the difficulty of using PLM to analyse individual imaging sessions, where limited data is available. We set the minimum sub-sample size in Fig. 8.3 equal to  $B = 4724$ , which is equivalent to the sample sizes collected for each of the 161 individuals in the dataset. We see immediately that  $T_{\text{individual}}^* \approx 0.5 T_{\text{aggregate}}^*$ . Halving the temperature of the inferred model can drastically alter the phase attributed to that model. State points inferred from individual imaging sessions are, therefore, unlikely to represent the true state point of the brain and are strongly influenced by the small sample size bias. For this reason, we focus our discussion entirely on the PLM model obtained from the full  $B = 760,564$  cross-participant sample.

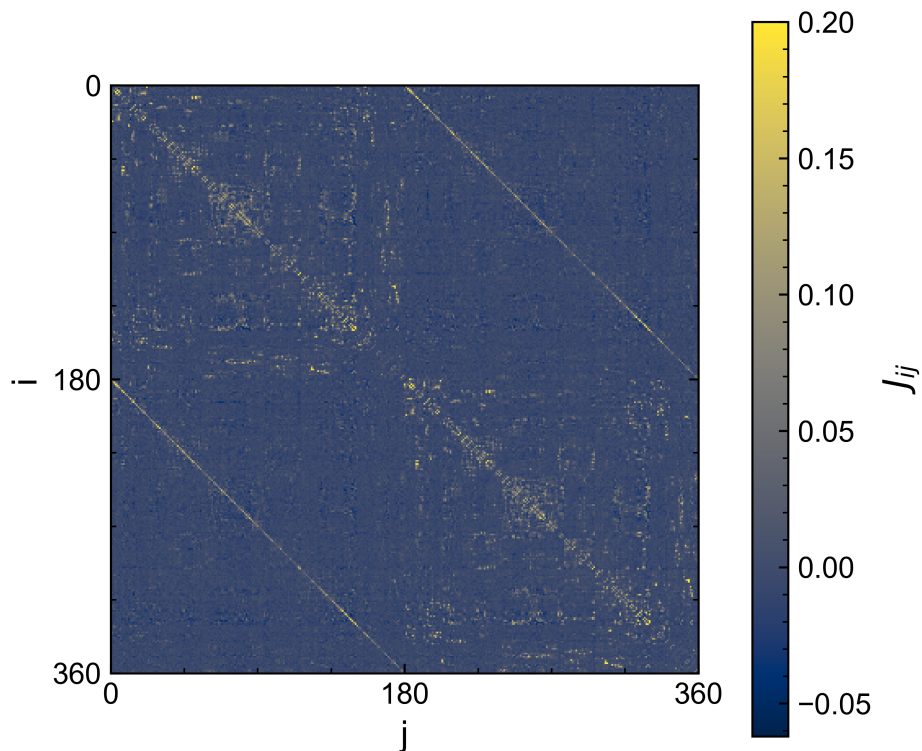
## 8.2.2 Overview of the Inferred Connectivity

We show the PLM model (i.e. the functional connectivity) inferred from the full aggregate HCP dataset in Fig. 8.4. As there is no longer a known “true” model to reference, we simplify our notation by dropping the  $*$  superscript and use  $J_{ij}$  to denote our inferred PLM couplings from here on. The two intra-hemisphere networks (i.e. L-L and R-R connections) are similar (correlating parameters in the left hemisphere  $\theta^L$  with those in the right hemisphere  $\theta^R$  yields  $R^2 = 0.811$ , see Fig. 8.5). This is further highlighted by



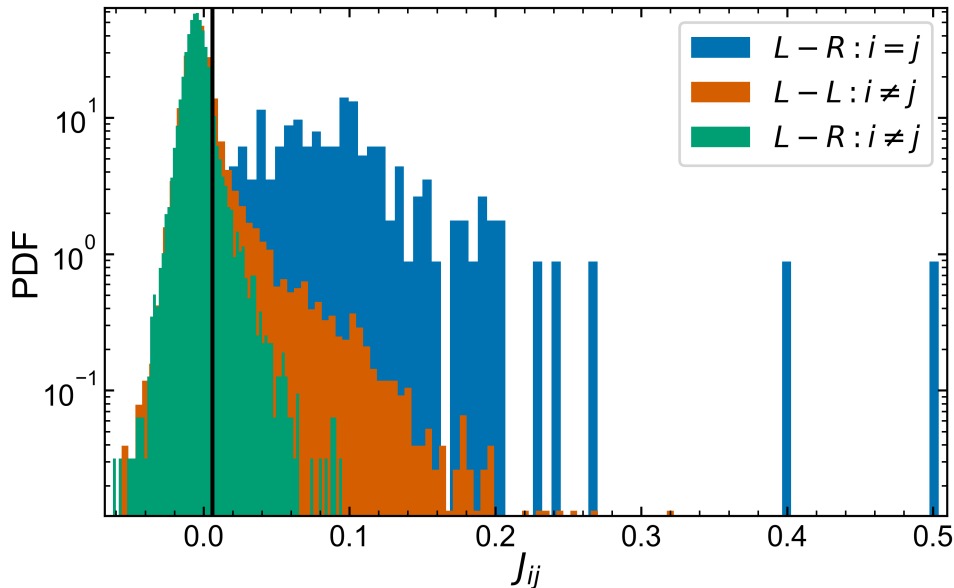
**Figure 8.5:** Correlation of parameters in the L and R hemispheres.

**Figure 8.4:** The coupling matrix  $J_{ij}$  inferred from the aggregate HCP resting-state dataset. A similar structure to the mindfulness dataset is observed. Left ( $i = 0$  to  $i = 179$ ) and right ( $i = 180$  to  $i = 359$ ) hemispheres appear highly symmetric, with many connections within each hemisphere, but few connections between hemispheres. The colour map is limited to 0.2 as few parameters exceed this value, see Fig. 8.6.



the large positive couplings seen along the diagonal of the inter-hemisphere network (L-R), i.e. the couplings  $J_{i,j=i+180}$ , telling us that nodes corresponding to the same ROIs in L and R hemispheres are strongly functionally connected (i.e. activate in unison). We, in fact, find that these couplings are the largest couplings in the network as a whole, which we further illustrate in Fig. 8.6. We identify a hierarchy of couplings. The strongest couplings in the network (the diagonal elements of the L-R inter-hemisphere network) encode the symmetry between identical ROIs in the L and R cortical hemispheres, e.g. between the first visual area of the left (V1-L) and the right (V1-R) hemispheres. Within each hemisphere (showing the left hemisphere as an example in Fig. 8.6), we find a skewed coupling distribution with a long positive tail. Strong positive (excitatory) intra-hemisphere couplings are also present within each hemisphere. We note that the bulk of L-L couplings (around 2/3) are negative (inhibitory) and weak. On average, the weakest couplings in the network are those contained in the off-diagonal elements of the L-R inter-hemisphere network. We find that the distribution of these couplings has a smaller variance than the L-L network and contains values much smaller than the diagonal L-R couplings.

Overall, this implies the resting-state brain is organised in the following way. Within each hemisphere, we find a non-trivial coupling (connection) distribution. The majority of these couplings are weak and inhibitory. We do however identify a long tail of positive couplings, implying a strong excitatory network exists within the hemispheres. This coupling structure is highly symmetric and is found in both L and R hemispheres. Turning now to the inter-hemisphere (L-R) connections, we find little evidence to suggest any complex functional connectivity between the hemispheres. The main identifiable source of interaction between the hemispheres is the “self”



**Figure 8.6:** The distribution of inferred PLM couplings  $J_{ij}$  for different components of the network. The distribution of couplings in the left-left (L-L) *intra*-hemisphere network is shown in orange, the diagonal elements of the left-right (L-R) *inter*-hemisphere network are shown in blue, and the off-diagonal elements of the L-R network are shown in green. Diagonal elements of the L-R network represent the coupling between regions of interest with the same label in each hemisphere, e.g. V1-Left to V1-Right. The biggest couplings in the network are the L-R diagonal elements. Off diagonal elements of the L-R network are, on average, smaller than those of the L-L network.

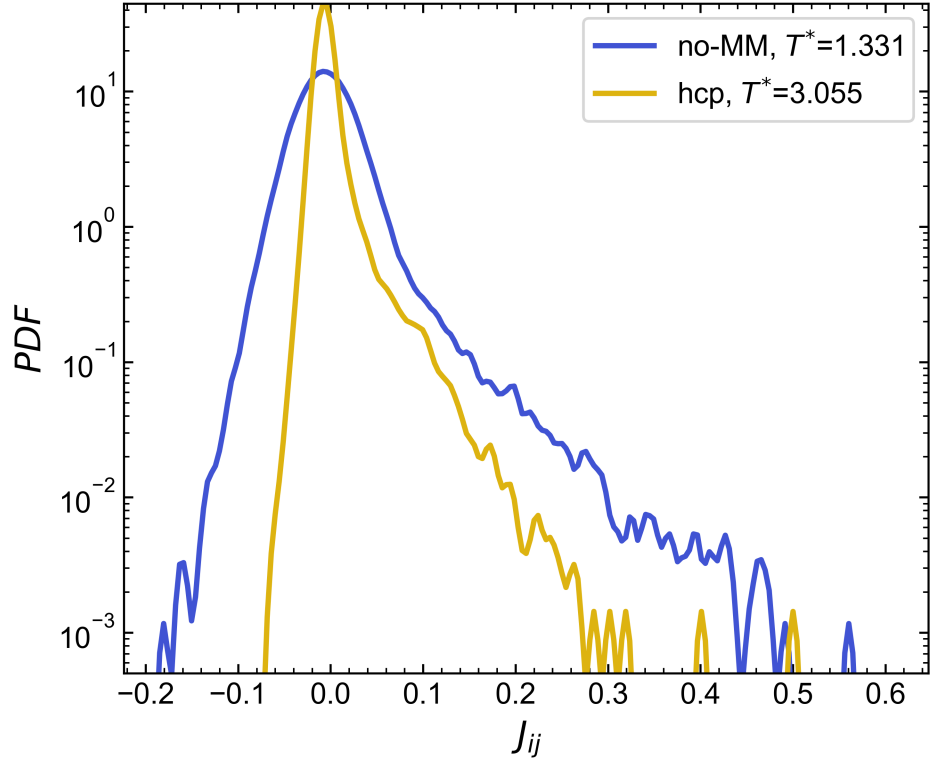
coupling of identical regions (i.e.  $J_{i,j=i+180}$ ) in both hemispheres. While connectivity within L and R hemispheres is mirrored, we find that the off-diagonal elements of the inter-hemisphere (L-R) network ( $J_{i,j \neq i+180}$ ) are small, implying that non-matched ROIs in one hemisphere have little effect on those in the other. This suggests that both hemispheres would be well described by decoupled sub-networks which activate in unison.

### 8.2.3 Characterising the Inferred Coupling Distribution

Now that we have established the broad structure of the inferred network, we want to devote some more time to understanding the distribution of the couplings as a whole. We begin by plotting the probability density function (PDF) of the coupling distribution inferred from the HCP dataset in Fig. 8.7. For reference, we also plot the coupling distribution corresponding to the noMM dataset from the last chapter. We clearly see that the noMM condition (for which  $B \approx 10,000$ ) and the HCP distribution share some features (including a peak at a small, negative  $J_{ij}$  and a long extended tail for positive couplings). The general magnitude of the couplings is larger in the noMM dataset leading to a distribution with heavier tails. We again stress that this is an effect of the small-sample bias of PLM. One can imagine, however, that the shape of the noMM distribution may collapse to that of the HCP distribution in the  $B \rightarrow \infty$  limit.

The HCP distribution itself is characterised by the following features. We note a sharp peak occurring when  $J_{ij}$  is small (i.e. weak) and negative. The distribution is non-symmetric about  $J_{ij} = 0$ , with a long slowly decaying tail for the positive couplings. We further find  $\min(J_{ij}) \approx -0.1$  and  $\max(J_{ij}) \approx 0.5$ , showing that negative interactions are more tightly constrained than positive interactions. The distribution for  $J_{ij} < 0$  appears approximately exponential (i.e. normal). We will now devote some time to characterising the behaviour of the non-standard tail for  $J_{ij} > 0$ . We look at the role that negative couplings play in mediating the system in the next section.

**Figure 8.7:** Comparison of the inferred coupling distributions found for the noMM (blue) dataset in the last chapter and the aggregate HCP dataset in this chapter (orange). The sample size  $B$  is much smaller in the noMM dataset (c.f.  $B \approx 10,000$  in the noMM dataset vs  $B \approx 1,000,000$  in the HCP dataset). Both distributions display long tails for  $J_{ij} > 0$ , but the HCP distribution is much more sharply peaked and contains smaller parameters. We believe that the difference between these two distributions is largely an effect of the small sample bias, highlighting the need to properly account for this effect.



### 8.2.4 Characterising the Positive Tail

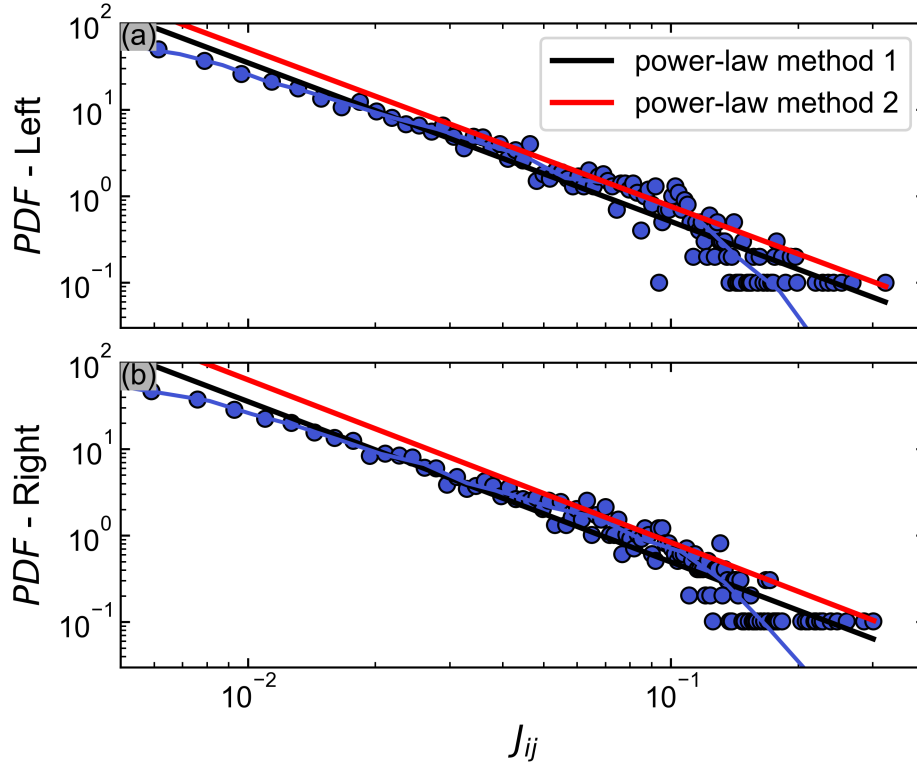
We now devote some time to characterising the positive tail of the coupling distribution. We previously noted in section 8.2.2 that the largest couplings in the network were the diagonal elements of the L-R network. These couplings represent a trivial component of the network (the symmetry between hemispheres) and contain no real information regarding the network structure found within each hemisphere. To better understand the true connectivity of each hemisphere we exclude these elements from our analysis and analyse the L-L and R-R sub-networks independently instead. The convergence of slowly decaying (heavy-tailed) distributions is more clearly represented on log-log axes, and as such, we plot PDFs of the coupling distribution from the L-L and R-R sub-networks in the top and bottom of Fig. 8.8 respectively. The tail of the coupling log-log PDF appears approximately linear, spanning roughly an order of magnitude in  $J_{ij}$ , implying that the positive tails of the coupling distributions in this region can be described by a (truncated) power law. We see this by noting that if the probability  $P(J_{ij})$  of observing a coupling with value  $J_{ij}$  is:

$$P(J_{ij}) = AJ_{ij}^{-\gamma}, \quad (8.2)$$

where  $A$  is a normalisation constant and  $\gamma$  is the exponent of a power-law, then

$$\log(P(J_{ij})) = \log(A) - \gamma \log(J_{ij}). \quad (8.3)$$

However, we also know that the core of the coupling distribution at  $J_{ij} \sim 0$  is well described by a Gaussian (see Fig. 8.7), and so expect there to be some



**Figure 8.8:** Probability density function (PDF) for positive couplings ( $J_{ij} > 0$ ) in the left (a) and right (b) hemispheres on double logarithmic axes. Linearly binned data is shown by blue dots, while logarithmically binned data is shown by the blue line. The approximate linearity of the tail on log-log axes leads us to fit the tail with power law functions via two methods (black and red lines), yielding the exponents shown in Table 8.1.

value of  $J_{ij}$  where we cross from Gaussian (exponential) behaviour to the power-law behaviour of the large  $J_{ij}$  tail. To properly characterise the tail we need to find two values: the cross-over value (which we denote by  $x_{\min}$ ) describing the minimum value of  $J_{ij}$  above which the coupling distribution is described by a power-law, and the exponent  $\gamma$  of the power-law. We implement two methods to extract the values  $x_{\min}$  and  $\gamma$ , and summarise the headline results in Table 8.1. We now describe both methods in more detail before discussing the values of the power-law parameters further.

Method	$x_{\min}$ (Left)	$x_{\min}$ (Right)	$\gamma$ (Left)	$\gamma$ (Right)
Method 1	$4 \times 10^{-3}$	$4 \times 10^{-3}$	1.8	1.9
Method 2	$6 \times 10^{-3}$	$7 \times 10^{-3}$	1.8	1.9

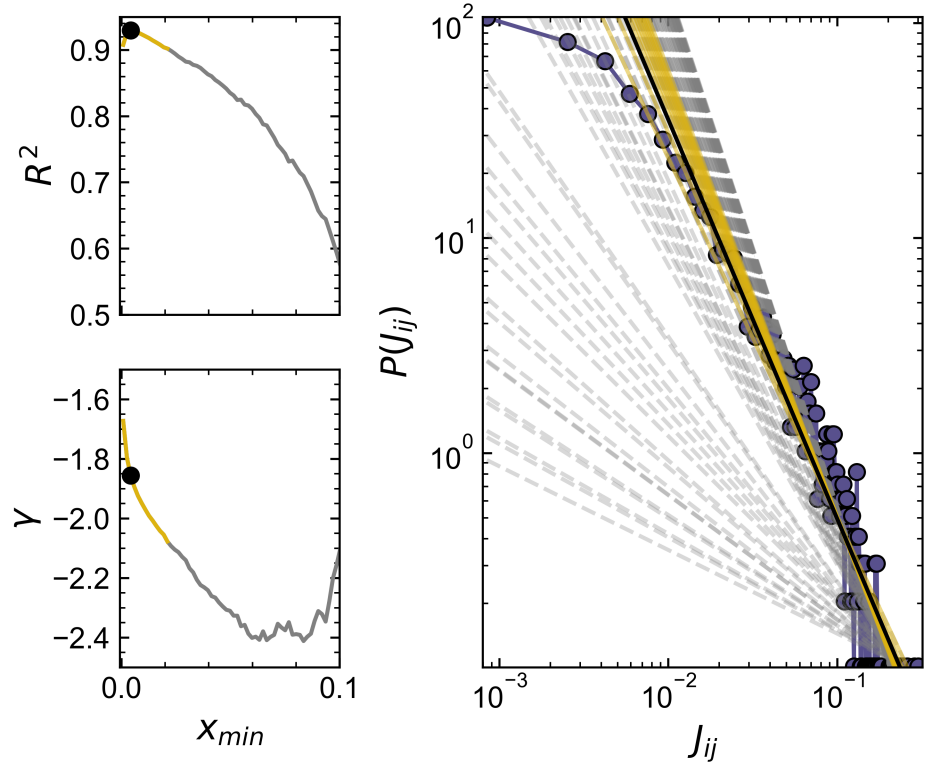
Table 8.1: Summary of results from the two power-law fitting methods for the positive tail of the couplings in the left and right hemispheres. The fitting methods are described in more detail below.

### Method 1: Simple Binning and Fitting

As established, if we plot the probability distribution of  $J_{ij}$  on double logarithmic axes we can perform a simple linear fit to extract the exponent  $\gamma$  from (8.3). This requires us to first discretize (i.e. sort into bins) our couplings. We bin our couplings by choosing 200 equally sized bins spaced evenly between 0 and  $\max(J_{ij})$ . We only fit Eq. 8.3 for bins which contain at least 1 coupling (i.e. where  $\text{PDF}(J_{ij}) \neq 0$ ). To determine the cross-over value  $x_{\min}$  we vary the range of  $J_{ij}$  (including only  $J_{ij} \geq x_{\min}$ ) over which we



**Figure 8.9:** Summary of method 1, where we bin our data and perform a linear fit to the binned data in double logarithmic space. The left panels show how our fitting metric  $R^2$  and the power-law exponent  $\gamma$  vary as functions of the cross-over value  $x_{\min}$  denoting the start of the heavy tail. The black dots show the values of  $R^2$  and  $\gamma$  at  $x_{\min}$  which maximises  $R^2$ . Orange regions show  $x_{\min}$  values where  $R^2 \geq 0.9$ , i.e. where there is very good agreement with the fit. The right panel shows how each fitted power law corresponds to the data. Blue dots show binned data. Black lines show the optimal fit, while orange lines show fits where  $R^2 \geq 0.9$ . Grey lines correspond to all other fits.

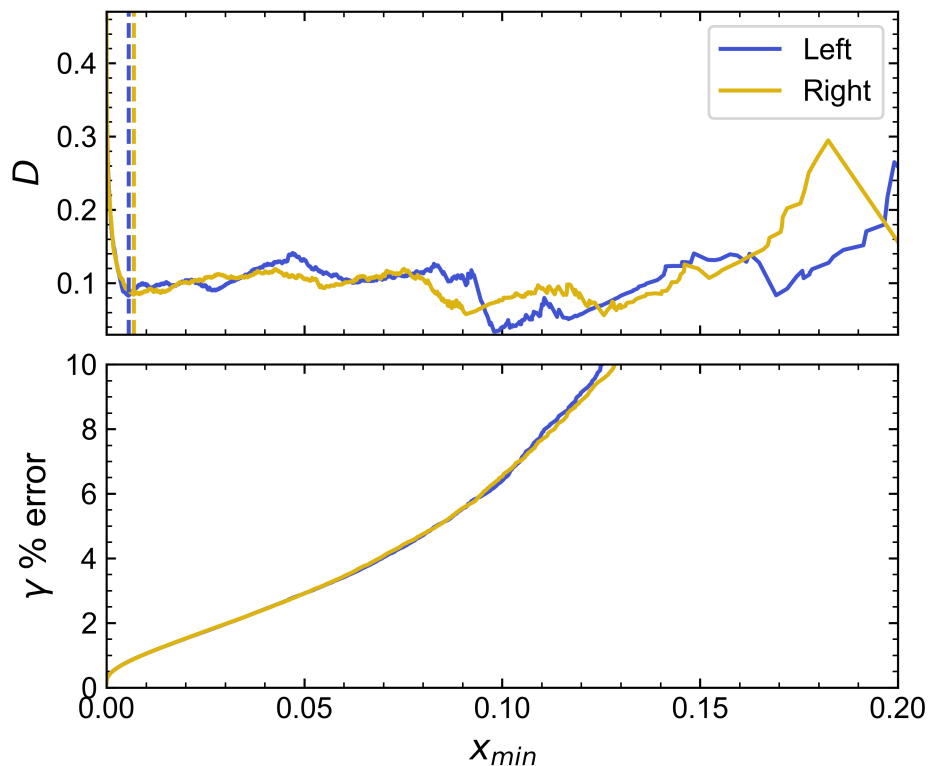


perform our fit and measure the coefficient of determination  $R^2$ . We define  $x_{\min}$  as the value of  $J_{ij}$  which maximises  $R^2$  in this process. We summarise this fitting procedure visually in Fig. 8.9 for the left hemisphere (although the analysis results are practically identical for the right hemisphere). The left two panels track how  $R^2$  and  $\gamma$  vary as functions of  $x_{\min}$ . The right panel shows the range of possible fits for each  $x_{\min}$ ; orange fits satisfy the condition  $R^2 \geq 0.9$  (i.e. agree very well with the data), while the black line shows the optimal solution. We clearly see that a broad range of  $x_{\min}$  (corresponding to a broad range of  $\gamma$ ) agree well with the data. Although we state the optimal value of the exponents as 1.8 in Table 8.1, we note that  $\gamma = 1.7$  to  $\gamma = 2.1$  also satisfy  $R^2 \geq 0.9$  and good produce agreement with linearity.

## Method 2: A Maximum Likelihood Method

The methodology described above is simple and widely used to test empirical data for power-law dependency. However, Refs. [226, 227] note that this process is subject to several systematic errors, for instance, that the extracted exponents depend on the size and spacing of the bins. Clauset *et al.* [226] therefore propose an alternate maximum likelihood-based approach to extracting the power-law exponent. Formally, we can express the correctly normalised probability of observing a variable  $x$  distributed according to a power-law with exponent  $\gamma$  above some minimum value  $x_{\min} > 0$  as

$$P(x) = \frac{\gamma - 1}{x_{\min}} \left( \frac{x}{x_{\min}} \right)^{-\gamma}. \quad (8.4)$$



**Figure 8.10:** Summary of method 2, where we use a maximum likelihood approach to extract the power law exponent of the heavy tail of the couplings. Analysis of the left hemisphere is shown in blue, and the right hemisphere is in orange. Top, the Kolmogorov-Smirnov distance metric  $D$  as a function of  $x_{\min}$ . Dashed vertical lines indicate the value of  $x_{\min}$  where  $D$  is minimized if we require the error in  $\gamma$  to be  $\leq 5\%$ . Bottom, the % error on the estimate obtained for  $\gamma$  at each  $x_{\min}$ . This increases monotonically as  $x_{\min}$  increases as fewer data points are included in the fit.

Given that we have a random variable  $x$  sampled  $n$  times, denoting each sample as  $x_i$ , Clauset *et al.* [226] show that the maximum likelihood estimate (MLE) obtained for  $\gamma$  from observations distributed according to (8.4) can be derived analytically as

$$\gamma = 1 + n \left[ \sum_{i=1}^n \ln \frac{x_i}{x_{\min}} \right]^{-1}. \quad (8.5)$$

When estimating the exponent of the tail of our coupling distribution, we note that  $n = N(N - 1)/2$  (the upper triangle of the coupling matrix  $\mathbf{J}$ ), with  $x_i$  corresponding to each unique coupling  $J_{ij}$ . Once again, one then needs to find the cutoff value  $x_{\min}$  defining the tail of the distribution by varying  $x_{\min}$  and minimizing some distance metric (c.f.  $R^2$  in the simple binning method). *We again stress that changing the value of  $x_{\min}$  changes the extracted exponent.*

Fortunately, an extensive Python package (aptly named *powerlaw*) already exists to perform this optimization [228]. We now use this to extract  $x_{\min}$  and  $\gamma$  as in the previous method. We show the Kolmogorov-Smirnov distance metric  $D$  and the estimated percentage error of  $\gamma$  as functions of  $x_{\min}$  in for both hemispheres in Fig. 8.10. We see that the landscape of  $D$  is rugged with many local minima and that the global minimum of  $D$  occurs at around  $x_{\min} = 0.1$  at which  $\gamma \approx 5$  for both L and R hemispheres. However, we also see that the error is a monotonically increasing function of  $x_{\min}$ , which arises naturally from our previous discussion of MLE biases as fewer data points are included in the optimisation, and that the global minima in  $D$  correspond to errors in the region of 8 to 10%. Moreover, inspecting the range of  $J_{ij}$  defined by  $x_{\min} \geq 0.1$  in Fig. 8.8 shows that the exponents



at the global minima of  $D$  correspond to the “truncated” component of the distribution. The global minima of  $D$ , therefore, do not capture the exponent of the main power-law scaling regime between ( $10^{-2} \leq J_{ij} \leq 10^{-1}$ ), and instead characterise the rate of decay of the truncation. We instead now look for the minima of  $D$  under the condition the percentage error of  $\gamma$  is  $\leq 5\%$ . These minima are marked with dashed vertical lines in Fig. 8.10, and their corresponding exponents are reported in Table 8.1. We note that  $D$  is flat in the vicinity of these minima, indicating that a range of  $x_{\min}$  produce equally good estimates for the onset of the tail. This provides a numeric explanation for the wide range of  $x_{\min}$  satisfying  $R^2 > 0.9$  in Fig. 8.9.

Power-law distributions are not the only distributions which lead to heavy tails; indeed there are many other candidate distributions that may be equally plausible for empirical data such as ours, e.g. stretched-exponential and log-normal distributions. The powerlaw package [228] also implements methods with which to test against these alternative distributions (by performing a log-likelihood ratio test). We now briefly discuss this and report p-values from the log-likelihood ratio test comparing the power law to a number of alternative candidate distributions. This discussion holds for both L and R hemispheres due to their similarity. The minimum requirement for a distribution to be heavy-tailed is that it differs from an exponential distribution (i.e. from a “normal” process). We find ( $p = 0.03$ ) in favour of a power law, indicating that the heavy-tailedness of the coupling distribution is statistically significant. The likelihood ratio test determines that other heavy-tailed alternatives including exponentially truncated power-laws, stretched-exponentials and log-normal distributions are all more likely candidate distributions ( $p \approx 0$ ). The pure power-law behaviour we have considered so far is therefore by no means the only appropriate description of the tail. However, as the tail only spans  $\sim 1$  order of magnitude of  $J_{ij}$  and as we cannot gather any additional data points, we cannot identify exactly which of these equally plausible distributions truly describes the data. Given our limited data, we claim that a simple power law provides the most intuitive description of the behaviour of the tail. Either way, testing against the exponential distribution clearly demonstrates that, regardless of its specific functional form, the strengths of positive functional connections in the brain are heavy-tailed.

## Fitting Results Summary

We implemented two methods with which to characterise the right (positive) tail of the coupling distribution. The fundamental assumption of these was those couplings in the tail are distributed according to a power law. Our fitting methods return two key parameters,  $x_{\min}$  which determines the cross-over from exponential to heavy-tailed behaviour, and  $\gamma$  the exponent of the power law. We found both methods to be in good agreement with each other, and moreover, that the general symmetry of L and R hemispheres also causes the tails of the distribution to be similar (see Table 8.1). Although we state the optimal exponents  $\gamma \approx 1.8$ , our discussion of both fitting methods has shown that the quality of our data in the tail is not sufficiently large to accurately determine this value. For method 1, this is shown by the large range of linear fits satisfying  $R^2$ , and the flatness of the Kolmogorov-Smirnov

distance  $D$  in the vicinity of the optimal solution in method 2 implies the same. We assess that (if we are to believe that the tail is truly distributed according to a power law) the exponent lies somewhere between  $\gamma = 1.7$  and  $\gamma = 2.1$ . Other heavy-tailed candidate distributions also provide equally good (or better) descriptions of the data. The only concrete conclusion we make regarding the tail is that it is indeed heavy (i.e. non-exponential). Nonetheless, we believe that the simple power-law description has value as it provides a simple single-parameter ( $\gamma$ ) interpretation of the decay of the tail.

### 8.3 Criticality of the Inferred PLM Model

In the previous section, we characterised the structure of the inferred coupling network, identifying symmetries between L and R hemispheres, and that the distribution of couplings is centred on a small negative value of  $J_{ij}$  but contains a heavy (power-law like) tail for positive  $J_{ij}$ . We now investigate the statistical physics of our inferred model. As a reminder, the original motivation for our work was the observation that complex systems derive a number of functional and computational advantages from operating at or near a critical point (i.e. near a second-order phase transition), see chapter 4. We are therefore particularly interested in characterising whether or not the resting state network we have inferred here, in a statistical physics sense, is positioned near a critical point. We will find that, as for the mindfulness data, the inferred model corresponds to a paramagnetic super-critical state-point, that is close to but offset from the phase transition. The ground state (i.e. low-temperature limit) of our model is a spin-glass (SG), and we will find that the small negative couplings play a vital role in ensuring that the model correctly reproduces the correlations of the data.

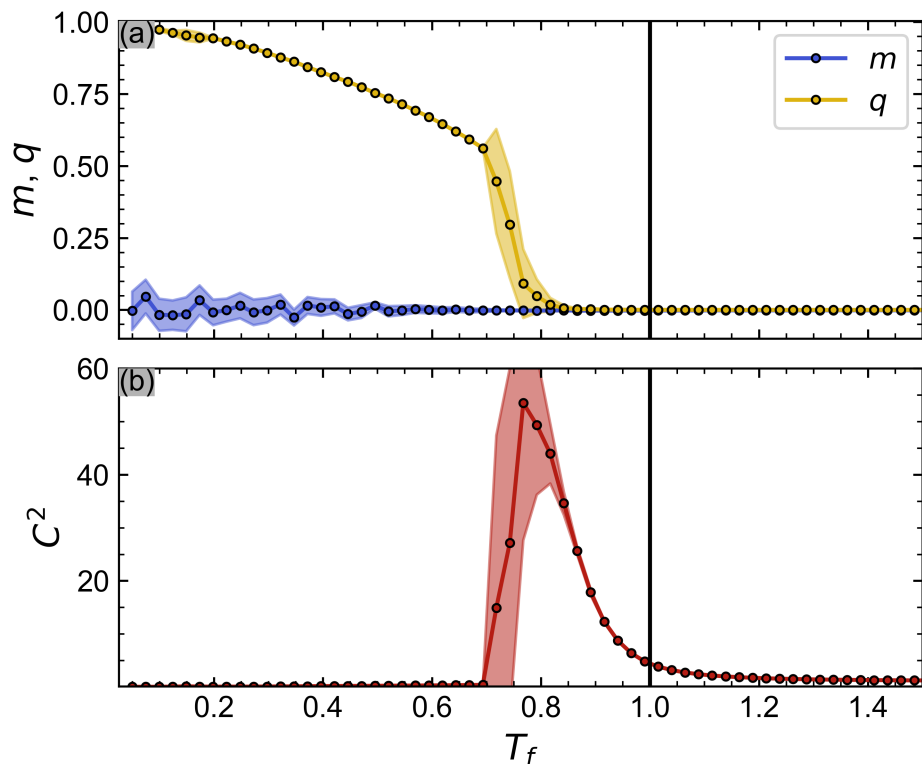
#### 8.3.1 Fictive Temperature Sweep

Following the steps of chapter 7, we will assess the proximity of our inferred model to a phase transition by introducing a fictive temperature  $T_f$ , which we vary to simulate different state points close to our model. Given  $\theta^*$  is the output model we receive from PLM, we will perform equilibrium Monte Carlo (MC) simulations of the modified model  $\frac{1}{T_f}\theta^*$  for each fictitious temperature<sup>20</sup>. The MC parameters for each simulation were  $p_{mc} = (10^5, 10^5, 10)$ , so that we produced  $B = 10^4$  samples from each simulation. We measure the ferromagnetic order parameter  $m$ , the spin-glass order parameter  $q$ , and the critical fluctuation measure  $C^2$  from these datasets. We performed 12 independent simulations for each  $T_f$ , and all presented results (errors) are averages (standard errors) over these independent simulations.

The calculated observables from this fictive temperature sweep are shown in Fig. 8.11: panel (a) tracks the order parameters  $m$  and  $q$ , while panel (b) shows the susceptibility  $C^2$ . The location of the “natural” PLM model (i.e. the unmodified model where  $T_f = 1$ ) is marked by the black vertical line. Panel (a) shows that although  $q$  varies from 1 at low  $T_f$  to 0 at high

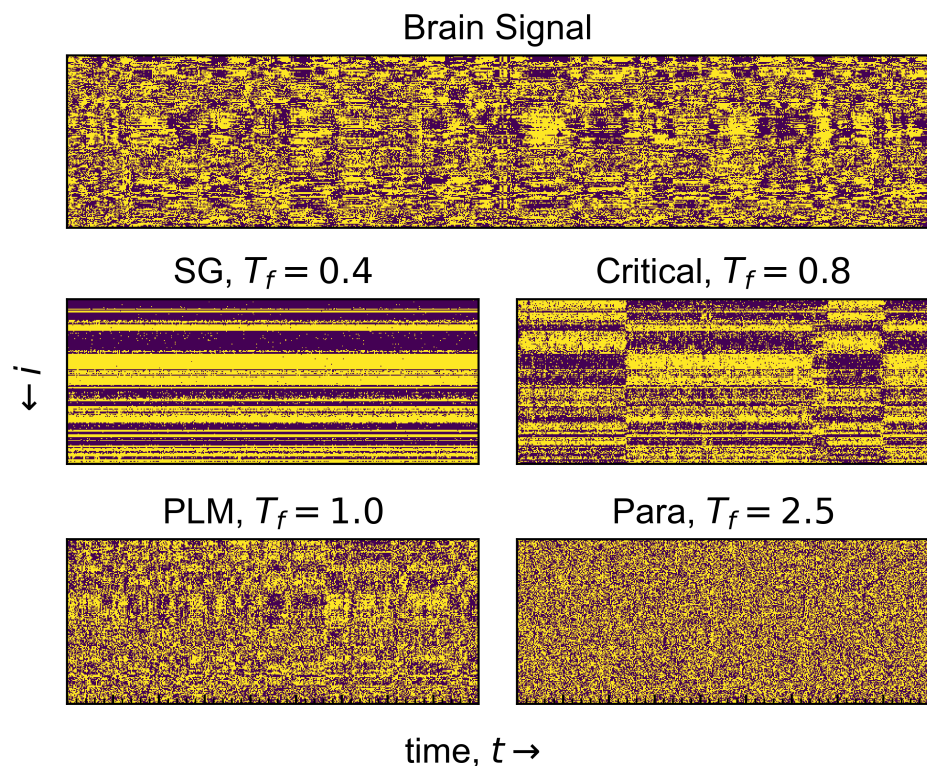
<sup>20</sup> The modified coupling distribution is therefore sharpened if  $T_f > 1$  and spread if  $T_f < 1$ .  $T_f = 1$  is equivalent to the “natural” PLM model of the resting-state data.

**Figure 8.11:** Proximity of the PLM model inferred from the HCP dataset to criticality. The top panel (a) shows the dependence of the ferromagnetic order parameter  $m$  and spin-glass order parameter  $q$  on the fictive temperature  $T_f$ . The bottom panel (b) shows the same for the susceptibility measure  $C^2$ . The originally inferred PLM model representing the binarized fMRI data corresponds to  $T_f = 1$  and is marked by a black vertical line. We observe that the model undergoes a transition from a paramagnet to a spin-glass at  $T_f \approx 0.8$ . The inferred PLM model corresponds to a super-critical paramagnetic state point. Shaded regions represent errors of 12 independent MC simulations.



$T_f$ ,  $m = 0$  throughout. The order-parameter  $m$  measures global alignment (i.e. whether *all* spins point in the same direction throughout the trajectory). In contrast,  $q$  measures individual alignment (i.e. whether each spin remains pointed in a single direction throughout the trajectory). The mismatch between these two thus tells us that the ground state (the low  $T_f$  limit) of the PLM model is a spin-glass, i.e. a system where there is no global ordering but where each spin remains frozen in a particular orientation throughout the trajectory. Tracking the observables from low  $T_f$  to high  $T_f$ , we further observe a single peak in  $C^2$  at  $T_f \approx 0.8$ , which coincides with a rapid, continuous drop in  $q$ . These two features together imply that the system undergoes a phase transition in the range of  $T_f$  considered here, with the critical temperature of the transition corresponding approximately to  $T_c \approx 0.8$ . For  $T_f > T_c$  both  $m = 0$  and  $q = 0$ , these unordered state points are paramagnetic. The PLM model at  $T_f = 1$  corresponds to a super-critical paramagnetic state-point. We find  $C^2(T_f = 1)/C^2(T_f = T_c) \approx 0.1$  so that the PLM model is within the regime of non-trivial enhanced fluctuations that surrounds the transition.

To gain a better understanding of how these phases appear in data, we plot sections of trajectories of the empirical brain signal and our MC simulations for several relevant temperatures in Fig. 8.12. We stress that the (to us unknown) dynamics of the brain signal are different from those of the MC simulations, but that it is nonetheless still interesting to compare the two to examine the temporal correlations of our simulated model. For the SG state point ( $T_f = 0.4$ ), we clearly see that there is no overall global ordering of the spins, but that each spin appears preferentially frozen in a given direction. Interestingly, there appear to be bands within the data, hinting that the couplings encode a clustered structure of highly connected



**Figure 8.12:** Spin trajectory segments from the empirical brain signal and from Monte Carlo simulations of the HCP PLM model at 4 different fictive temperatures. We only plot ROIs in the left hemisphere ( $i = 0$  to  $i = 179$ ) and show  $t = 450$  time points. The ROI index runs vertically, while time is shown horizontally. The simulated dynamics of the PLM model ( $T_f = 1$ ) most closely resemble the real dynamics of the empirical brain signal. We identified spin-glass (SG), critical and paramagnetic state points using Fig. 8.11. Purple pixels are deactivated  $s_i(t) = -1$ , while yellow regions are activated  $s_i(t) = +1$ .

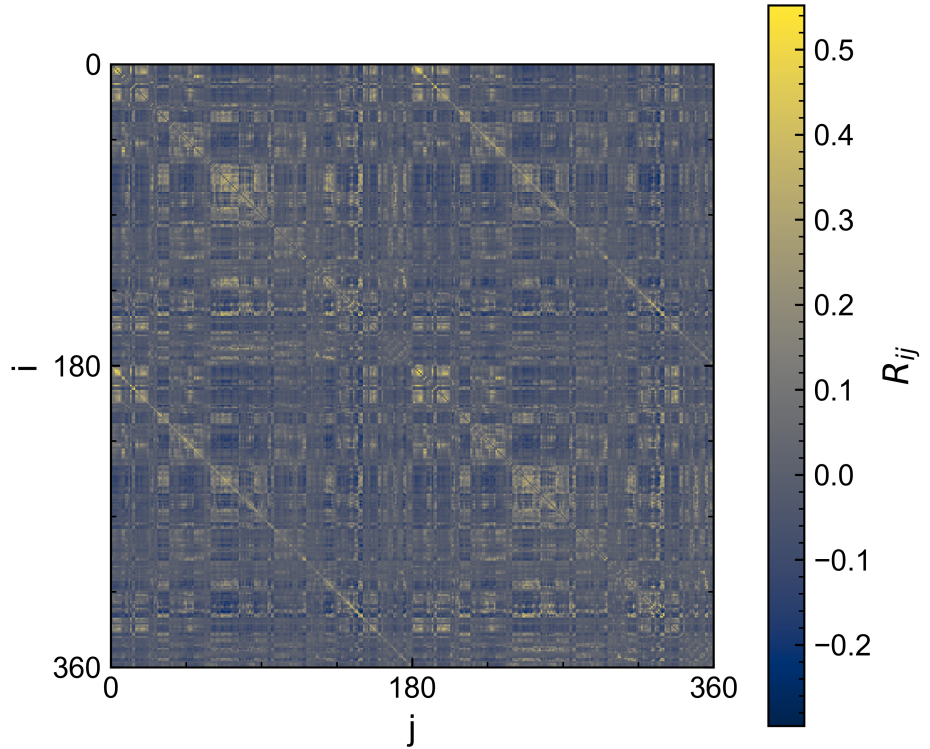
sub-networks. We take this as evidence for the modular organisation of the brain [188]. As the temperature increases, the spins start to unfreeze and fluctuate. At the critical temperature  $T_f \approx 0.8$ , we find large rearrangements involving the entire system<sup>21</sup>. This highly correlated behaviour is exemplified by the observation of a number of system-wide “flipping” events, where the sign of almost all spins in the system simultaneously reverses, within the trajectory except. We still observe the same banding structure, but that activity within the bands now fluctuates between average periods of activity (+1) and inactivity (−1), delineated by the flipping events. At the natural temperature of the PLM model ( $T_f = 1$ ), we no longer find any system-wide rearrangements but continue to observe shorter periods of activity/inactivity for clusters of specific spins. This data is visually most consistent with the empirical binarised fMRI signal recorded from the brain. In the high-temperature limit,  $T_f = 2.5$ , the data resembles Gaussian white noise, with spins fluctuating randomly as expected from a paramagnetic state-point.

In summary, our analysis of the order parameters and susceptibility of the inferred model at different fictive temperatures revealed that the resting-state PLM model of the brain sits near an SG-P phase transition. The PLM model corresponds to a super-critical paramagnetic state-point, which still exhibits a degree of heightened correlations. By examining simulated MC trajectories from the PLM model and the critical point in detail, we find that the PLM model lacks features such as the system-wide cooperation characteristic of a system at criticality. We conclude that, although the correlations in the natural PLM model ( $T_f = 1$ ) are non-trivial, the resting-state brain network itself does not correspond to an equilibrium (statistical physics) critical point. We find that the ground state of the brain (i.e. the

<sup>21</sup> We note that system-wide (i.e. scale-free) cooperation is a key feature of the critical state.



**Figure 8.13:** Pearson’s correlation coefficients  $R_{ij}$  calculated from the aggregate HCP resting-state dataset. The overall correlation structure implies much more connectivity than the coupling structure we infer using our inverse method (see. Fig. 8.4).



low  $T_f$  limit) corresponds to a spin-glass phase.

## 8.4 Thresholding and Network Construction

The general goal of most neuroimaging studies is to understand the connectivity of the brain in the hope that this will reveal how organisation facilitates various cognitive functions. Most commonly, functional connectivity (FC) is established by assessing *correlations* between ROIs, e.g. through the linear (Pearson) correlation coefficient

$$R_{ij} = \frac{C_{ij}}{\sigma_i \sigma_j} = \frac{\langle s_i s_j \rangle - \langle s_i \rangle \langle s_j \rangle}{\sigma_i \sigma_j}, \quad (8.6)$$

which we show for the aggregate HCP dataset in Fig 8.13. Comparing the correlations in Fig 8.13 to the couplings we inferred from the same data in Fig 8.4 shows that correlations generally appear more connected (i.e. denser networks) than the couplings. We also noted the same effect when analysing the “small” dataset in the previous chapter (chapter 7), where we showed that a range of correlation structures could be produced by tuning the inferred PLM model towards the critical point. We thus expect that networks constructed from correlation measures such as (8.6) are likely to over-represent the connectivity of the data. This effect has been explored in detail in Ref [137], where it was shown that simple coupling topologies such as the two 2D nearest-neighbour Ising model, can produce complex correlations that mimic those observed in functional connectivity analyses of neuro-imaging data [137]. Some of the ubiquitous claims about the functional connectivity of the brain established from correlations, such as small world properties and short path lengths [25, 67], may therefore be

the result of much simpler coupling topologies. In our view, inverse methods such as PLM, provide a vital tool with which to discern whether two regions of interest are truly connected or whether they simply co-vary through some tertiary connection<sup>22</sup>, and we explore their use for this application in this section.

Usual correlative functional connectivity neuroimaging analyses begin first by measuring correlations, and then thresholding these to create links (edges) between ROIs (nodes). Values below the threshold  $\delta$  are discarded, while those above  $\delta$  are used to construct the FC network. A variety of schemes are employed when thresholding, with authors constructing networks of positive correlations [66] ( $R_{ij} \geq \delta$ ), negative correlations [132] ( $R_{ij} \leq -\delta$ ) and absolute correlations [62, 133] ( $|R_{ij}| \geq \delta$ ) throughout the literature. These choices are highly contentious [128, 135]. For instance, only considering positive correlations, or correlation magnitudes does not allow for the excitatory/inhibitory nature of the connection to be assessed. Moreover, negative interactions are known to help modulate brain dynamics, and their functional importance has been demonstrated in neuroimaging studies [132, 136]. Discarding them clearly discards an important component of the network. We now wish to apply the same type of threshold to the resting-state couplings  $\mathbf{J}$  we inferred for the HCP dataset. Our analysis will demonstrate the important role negative couplings play in mediating and maintaining correlations within the inferred brain network.

<sup>22</sup> E.g. we can imagine that both regions  $i$  and  $j$  and regions  $i$  and  $k$  are strongly coupled. Regions  $j$  and  $k$  would then be highly correlated, and networks built from correlation measures would associate the two. Inverse methods such as PLM prevent this, and produce an overall sparser functional connectivity

#### 8.4.1 Statistics of Thresholded Coupling Networks

We apply two common thresholding schemes to our inferred PLM model. We again measure  $q$  and  $C^2$  as we vary the threshold  $\delta$  to assess the impact that discarding a given set of couplings has on the model. As a reminder, we found that 2 out of 3 of the inferred HCP couplings were negative, but that the distribution of couplings itself is skewed, with a heavy tail for positive  $J_{ij}$ . We will investigate two thresholding schemes, which emphasise different properties of the network. The positive thresholding scheme

$$J_{ij} = \begin{cases} J_{ij}, & \text{if } J_{ij} \geq \delta \\ 0, & \text{otherwise} \end{cases}, \quad (8.7)$$

discards all negative couplings from the network, and is equivalent to constructing a network containing only the excitatory couplings. The symmetric thresholding scheme

$$J_{ij} = \begin{cases} J_{ij}, & \text{if } |J_{ij}| \geq \delta \\ 0, & \text{otherwise} \end{cases}, \quad (8.8)$$

discards all “small” couplings, and therefore builds a network where only the most important couplings are retained. Note that we do not set  $J_{ij} = 1$ , i.e. we keep the weights of the couplings. We now vary the threshold  $\delta$  and perform MC simulations of the thresholded models. The parameters of these simulations were  $p_{mc} = (10^5, 10^5, 10)$ , producing  $B = 10^4$  samples from which we calculated  $q$  and  $C^2$ . We again performed 12 independent repeat simulations for each  $\delta$  and report the averages over these runs. As we previously noted a power-law tail, we chose to vary  $\delta$  logarithmically.

**Figure 8.14:** We track the dependence of the spin-glass order parameter  $q$  (panel a), the susceptibility measure  $C^2$  (panel b) and the remaining number of edges in the network (panel c) as functions of the threshold  $\delta$ . Results for the symmetric thresholding scheme are shown in blue, while the positive thresholding scheme is shown in yellow. We identify 3 threshold values of special importance.  $\delta_1$  shows the largest  $\delta$  for which  $q$  and  $C^2$  of the symmetrically thresholded model agree with those of the fully connected PLM model.  $\delta_2$  shows where there is a peak in  $C^2$  for the positive thresholding scheme.  $\delta_3$  denotes the onset of network fragmentation (i.e. the largest value of  $\delta$  for which all ROIs form a single connected network). Above this multiple fragmented sub-networks begin to form (grey region).

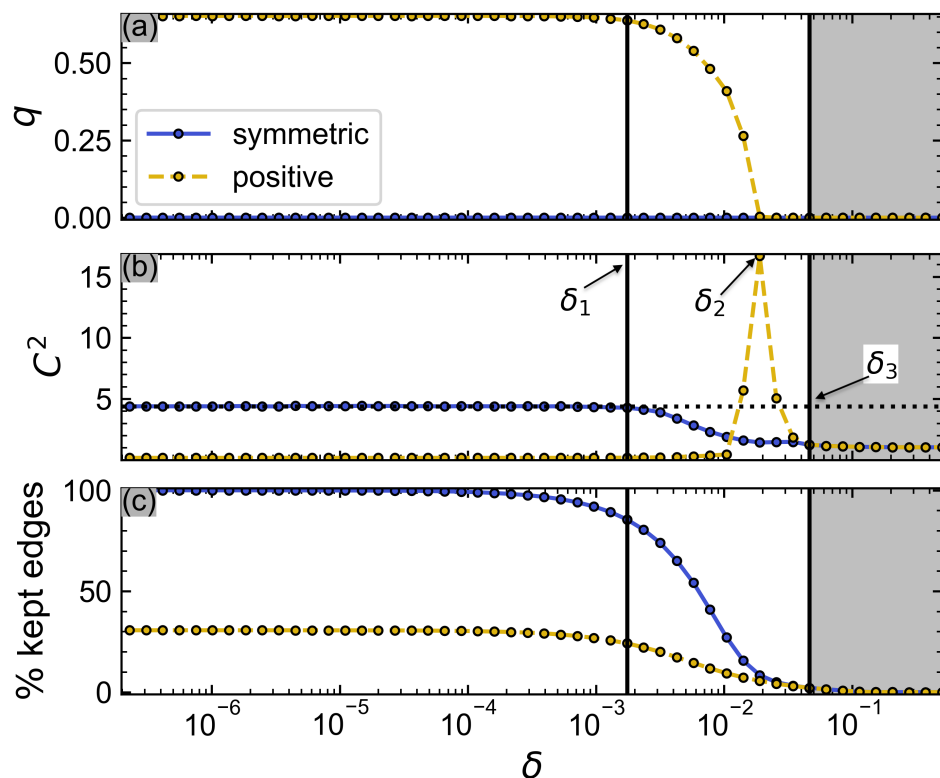


Fig. 8.14 shows the results of this analysis. Panels (a) and (b) show  $q$  and  $C^2$  respectively. In panel (c) we track the % of edges (i.e. number of non-zero  $J_{ij}$ ) present in the thresholded network compared to the initial fully connected PLM model.

**Assessing the two thresholding schemes in the limit  $\delta \rightarrow 0$ .** For the symmetric scheme, this corresponds exactly to the unmodified PLM model, and we see good agreement with  $q$  and  $C^2$  when  $\delta$  is small. But for the positive scheme, by design, the threshold automatically discards the majority ( $\sim 66\%$ ) of couplings as they were negative. Panels (a) and (b) show that sampling from the positively thresholded models distorts the phase of the system, indeed both  $q$  and  $C^2$  are far from the values observed for the un-thresholded PLM model. These correspond to a highly coupled ferromagnetic-like state-point, where due to a lack of connections, some nodes continue to fluctuate, leading to  $q \neq 1$ . Comparing this behaviour with the trajectory excerpts of the empirical brain data and the simulated PLM data Fig. 8.12 shows that the statistics of these positively thresholded networks do not correspond to those of the data. Although the negative couplings in the model are on-average smaller than the positive couplings (see Fig. 8.7), this highlights the importance negative couplings play in mediating correlations within the system. Negative correlations are required to maintain the state-point of the system, and network analyses of the positive connections alone are therefore unlikely to be representative of the true connectivity of the brain.

**Assessing the two thresholding schemes when  $\delta \neq 0$ .** Until  $\delta \approx 10^{-3}$ , we observe little change in  $q$ ,  $C^2$  or the edge density. We then identify 3 key threshold values at which significant changes in behaviour occur. The first of these,  $\delta_1$ , denotes the largest  $\delta$  for which the correlation measure  $C^2$  of the symmetrically thresholded model is in agreement with  $C^2$  measured from simulations of the un-modified PLM model<sup>23</sup>. Interestingly, the number of retained edges in the symmetrically thresholded model at  $\delta_1$  is around 86%, showing that 14% of couplings can be discarded (i.e. set to 0) without loss. Similarly to the self-consistency correction we proposed for PLM in chapter 6,  $\delta_1$  allows us to define a self-consistent threshold which conserves the correlations of the data (i.e. matches  $C^2$ ). This circumvents another critique of the thresholding process, which is that the choice of  $\delta$  is usually an arbitrarily defined hyperparameter. For  $\delta > \delta_1$ , we find that both  $C^2$  and the edge density monotonically decrease for the symmetric scheme; as we reduce the number of connections in the network, the overall coupling strength decreases and the data becomes more paramagnetic. On the contrary, models obtained from the positive thresholding scheme appear to undergo a phase transition, and we find a peak in  $C^2$  at  $\delta_2$ . For now, we explain this by the following reasoning. Increasing  $\delta$  decreases the number of couplings in the system, which lowers the overall coupling strength of the model. As the positively thresholded model at  $\delta = 0$  is in a ferromagnetic-like state, there must therefore be a value of  $\delta$  where the overall coupling strength becomes weak enough for the system to “melt” and transition to a paramagnetic state. As shown in Fig. 8.11, in the  $\delta \rightarrow 0$  limit, the symmetric model (i.e. the raw PLM model) already corresponds to a paramagnetic state-point. Increasing  $\delta$  further can only turn the system more paramagnetic. We explore the inter-dependence between  $\delta$  and  $T_f$  further in section 8.4.2. Interestingly, the location of the peak for the positive scheme corresponds to a plateau in  $C^2$  for the symmetric scheme. This again highlights that negative couplings are essential in controlling the system’s statistics. The point  $\delta_3$  denotes the largest value of  $\delta$  for which we retain a single network<sup>24</sup>. The network fractures into sub-networks for  $\delta > \delta_3$ , and we shade this regime in grey in Fig. 8.14.

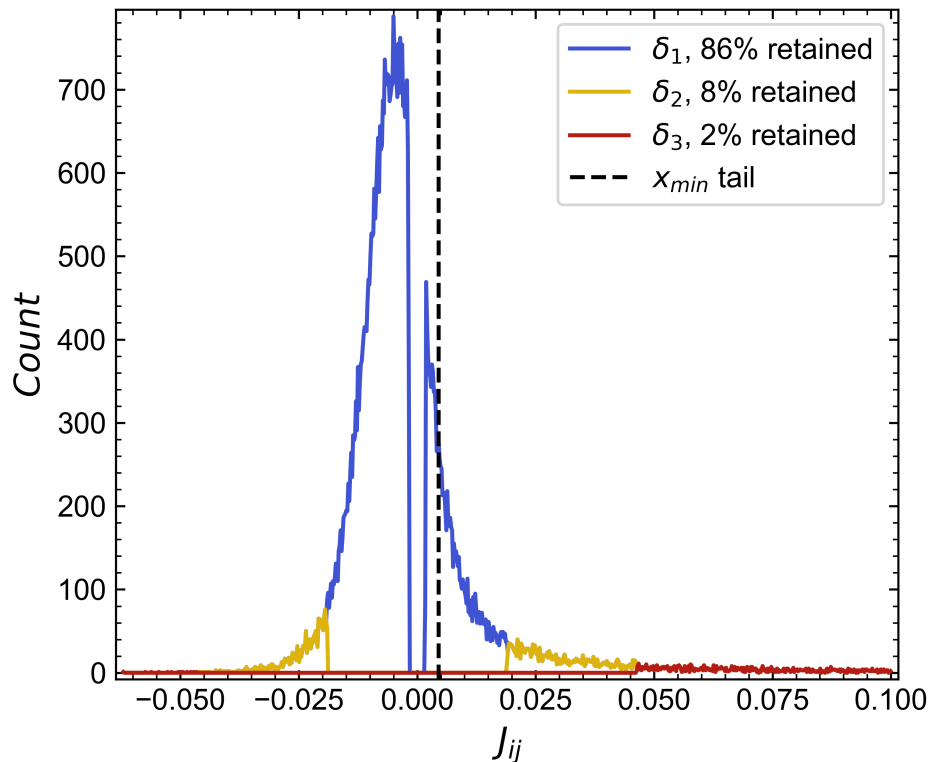
In summary, networks constructed by discarding all negative (i.e. “positive thresholding”) interactions fail to capture the statistics of the binarised fMRI input data. Network analyses which remove negative couplings will therefore not capture the true connectivity of the brain. By thresholding symmetrically (i.e. removing all  $J_{ij}$  for which  $|J_{ij}| < \delta$ ) we show that up to 15% of the smallest couplings may be removed without impacting the statistics of the data. This provides us with a self-consistent route towards identifying an appropriate threshold  $\delta$ , couplings below which may safely be viewed as artefacts of the inference and removed without issue. The remaining network is still close to fully connected ( $\sim 86\%$  out of the original fully connected  $(360 \times 359)/2 = 64,620$  coupling are retained). We find that the last  $\delta$  for we retain a single connected component (i.e. for the network does not fragment into sub-networks) occurs at a large  $\delta = \delta_3 = 0.04$ . The strongest couplings in the network form a sufficient “backbone” of connectivity through which information can pass from any node to any other node. We briefly mentioned that varying  $\delta$  induced a phase transition in the positive thresholding scheme. We attempt to formalise this observation

<sup>23</sup> Note that we defined this by measuring whether  $C^2$  for the symmetrically thresholded model was within 5% of the un-modified PLM model.

<sup>24</sup> I.e. in which a path exists between all  $N = 360$  ROIs.



**Figure 8.15:** Coupling distributions obtained after applying a symmetric threshold to the HCP PLM model for the three values of  $\delta$  shown in Fig. 8.14. Blue shows  $\delta_1$ , orange shows  $\delta_2$  and red shows  $\delta_3$ . We note that for  $\delta_1$  the distribution is dominated by small negative couplings. The legend further shows how many of the original couplings each of these thresholded models retain. The approximate value of the onset of the heavy tail (see Table. 8.1) is shown by the dashed vertical line.



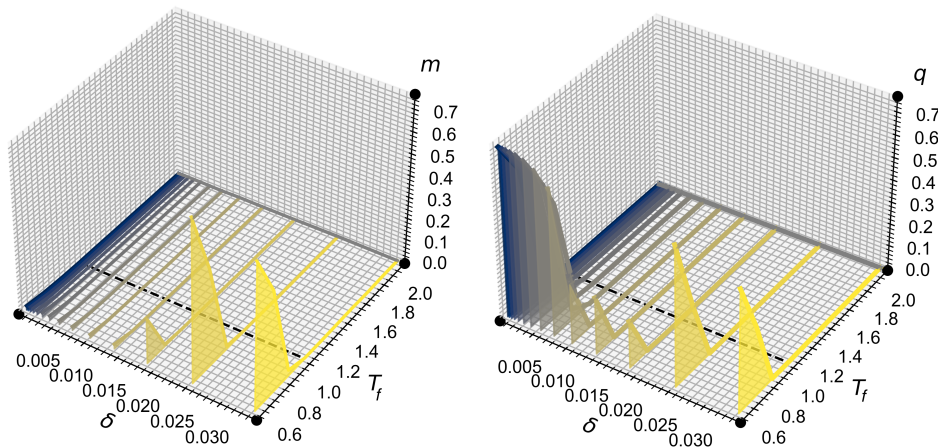
in the next section.

## 8.4.2 Thresholds and Phase Transitions

Although the spike at  $\delta_2$  for the positive scheme appears like a second-order phase transition in  $\delta$ , we have no reference values of  $C^2$  to establish whether  $C^2(\delta_2) \approx 15$  is truly maximal for this particular  $\delta$ . We will now investigate this formally by re-introducing the fictive temperature and simulating each thresholded model for a range of  $T_f$ , in a similar process as in Fig. 8.11. We again perform 6 independent Monte Carlo (MC) simulations, using  $p_{mc} = (10^5, 10^4, 10)$  to collect  $B = 10^3$  samples, for each  $(\delta, T_f)$  pairing and present averages over these. We acknowledge that these are relatively short simulation times, and that future work may aim to increase these to improve the reliability of our conclusions. Nonetheless, we believe that these sampling conditions are good enough to provide an illustrative example of the phase behaviour of the thresholded models. We begin by looking at the symmetric thresholding scheme.

### Fictive Temperature Sweeps for Symmetric Thresholds

We begin by showing how thresholding at the three key values ( $\delta_1$ ,  $\delta_2$  and  $\delta_3$ ) identified in Fig. 8.14 impacts the distribution of couplings in the symmetric thresholding scheme in Fig. 8.15. As a reminder, this scheme corresponds to setting all  $|J_{ij}| < \delta$  to zero. For  $\delta_1$  we again see that, by count, the distribution is dominated by negative couplings, and that this is the last threshold value for which the thresholded model reproduces the statistics of the input data. Increasing  $\delta$  places a greater emphasis on the positive



**Figure 8.16:** The ferromagnetic order parameter  $m$ , and the spin-glass order parameter  $q$  as functions of  $\delta$  and  $T_f$  for the symmetric thresholding scheme. Projection (grey) shows the dependence of  $m$  and  $q$  on  $\delta$  at  $T_f = 1$  (shown by the black dashed line), i.e. the trace shown in Fig. 8.14. As  $\delta$  increases the transition morphs from a spin-glass-paramagnetic transition to a ferromagnetic-paramagnetic transition. Plotted results are averages from 6 independent MC simulations.

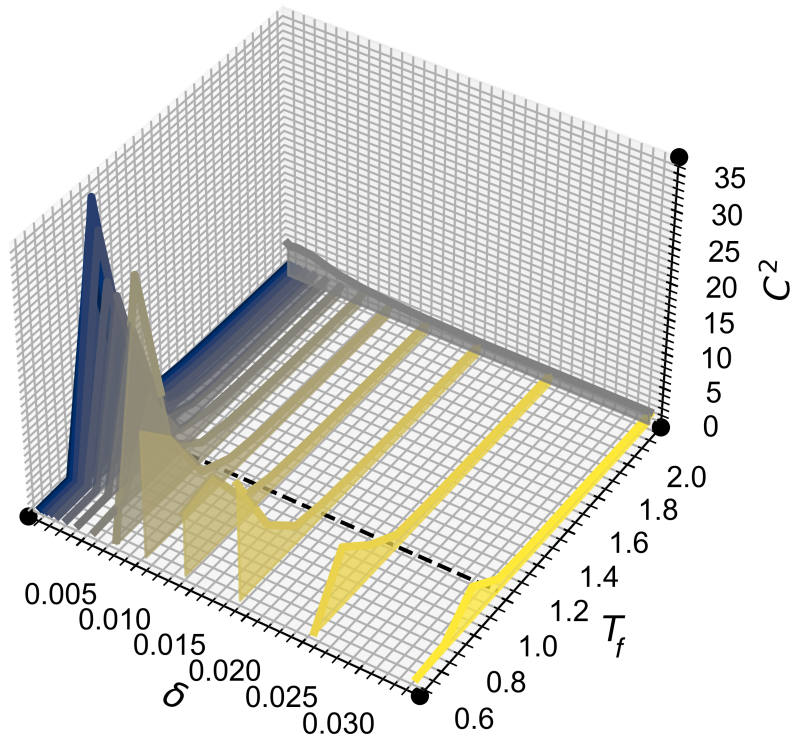
couplings in the tail as the distribution is skewed, with  $\delta_3$  composed almost exclusively of large positive couplings. We show the approximate value of  $x_{\min}$  defining the start of the heavy tails for reference (see section 8.2.4).

In Fig. 8.14 we found that increasing  $\delta$  monotonically decreased  $C^2$ . We stated that this was due to the raw PLM model already corresponding to a paramagnetic state-point, and so weakening the interactions within this model could only shift the state-point further into the paramagnetic regime. We now expand on this concept by performing the aforementioned temperature sweeps for a variety of  $\delta$ s. The calculated values for  $m$  and  $q$  are shown in Fig. 8.16, while those for  $C^2$  are shown in Fig. 8.17. For small  $\delta$ , these simulations reinforce our view that the HCP PLM model (regardless of the chosen threshold) is paramagnetic, as we measure a peak in  $C^2$  at  $T_f < 1$  for all  $\delta$ . The dependence of  $m$  and  $q$  on  $\delta$  further shows that increasing  $\delta$  changes the ground state of the model (i.e. the low  $T_f$  limit), from a spin-glass ( $m = 0$ ,  $q \neq 0$ ) to a ferromagnet ( $m \neq 0$ ,  $q \neq 0$ ). This is a result of the positive skew of the coupling distribution; as  $\delta$  increases we begin considering coupling distributions which are dominated by one sign (i.e. positive couplings), in turn destroying the quenched disorder required to form a spin-glass. An interesting side effect of this is that the peak of  $C^2$  along each  $T_f$  plane decreases with  $\delta$ . This is because values of  $C^2$  associated with the spin-glass paramagnetic transition are larger than those associated with the ferromagnetic paramagnetic transition (see e.g. Fig. 5.21 of section 5.6.3). In summary, the transition behaviour of each thresholded model has a complex dependence on  $\delta$ . Networks constructed from arbitrary choices of  $\delta$  are therefore likely not only to mischaracterise the statics of the data but can also lead to the incorrect identification of the ground state associated with the model. We further find that the height of the peak of  $C^2$  depends on  $\delta$  in a non-trivial way, meaning that directly comparing values of  $C^2$  at different  $\delta$  is not a good metric for establishing proximity to a phase transition.

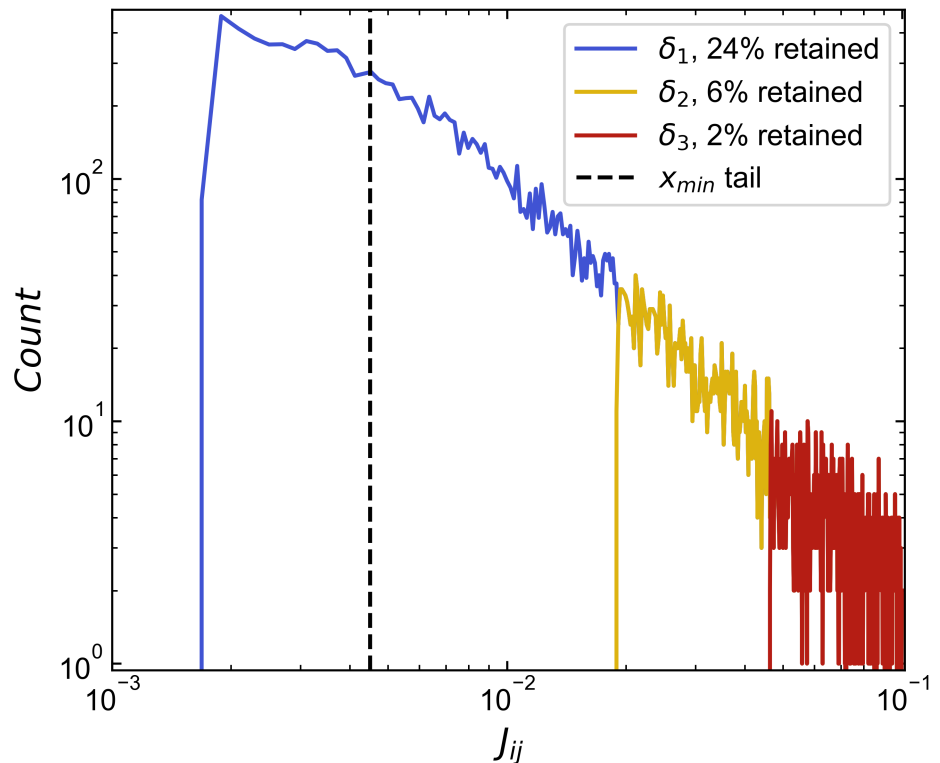
### Fictive Temperature Sweeps for Positive Thresholds

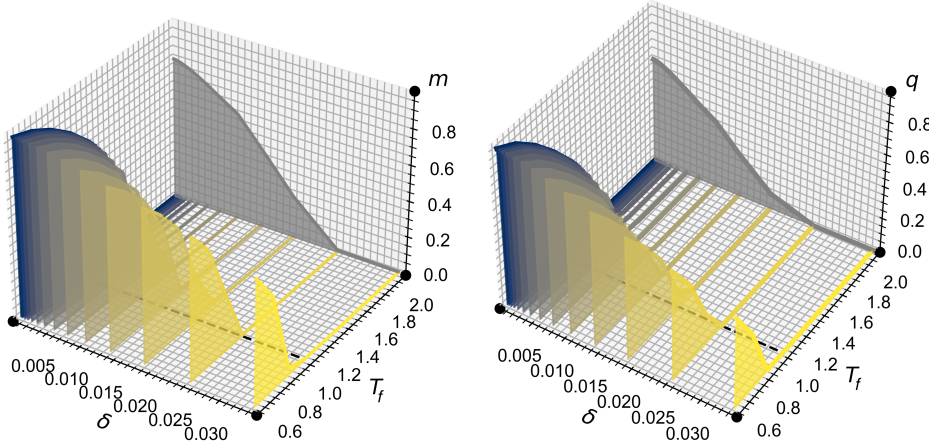
We now repeat the above analysis for the positive thresholding scheme, where again for clarity, we set all  $J_{ij} < \delta$  to zero. This is equivalent to constructing a network representing the strongest excitatory functional con-

**Figure 8.17:** The susceptibility measure  $C^2$  as a function of  $\delta$  and  $T_f$  for the symmetric thresholding scheme. Projection (grey) shows the dependence of  $C^2$  on  $\delta$  at  $T_f = 1$  (shown by the black dashed line), i.e. the trace shown in Fig. 8.14. Reduction in size of the  $C^2$  peak as  $\delta$  increases is the result of the changing ground state (from spin-glass to ferromagnetic) of the model. Plotted results are averages from 6 independent MC simulations.



**Figure 8.18:** Coupling distributions obtained after applying the positive thresholding scheme to the HCP PLM model for the three values of  $\delta$  shown in Fig. 8.14. We plot these on a log-log axis as the symmetric thresholding scheme discards all negative couplings, essentially constructing a network consisting only of the heavy tail discussed in section 8.2.4. Blue shows  $\delta_1$ , orange shows  $\delta_2$  and red shows  $\delta_3$ . The legend further shows how many of the original couplings each of these thresholded models retain. The approximate value of the onset of the heavy tail (see Table. 8.1) is shown by the dashed vertical line.





**Figure 8.19:** The ferromagnetic order parameter  $m$ , and the spin-glass order parameter  $q$  as functions of  $\delta$  and  $T_f$  for the positive thresholding scheme. Projection (grey) shows the dependence of  $m$  and  $q$  on  $\delta$  at  $T_f = 1$  (shown by the black dashed line), i.e. the trace shown in Fig. 8.14. Transitions found for all  $\delta$  are ferromagnetic-paramagnetic. Plotted results are averages from 6 independent MC simulations.

nections in the brain. For the positive thresholding scheme, increasing  $\delta$  corresponds to selecting different components of the positive tail of the distribution. We show the coupling distribution for 3 example values of  $\delta$  along with  $x_{\min}$  in Fig. 8.18.

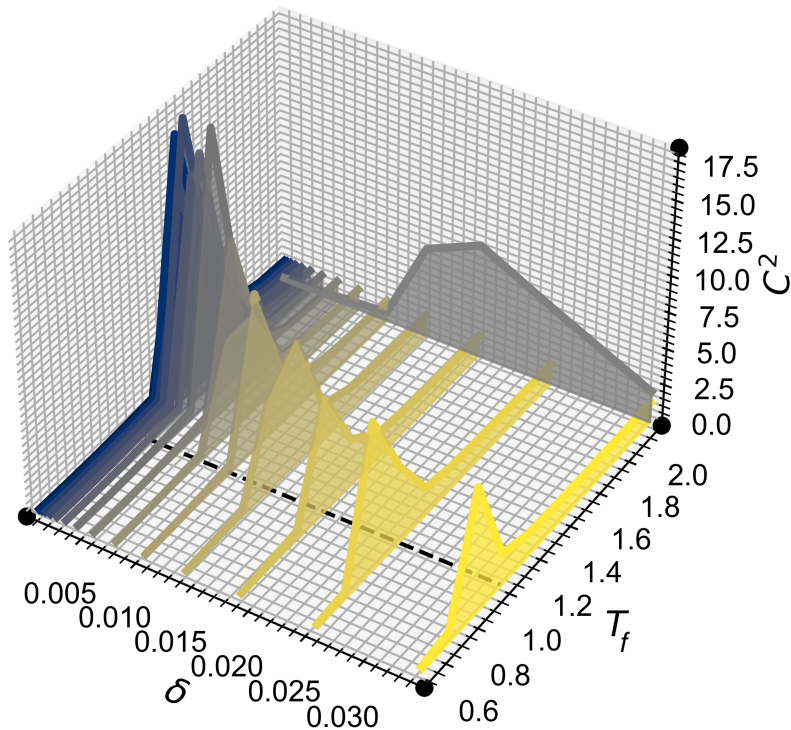
We plot  $m$  and  $q$  for the positive thresholding scheme for  $\delta$  and  $T_f$  in Fig. 8.19. In contrast to the symmetric scheme, we find that both  $m$  and  $q$  are similar ( $m \propto q$ ) for all  $\delta$  and that in the small  $T_f$  limit both  $m \neq 0$  and  $q \neq 0$ , implying that the ground state of all positively thresholded models is a ferromagnet. As before, Fig. 8.20 shows the dependence of  $C^2$  on  $\delta$  and  $T_f$ . Comparing the limits of the axes in Fig. 8.20 and Fig. 8.17 shows that the phase transitions of the positively thresholded models have smaller maximal values of  $C^2$ . This is in line with our previous argument concerning the values of  $C^2$  at the SG-P and F-P transition lines in the SK model. Moreover, we find that as  $\delta$  increases the temperature corresponding to the peak of  $C^2$  crosses the line  $T_f = 1$ , leading to the  $C^2$  dependence we observe in Fig. 8.14. This means that when applying the positive thresholding scheme one can always identify a value of  $\delta$  at which  $C^2$  is maximised at  $T_f = 1$ , i.e. that corresponds exactly to a (finite sized) thermodynamic critical point. The threshold  $\delta_2$  identified in Fig. 8.14 is close to this value. The average peak value of  $C^2$  decreases as  $\delta$  increases. This reinforces our finding that values of  $C^2$  at different threshold levels cannot be compared directly to assess proximity to the phase transition and that an investigation of e.g.  $T_f$  is required to contextualise whether a given value of  $C^2$  is maximal. In the limit of large  $\delta$  both thresholding schemes converge to models exhibiting similar thermodynamics (paramagnetic models with ferromagnetic ground states), with exact equality occurring once  $\delta > |\min(J_{ij})| \approx 0.05$ .

### 8.4.3 Thresholding Analysis Summary

Removing the negative interactions from the network (somewhat unsurprisingly) fundamentally alters the statistical physics of the originally inferred PLM model. As discussed, such thresholding schemes are however still commonly employed to extract network topologies within the neuroimaging literature [128]. We performed a detailed analysis of the statistical physics of both symmetrically thresholded (retaining coupling of high magnitude) and positively thresholded (keeping only positive couplings) PLM models. We



**Figure 8.20:** The susceptibility measure  $C^2$  as a function of  $\delta$  and  $T_f$  for the positive thresholding scheme. Projection (grey) shows the dependence of  $C^2$  on  $\delta$  at  $T_f = 1$  (shown by the black dashed line), i.e. the trace shown in Fig. 8.14. The maximum value of  $C^2$  is  $\approx 1/2$  that is seen in the symmetric scheme, see Fig 8.17. As  $\delta$  increases  $T_c$  (which we define by the peak in  $C^2$ ) passes through  $T_f = 1$  so that there exists a value of  $\delta$  at which the model is perfectly at the ferromagnetic-paramagnetic critical point. Plotted results are averages from 6 independent MC simulations.



established that the symmetric thresholding scheme allowed us to identify a threshold value  $\delta$  which preserved self-consistency with the empirical correlations in the input data, while also removing around 15% of the couplings. This model corresponds to a paramagnetic state. All networks constructed from only positive couplings fail to preserve the correlations within the input data and change the ground state of the model from a spin-glass to a ferromagnet. This highlights the importance of negative couplings in mediating correlations within the brain. Complex networks which disregard these are therefore unlikely to represent the true connectivity of the brain. Our work further shows that for positive threshold models there always exists some  $\delta$  at which the model is perfectly tuned to a ferromagnetic-paramagnetic critical point. Claims of criticality arising from thresholded models should therefore be scrutinised thoroughly.

## 8.5 Conclusion

In this chapter, we applied pseudo-likelihood maximisation (PLM) to infer the functional connectivity of the brain from the Human Connectome Project resting-state fMRI dataset. We established that this dataset is large enough to minimize contributions from the inference bias by sub-sampling the data and re-performing the inference. The connections within the extracted PLM model are predominately negative, although the distribution is skewed with a heavy, non-exponential positive tail. We find that this tail can be approximately described by a power law, with an exponent in the range  $\gamma = 1.7$  to  $\gamma = 2.1$ . The PLM model further reveals strong symmetry between the left and right cortical hemispheres. We find a general

lack of inter-hemisphere connectivity, with the only exception being regions corresponding to the same label, e.g. L-V1 with R-V1, which are also on average the strongest connections in the network. This suggests a functional organisation whereby both hemispheres act “copies” of each other which activate simultaneously but are otherwise independent. We stress that the functional connectivity implied by the couplings of PLM is much sparser than that of correlative measures, such as the linear correlation coefficient. Thresholded PLM models are therefore less likely to re-produce small-world features such as short path lengths ubiquitously reported in previous functional connectivity analyses, although more work is needed to establish this numerically. We were also able to probe whether the resting state corresponded to a critical state, as the PLM representation of the brain is equivalent to an equilibrium model in statistical physics. We found that, although non-trivial fluctuations are present in the data, the resting-state brain does not operate at a critical point, instead corresponding to a super-critical paramagnetic state point. We also investigated how removing weak couplings (i.e. thresholding) from the model impacted the thermodynamics of the system. We found that schemes which remove all negative couplings (i.e. those that only consider excitatory connections) cannot reproduce the correlations of the empirical input data. However, applying a symmetric threshold (i.e. extracting the most impactful inhibitory and excitatory connections) allowed us to remove  $\sim 15\%$  of couplings without altering the state-point of the model. These results together highlight the essential role negative couplings play in mediating correlations within the brain.

## Summary of Chapter 8

- We applied pseudo-likelihood maximisation (PLM) to the young adult dataset from the Human Connectome Project (HCP). This is one of the highest quality openly available resting-state functional magnetic resonance imaging (fMRI) datasets.
- We performed a sub-sampling analysis of this dataset and concluded that the effect of the small sample size bias was negligible. We thus believe that our inferred model is truly representative of the data.
- We find that the inferred PLM network (i.e. the functional connectivity) of the resting human mind has the following features:

### **Hemispherical symmetry**

Connections between regions of interest within the left and right cortical hemispheres are highly symmetric. This is evident as the strongest couplings in the network are those connecting self-similar regions, e.g. the couplings between the left visual 1 region and the right visual 1 region.

### **Low inter-hemisphere connectivity**

Bar self-similar connections, the functional connectivity between cortical hemispheres is weak. The inter-hemisphere couplings are on average weaker than the intra-hemisphere couplings.

### **Couplings are predominantly small and negative (inhibitory)**

Two out of three couplings within the network are negative.

### **The distribution of positive (excitatory) couplings is heavy-tailed**

The positive couplings form a power law like heavy tail. We approximate the exponent of this power law as  $\gamma \approx 1.7$  to  $2.1$ .

- We analysed the proximity of the PLM model to a phase transition by performing Monte Carlo simulations of the model at a range of *fictive temperatures*.

### **The HCP dataset is paramagnetic and close to criticality**

We find that the PLM model of the HCP dataset corresponds to a paramagnetic state point. This state point is close to a spin-glass-paramagnetic phase transition and exhibits non-trivial correlations.

- We thresholded the inferred PLM model to emulate standard procedures for network analyses in neuroimaging studies. We find that:

### **Negative couplings mediate correlations**

Thresholding schemes which discard all negative couplings are unable to reproduce the correlations of the input data. We argue that thresholding schemes which discard these discard an essential feature of the network.

### **There is a self-consistent symmetric threshold**

Applying a symmetric thresholding scheme to the couplings allows us to extract a simplified representation of the model (with approximately 15% of connections discarded) that conserves the correlations found within the data.



# Chapter 9

## Conclusions

Increasingly, biologists are able to collect “big data” from experiments [7, 29], allowing us to probe complex systems at resolutions that were previously unobtainable. Inverse methods, such as the pairwise maximum entropy modelling framework investigated here, provide a valuable tool set with which to digest these datasets, and allow us to extract and examine interactions that are not directly measurable from the data. Maximum entropy modelling further allows us to assess these interactions within the established framework of statistical physics and provides a tool for network reconstruction when up-to-pairwise interactions are considered. In this thesis, we performed an in-depth characterisation of a particular implementation of pairwise maximum entropy modelling termed pseudo-likelihood maximisation (PLM), which is considered a state-of-the-art solution. We then applied PLM to datasets obtained from neuroimaging studies, constructing spin-glass models of these systems. Our over-arching aim was to use PLM to contribute to the “critical brain hypothesis”, which asserts that the brain derives some of its advanced computational properties from operating near an order-disorder phase transition. We found that no previous studies had systematically established how models inferred via PLM depend on the size of the input dataset and aimed to address this gap in knowledge. We deemed this important as understanding how the inference is affected by limited dataset sizes is essential if one wishes to properly interpret fitting results from real-world experimental data, where sample numbers are naturally limited.

We began by characterising PLM on simulated datasets from the Sherrington - Kirkpatrick (SK) spin-glass model in chapter 5. We demonstrated that estimates of important physical quantities (such as the temperature) which define the state of the inferred model depend linearly on the inverse sample number  $1/B$ , similar to the standard bias of maximum likelihood estimators. This causes models inferred from paramagnetic datasets to exhibit enhanced critical fluctuations, with this effect worsening with decreasing  $B$ . Paramagnetic data is therefore misclassified as near-critical for finite  $B$ , i.e. PLM underestimates the distance from criticality. The inference error is minimized close to but offset from the phase transition in the paramagnetic regime. The development of strong correlations on the approach to the critical point means that PLM fails due to separation at lower  $T$ . We note that, although information-theoretic arguments suggest otherwise [204], for

small or intermediate  $B$  the regime of failure occurs *before* the finite size critical temperature is reached.

In chapter 6, we describe data-driven approaches to mitigate these effects. The self-consistent correction we propose improves the temperature estimate while matching the critical fluctuations  $C^2$  of the dataset. It performs well when a PLM solution can be found, i.e. when separation does not occur. Firth’s penalized logistic regression may be used to estimate the state point when standard PLM fails due to separation. Although Frith’s correction produces reasonable parameter estimates at low temperatures, we caution that the critical fluctuations inferred using Firth’s correction are not representative of the data. These models thus fail to capture an essential property of the system. Both the self-consistency correction and Firth’s correction provide biased estimates of the temperature, with  $T^* \rightarrow T^0$  from below as  $B \rightarrow \infty$ . The estimated temperatures  $T^*$  should therefore be considered as lower bounds on the true temperature of the dataset. In contrast to other regularization techniques, neither correction requires a hyperparameter to be tuned.

In chapter 7, we show that the bias profoundly impacts the estimation of criticality in a typical finite  $B$  dataset from neuroscience. Not accounting for small sample size effects causes state points to be incorrectly classified. For fluctuating, i.e. dynamically varying data, this corresponds to inferring models which are falsely tuned towards the critical point. Applying the self-consistency correction to this dataset allows us to counteract this and establish that, as a lower bound, the data is paramagnetic. We conclude that any PLM study claiming criticality in a real dataset must carry out a proper analysis of the dependence of the inference on  $B$ , e.g. through the subsampling scheme we describe. Otherwise, small sample size biases cannot be ruled out as the primary cause of the criticality of the inferred model. This is especially important as we have shown that the bias prefactors, e.g.  $\bar{B}$ , setting the learning difficulty of the model are functions of the state point (and also topology [118]), and so one cannot establish a priori what constitutes a “small” sample size.

In chapter 8, we apply PLM to the human connectome project (HCP) young adult dataset. As a result of our previous investigations, we subsample this dataset and establish that it is large enough to safely disregard the small-sample size bias. We find that the resting-state network of the human brain predominantly contains small negative couplings, although the distribution of couplings is skewed, with the positive couplings forming a power-law-like tail. We identified strong symmetry between the connectivity of the two hemispheres. We probed whether the PLM model of the HCP dataset corresponded to a critical point by simulating the model at a range of fictive temperatures, finding the resting-state brain to operate at a paramagnetic state point close to the critical point. We also investigated PLM as a general network reconstruction method, with the aim of superseding correlation-based measures that are currently used in the neuroimaging community. Through simulations, we show that thresholding schemes which construct network topologies by removing all negative couplings (i.e. those that only consider excitatory connections) cannot reproduce the correlations of the empirical input data. Conversely, applying symmetric thresholds (i.e. extracting impactful inhibitory and excitatory connections) allowed us to

remove  $\sim 15\%$  of couplings without altering the statistics of the model. This allows us to establish a self-consistent thresholding scheme with which to simplify the network.

Based on our investigations, we highlight the following avenues for future work. We focused entirely on the study of binary data, as binary representations of neural activity are common, and as computationally efficient solutions such as PLM exist with which to solve the inverse Ising problem. Many datasets, including the raw fMRI data we study, however, are not binary. As such, it would be interesting to extend PLM to e.g. XY-like models (where spins are allowed to rotate continuously within a 2D plane) [229]. Furthermore, in our study of the small sample bias we exclusively generated data from the SK model, which perfectly matches the form of the generalised Ising models we infer with PLM (i.e. we only generated data from binary pairwise models). It is not immediately clear how our conclusions would translate to the case of a model mismatch. Say we were to generate data from models more representative of true neural dynamics, such as the coupled oscillator Kuramoto model [230, 231], would we still observe the same biasing towards criticality? Analysing PLM for such models would help provide a bridge between the inference errors we identify in the best-case SK model sampling, and the unknown true model which generated the experimental data. Finally, although we highlight PLM as a graph reconstruction method and qualitatively discuss the structure of the inferred couplings for the HCP dataset, we were unable to carry out a detailed graph-theoretic analysis of the resulting networks. Doing so would allow us to strengthen our arguments regarding the difference between correlation and coupling networks and assess whether coupling networks also display the key small-world and scale-free properties previously reported in the neuroimaging literature. For anyone wishing to continue this work, the code used to generate, simulate and infer the datasets throughout this thesis is available as a Python package at <https://github.com/maxkloucek/pyplm>.

# Appendices

# Appendix A

## Determining the Linear Bias Regime

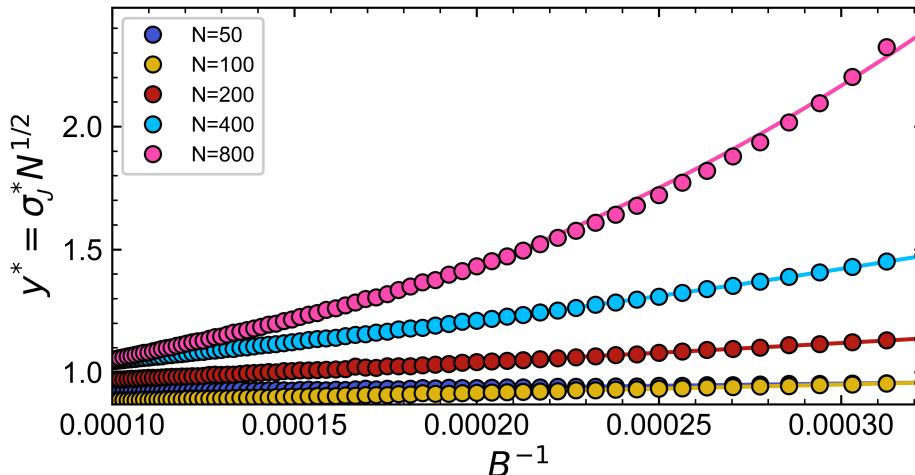
In our main analysis, we identify that statistical averages of the inferred parameters depend linearly on  $1/B$ , linking this with the standard first-order bias of maximum likelihood estimators. In this appendix, we present a method to identify the sample size  $B$  where this approximation is true. We will look at the convergence of  $\sigma_J N^{1/2}$  (i.e. the inverse temperature) as a function of  $1/B$ , using our varying  $N$  minimum error sub-sampling dataset as a case study (see section 5.6). Setting  $y = \sigma_J N^{1/2}$  for convenience, we can write the maximum likelihood estimate of  $y$  as

$$y^* = y^0 + \frac{b_1}{B} + \mathcal{O}(B^{-2}), \quad (\text{A.1})$$

where  $b_1$  is the first-order bias. If we are in a  $B$  regime dominated by the first-order term, we then expect

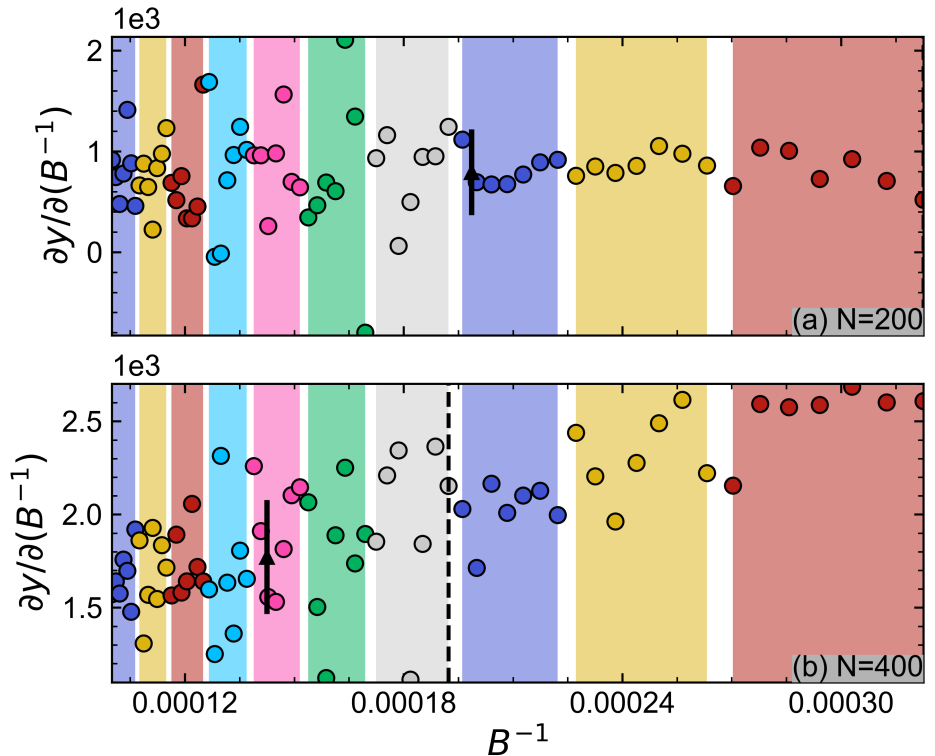
$$\frac{\partial y^*}{\partial B^{-1}} \approx b_1 = \text{const.} \quad (\text{A.2})$$

We plot  $y^*$  as a function of  $B^{-1}$  in Fig.A.1 to exemplify the issue at hand. The small ( $N = 50$ ,  $N = 100$  and  $N = 200$ ) system sizes are well approximated by a linear dependence. But for e.g.  $N = 800$  we see that higher



**Figure A.1:** Dependence of the inverse inferred temperature  $\sigma_J^* N^{1/2}$  on the inverse sample size  $B^{-1}$  for sub-samples of datasets from the minimum error state-points shown in Fig. 5.17.

**Figure A.2:** Statistical method to establish if an observable depends linearly on  $1/B$ . Gradients are split into windows of equal width in  $B$  (shown by shaded rectangles). In (a)  $N = 200$  and all samples were found to belong to the same distribution (i.e. the gradient is constant). In (b)  $N = 400$  and we find a boundary above which the gradient was significantly different from the gradient in the coalesced window. We mark the boundary with a dashed line and the means and standard deviations within each coalesced window with black triangles and error bars.



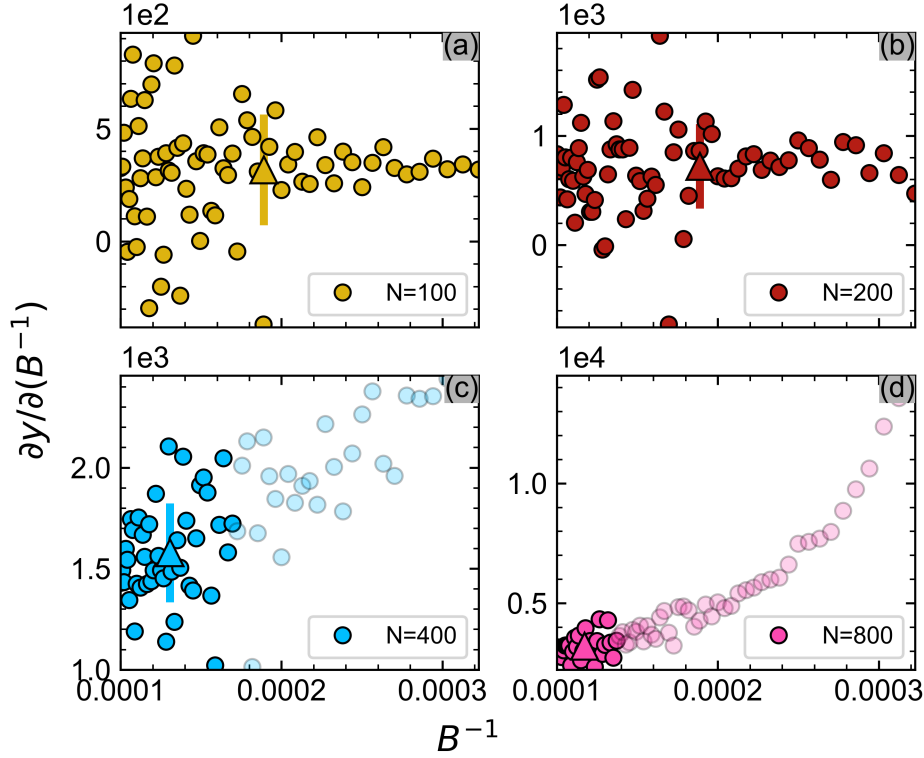
order terms also contribute strongly to the bias.

## A.1 Statistically Identifying a Constant Gradient

We develop a statistical method, with the aim of determining the region of  $B$  in which  $\partial y^* / \partial B^{-1}$  is approximately constant (i.e. in which  $y^*$  is linear in  $1/B$ ). We proceed as follows. We calculate numerical gradients  $\partial y^* / \partial B^{-1}$  for each sub-sample of size  $B$ . We then split the measured numerical gradients into  $N_w = 10$  windows of equal length in  $B$ . We now treat the gradients in each window as samples from some probability distribution. We denote the start of the  $i^{\text{th}}$  window as

$$w_i = B_{\max} \left( \frac{N_w - i}{N_w} \right), \quad (\text{A.3})$$

so that e.g.  $w_0$  runs from  $B_{\text{start}} = B_{\max}$  to  $B_{\text{stop}} = 0.9B_{\max}$ . This process is shown in Fig. A.2 for  $N = 200$  and  $N = 400$ , with each coloured region corresponding to a separate window. We start by performing a two-sample Kolmogorov–Smirnov (KS) test between  $w_0$  and the next window  $w_1$ . This test assesses how likely it is that the samples from both windows were drawn from the same probability distribution. If we find the samples of the gradient in  $w_0$  and  $w_1$  to be from the same distribution we “coalesce” the window. We then test the coalesced window  $w_{<2}$  against the next window  $w_2$ , continuing this process until we find  $w_{<i}$  to be different from  $w_i$ . We define this point as the boundary between the linear asymptotic  $1/B$  regime and the under-sampled regime where higher order terms ( $\mathcal{O}(B^{-2})$  etc.) strongly affect the bias. This boundary is shown for  $N = 400$  by the dotted vertical line in



**Figure A.3:** Gradients for different system sizes. We use our statistical coalescence scheme to identify whether gradients were drawn from the same underlying population. Each panel corresponds to a different system size which is shown in the legend. Full points denote the sample where the gradient is constant (in a statistical sense). Transparent points are those where the gradient is different from the constant sample. Triangles and error bars show the means and standard deviations for each constant gradient sample.

Fig. A.2. In the linear regime, we can extrapolate observables like  $y^*$  to the  $B \rightarrow \infty$  limit and look at the intercept to obtain an “unbiased” estimate for the observable in question. We can also calculate the bias parameters  $b_1$  as the average in our sample  $w_{<i}$ , which we indicate with black triangles in Fig. A.2. This then allows us to assess if enough data has been collected to trust our estimate of the state point. We acknowledge that this process implicitly assumes that  $w_0$  belongs to the linear regime. We counter this by noting that if this were not the case, we would find the boundary to lie between  $w_0$  and  $w_1$  giving us a clear indication that more data is needed.

The results of this coalescence process for the gradients of  $y^* = \sigma_j^* N^{1/2}$  are shown for  $N \geq 50$  in Fig. A.3. Transparent points indicate those outside of the first-order bias regime. We find that for  $N = 50$ ,  $N = 100$  and  $200$  all sub-sample sizes can be approximated by the linear regime and that the inferred inverse temperature is well described by a  $1/B$  dependence on the number of sub-samples. We find that higher-order terms are relevant for the two bigger systems. The average values of  $b_1 = \partial y^* / \partial B^{-1}$  within the linear regime are shown by triangles. Increasing  $N$  corresponds to shifting the onset of the  $1/B$  regime to higher  $B$ , i.e. more data is required to learn bigger systems.



# Bibliography

- [1] M. B. Klouček, T. Machon, S. Kajimura, C. P. Royall, N. Masuda, and F. Turci, *Biases in Inverse Ising Estimates of Near-Critical Behaviour*, Jan. 2023.
- [2] D. R. Chialvo, “Emergent complex neural dynamics”, *Nature Phys* **6**, 744 (2010).
- [3] W. Bialek, “Perspectives on theory at the interface of physics and biology”, *Rep. Prog. Phys.* **81**, 012601 (2018).
- [4] P. W. Anderson, “More Is Different”, *Science* **177**, 393 (1972).
- [5] A. L. Hodgkin and A. F. Huxley, “A quantitative description of membrane current and its application to conduction and excitation in nerve”, *J. Physiol.* **117**, 500 (1952).
- [6] T. Mora and W. Bialek, “Are Biological Systems Poised at Criticality?”, *J Stat Phys* **144**, 268 (2011).
- [7] M. A. Muñoz, “Colloquium : Criticality and dynamical scaling in living systems”, *Rev. Mod. Phys.* **90**, 031001 (2018).
- [8] O. Kinouchi and M. Copelli, “Optimal dynamical range of excitable networks at criticality”, *Nature Phys* **2**, 348 (2006).
- [9] P. Rämö, J. Kesseli, and O. Yli-Harja, “Perturbation avalanches and criticality in gene regulatory networks”, *Journal of Theoretical Biology* **242**, 164 (2006).
- [10] G. Deco and V. K. Jirsa, “Ongoing Cortical Activity at Rest: Criticality, Multistability, and Ghost Attractors”, *Journal of Neuroscience* **32**, 3366 (2012).
- [11] J. O’Byrne and K. Jerbi, “How critical is brain criticality?”, *Trends in Neurosciences* **45**, 820 (2022).
- [12] V. Zimmern, “Why Brain Criticality Is Clinically Relevant: A Scoping Review”, *Front. Neural Circuits* **14** (2020).
- [13] T. Ezaki, E. Fonseca dos Reis, T. Watanabe, M. Sakaki, and N. Masuda, “Closer to critical resting-state neural dynamics in individuals with higher fluid intelligence”, *Commun Biol* **3**, 52 (2020).
- [14] E. Schneidman, M. J. Berry, R. Segev, and W. Bialek, “Weak pairwise correlations imply strongly correlated network states in a neural population”, *Nature* **440**, 1007 (2006).
- [15] E. Aurell and M. Ekeberg, “Inverse Ising Inference Using All the Data”, *Phys. Rev. Lett.* **108**, 090201 (2012).

- [16] H. C. Nguyen, R. Zecchina, and J. Berg, “Inverse statistical problems: from the inverse Ising problem to data science”, *Advances in Physics* **66**, 197 (2017).
- [17] P. Mehta, M. Bukov, C.-H. Wang, A. G. Day, C. Richardson, C. K. Fisher, and D. J. Schwab, “A high-bias, low-variance introduction to Machine Learning for physicists”, *Physics Reports* **810**, 1 (2019).
- [18] W. S. McCulloch and W. Pitts, “A logical calculus of the ideas immanent in nervous activity”, *Bulletin of Mathematical Biophysics* **5**, 115 (1943).
- [19] K. Binder and A. P. Young, “Spin glasses: Experimental facts, theoretical concepts, and open questions”, *Rev. Mod. Phys.* **58**, 801 (1986).
- [20] E. Ising, “Beitrag zur Theorie des Ferromagnetismus”, *Z. Physik* **31**, 253 (1925).
- [21] J. J. Hopfield, “Neural networks and physical systems with emergent collective computationalabilities”, *Proc. Natl. Acad. Sci.* **79**, 2554 (1982).
- [22] D. J. Amit, H. Gutfreund, and H. Sompolinsky, “Spin-glass models of neural networks”, *Phys. Rev. A* **32**, 1007 (1985).
- [23] M. F. Glasser, T. S. Coalson, E. C. Robinson, C. D. Hacker, J. Harwell, E. Yacoub, K. Ugurbil, J. Andersson, C. F. Beckmann, M. Jenkinson, S. M. Smith, and D. C. Van Essen, “A multi-modal parcellation of human cerebral cortex”, *Nature* **536**, 171 (2016).
- [24] S. Platek, T. Shackelford, and J. Keenan, *Evolutionary Cognitive Neuroscience*, 1st edition (MIT Press, Cambridge, Mass, Nov. 2006).
- [25] E. Bullmore and O. Sporns, “Complex brain networks: graph theoretical analysis of structural and functional systems”, *Nat Rev Neurosci* **10**, 186 (2009).
- [26] D. S. Bassett and O. Sporns, “Network neuroscience”, *Nat Neurosci* **20**, 353 (2017).
- [27] J. G. White, E. Southgate, J. N. Thomson, and S. Brenner, “The structure of the nervous system of the nematode *Caenorhabditis elegans*”, *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **314**, 1 (1997).
- [28] E. Jonas and K. P. Kording, “Could a Neuroscientist Understand a Microprocessor?”, *PLOS Computational Biology* **13**, e1005268 (2017).
- [29] T. J. Sejnowski, P. S. Churchland, and J. A. Movshon, “Putting big data to good use in neuroscience”, *Nat Neurosci* **17**, 1440 (2014).
- [30] S. Ogawa, T. M. Lee, A. R. Kay, and D. W. Tank, “Brain magnetic resonance imaging with contrast dependent on blood oxygenation.”, *Proc. Natl. Acad. Sci.* **87**, 9868 (1990).
- [31] N. K. Logothetis and B. A. Wandell, “Interpreting the BOLD Signal”, *Annu. Rev. Physiol.* **66**, 735 (2004).
- [32] C. Chang, J. P. Cunningham, and G. H. Glover, “Influence of heart rate on the BOLD signal: The cardiac response function”, *NeuroImage* **44**, 857 (2009).

- [33] M. E. Raichle and M. A. Mintun, “Brain Work and Brain Imaging”, *Annu. Rev. Neurosci.* **29**, 449 (2006).
- [34] M. P. van den Heuvel and H. E. Hulshoff Pol, “Exploring the brain network: A review on resting-state fMRI functional connectivity”, *European Neuropsychopharmacology* **20**, 519 (2010).
- [35] D. Sherrington and S. Kirkpatrick, “Solvable Model of a Spin-Glass”, *Phys. Rev. Lett.* **35**, 1792 (1975).
- [36] E. T. Jaynes, “Information Theory and Statistical Mechanics”, *Phys. Rev.* **106**, 620 (1957).
- [37] D. Firth, “Bias Reduction of Maximum Likelihood Estimates”, *Biometrika* **80**, 27 (1993).
- [38] D. Chandler, *Introduction to modern statistical mechanics* (Oxford University Press, 1987).
- [39] G. Parisi, *Statistical field theory* (Addison-wesley, 1988).
- [40] A. Onuki, *Phase Transition Dynamics* (Cambridge University Press, 2002).
- [41] M. E. Fisher and M. N. Barber, “Scaling Theory for Finite-Size Effects in the Critical Region”, *Phys. Rev. Lett.* **28**, 1516 (1972).
- [42] K. Binder, “Finite size effects on phase transitions”, *Ferroelectrics* **73**, 43 (1987).
- [43] D. J. Amit and V. Martin-Mayor, *Field Theory, the Renormalization Group, and Critical Phenomena: Graphs to Computers*, Third (WORLD SCIENTIFIC, June 2005).
- [44] T. Castellani and A. Cavagna, “Spin-glass theory for pedestrians”, *J. Stat. Mech.* **2005**, P05012 (2005).
- [45] P. Gallo, K. Amann-Winkel, C. A. Angell, M. A. Anisimov, F. Caupin, C. Chakravarty, E. Lascaris, T. Loerting, A. Z. Panagiotopoulos, J. Russo, J. A. Sellberg, H. E. Stanley, H. Tanaka, C. Vega, L. Xu, and L. G. M. Pettersson, “Water: A Tale of Two Liquids”, *Chem. Rev.* **116**, 7463 (2016).
- [46] L. Onsager, “Crystal Statistics. I. A Two-Dimensional Model with an Order-Disorder Transition”, *Phys. Rev.* **65**, 117 (1944).
- [47] J. Cardy, *Scaling and Renormalization in Statistical Physics*, First (Cambridge University Press, Apr. 1996).
- [48] K. Hu, J. H. Meijer, S. A. Shea, H. T. vanderLeest, B. Pittman-Polletta, T. Houben, F. van Oosterhout, T. Deboer, and F. A. J. L. Scheer, “Fractal Patterns of Neural Activity Exist within the Suprachiasmatic Nucleus and Require Extrinsic Network Interactions”, *PLoS ONE* **7**, edited by E. C. Johnson, e48927 (2012).
- [49] A.-L. Barabási, “The origin of bursts and heavy tails in human dynamics”, *Nature* **435**, 207 (2005).
- [50] R. Albert and A.-L. Barabási, “Statistical mechanics of complex networks”, *Rev. Mod. Phys.* **74**, 47 (2002).

- [51] P. C. Hohenberg and B. I. Halperin, “Theory of dynamic critical phenomena”, *Rev. Mod. Phys.* **49**, 435 (1977).
- [52] Y. C. Kim, M. A. Anisimov, J. V. Sengers, and E. Luijten, “Crossover Critical Behavior in the Three-Dimensional Ising Model”, *J. Stat. Phys.* **110**, 19 (2003).
- [53] F. Krzakala, M. Tarzia, and L. Zdeborová, “Lattice Model for Colloidal Gels and Glasses”, *Phys. Rev. Lett.* **101**, 165702 (2008).
- [54] C. P. Royall, D. G. A. L. Aarts, and H. Tanaka, “Bridging length scales in colloidal liquids and interfaces from near-critical divergence to single particles”, *Nature Phys* **3**, 636 (2007).
- [55] P. Gallo and F. Sciortino, “Ising Universality Class for the Liquid-Liquid Critical Point of a One Component Fluid: A Finite-Size Scaling Test”, *Phys. Rev. Lett.* **109**, 177801 (2012).
- [56] L. E. Ismail and W. Karwowski, “A Graph Theory-Based Modeling of Functional Brain Connectivity Based on EEG: A Systematic Review in the Context of Neuroergonomics”, *IEEE Access* **8**, 155103 (2020).
- [57] S. H. Strogatz, “Exploring complex networks”, *Nature* **410**, 268 (2001).
- [58] M. E. Newman, “The Structure and Function of Complex Networks”, *SIAM Rev.* **45**, 165 (2003).
- [59] D. J. Watts and S. H. Strogatz, “Collective dynamics of ‘small-world’ networks”, *Nature* **393**, 440 (1998).
- [60] A.-L. Barabási and R. Albert, “Emergence of Scaling in Random Networks”, *Science* **286**, 509 (1999).
- [61] R. Albert, H. Jeong, and A.-L. Barabási, “Error and attack tolerance of complex networks”, *Nature* **406**, 378 (2000).
- [62] S. Achard, R. Salvador, B. Whitcher, J. Suckling, and E. Bullmore, “A Resilient, Low-Frequency, Small-World Human Brain Functional Network with Highly Connected Association Cortical Hubs”, *J. Neurosci.* **26**, 63 (2006).
- [63] R. Salvador, J. Suckling, M. R. Coleman, J. D. Pickard, D. Menon, and E. Bullmore, “Neurophysiological Architecture of Functional Magnetic Resonance Images of Human Brain”, *Cerebral Cortex* **15**, 1332 (2005).
- [64] D. S. Bassett, A. Meyer-Lindenberg, S. Achard, T. Duke, and E. Bullmore, “Adaptive reconfiguration of fractal small-world human brain functional networks”, *Proc. Natl. Acad. Sci.* **103**, 19518 (2006).
- [65] D. R. Chialvo, “Critical brain networks”, *Physica A: Statistical Mechanics and its Applications, Complexity and Criticality: In Memory of Per Bak (1947–2002)* **340**, 756 (2004).
- [66] V. M. Eguíluz, D. R. Chialvo, G. A. Cecchi, M. Baliki, and A. V. Apkarian, “Scale-Free Brain Functional Networks”, *Phys. Rev. Lett.* **94**, 018102 (2005).
- [67] O. Sporns, D. R. Chialvo, M. Kaiser, and C. C. Hilgetag, “Organization, development and function of complex brain networks”, *Trends in Cognitive Sciences* **8**, 418 (2004).

- [68] D. S. Bassett and E. Bullmore, “Small-World Brain Networks”, *Neuroscientist* **12**, 512 (2006).
- [69] L. F. Lago-Fernández, R. Huerta, F. Corbacho, and J. A. Sigüenza, “Fast Response and Temporal Coherent Oscillations in Small-World Networks”, *Phys. Rev. Lett.* **84**, 2758 (2000).
- [70] D. Simard, L. Nadeau, and H. Kröger, “Fastest learning in small-world neural networks”, *Physics Letters A* **336**, 8 (2005).
- [71] N. Masuda and K. Aihara, “Global and local synchrony of coupled neurons in small-world networks”, *Biol. Cybern.* **90**, 302 (2004).
- [72] V. Latora and M. Marchiori, “Efficient Behavior of Small-World Networks”, *Phys. Rev. Lett.* **87**, 198701 (2001).
- [73] C. P. Herrero, “Ising model in small-world networks”, *Phys. Rev. E* **65**, 066110 (2002).
- [74] C. P. Herrero, “Ising model in scale-free networks: A Monte Carlo simulation”, *Phys. Rev. E* **69**, 067109 (2004).
- [75] C. van Vreeswijk and H. Sompolinsky, “Chaos in Neuronal Networks with Balanced Excitatory and Inhibitory Activity”, *Science* **274**, 1724 (1996).
- [76] L. J. Herstel and C. J. Wierenga, “Network control through coordinated inhibition”, *Current Opinion in Neurobiology, Neurobiology of Learning and Plasticity* **67**, 34 (2021).
- [77] V. S. Sohal and J. L. R. Rubenstein, “Excitation-inhibition balance as a framework for investigating mechanisms in neuropsychiatric disorders”, *Mol Psychiatry* **24**, 1248 (2019).
- [78] M. Mezard, G. Parisi, and M. Virasoro, *Spin Glass Theory and Beyond: An Introduction to the Replica Method and Its Applications*, Vol. 9, World Scientific Lecture Notes in Physics (WORLD SCIENTIFIC, Nov. 1986).
- [79] J. Hopfield and D. Tank, “Computing with neural circuits: a model”, *Science* **233**, 625 (1986).
- [80] S. Kirkpatrick and D. Sherrington, “Infinite-ranged models of spin-glasses”, *Phys. Rev. B* **17**, 4384 (1978).
- [81] V. Cannella and J. A. Mydosh, “Magnetic Ordering in Gold-Iron Alloys”, *Phys. Rev. B* **6**, 4220 (1972).
- [82] S. F. Edwards and P. W. Anderson, “Theory of spin glasses”, *J. Phys. F* **5**, 965 (1975).
- [83] J. R. L. de Almeida and D. J. Thouless, “Stability of the Sherrington-Kirkpatrick solution of a spin glass model”, *J. Phys. Math. Gen.* **11**, 983 (1978).
- [84] G. Parisi, “TOWARD A MEAN FIELD THEORY FOR SPIN GLASSES”, *Phys. Lett.* **73**, 3 (1979).
- [85] G. Parisi, “A sequence of approximated solutions to the S-K model for spin glasses”, *J. Phys. Math. Gen.* **13**, 8 (1980).

- [86] G. Parisi, “Order Parameter for Spin-Glasses”, *Phys. Rev. Lett.* **50**, 1946 (1983).
- [87] D. Panchenko, “Introduction to the SK model”, *Curr. Dev. Math.* **2014**, 231 (2014).
- [88] P. Carmona and Y. Hu, “Universality in Sherrington–Kirkpatrick’s spin glass model”, *Annales de l’Institut Henri Poincaré (B) Probability and Statistics* **42**, 215 (2006).
- [89] K. H. Fischer and J. A. Hertz, *Spin Glasses*, First (Cambridge University Press, May 1991).
- [90] L. Berthier, G. Biroli, P. Charbonneau, E. I. Corwin, S. Franz, and F. Zamponi, “Gardner physics in amorphous solids and beyond”, *J. Chem. Phys.* **151**, 010901 (2019).
- [91] C. E. Shannon, “A mathematical theory of communication”, *Bell Syst. Tech. J.* **27**, 379 (1948).
- [92] E. T. Jaynes, “Information Theory and Statistical Mechanics. II”, *Phys. Rev.* **108**, 171 (1957).
- [93] M. Brokate, P. Manchanda, and A. H. Siddiqi, *Calculus for Scientists and Engineers*, Industrial and Applied Mathematics (Springer Singapore, Singapore, 2019).
- [94] D. H. Ackley, G. E. Hinton, and T. J. Sejnowski, “A Learning Algorithm for Boltzmann Machines\*”, *Cogn. Sci.* **9**, 147 (1985).
- [95] S. Kullback, *Information theory and statistics* (Wiley, New York, 1959).
- [96] S. Geman and D. Geman, “Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images”, *IEEE PAMI-6*, 21 (1984).
- [97] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, “Equation of State Calculations by Fast Computing Machines”, *The Journal of Chemical Physics* **21**, 1087 (1953).
- [98] W. K. Hastings, “Monte Carlo sampling methods using Markov chains and their applications”, *Biometrika* **57**, 97 (1970).
- [99] G. Tkacik, E. Schneidman, M. J. Berry II, and W. Bialek, “Ising models for networks of real neurons”, arXiv (2006).
- [100] T. Broderick, M. Dudik, G. Tkacik, R. E. Schapire, and W. Bialek, *Faster solutions of the inverse pairwise Ising problem*, Dec. 2007.
- [101] E. J. G. Pitman, “Sufficient statistics and intrinsic accuracy”, *Math. Proc. Camb. Philos. Soc.* **32**, 567 (1936).
- [102] B. O. Koopman, “On distributions admitting a sufficient statistic”, *Trans. Amer. Math. Soc.* **39**, 399 (1936).
- [103] E. L. Lehmann and G. Casella, *Theory of point estimation*, Second, Springer Texts in Statistics (Springer-Verlag, New York, 1998).
- [104] A. Montanari, “Computational implications of reducing data to sufficient statistics”, *Electron. J. Stat.* **9**, 2370 (2015).

- [105] G. Tkačik, O. Marre, D. Amodei, E. Schneidman, W. Bialek, and M. J. Berry, “Searching for Collective Behavior in a Large Network of Sensory Neurons”, *PLoS Comput Biol* **10**, edited by O. Sporns, e1003408 (2014).
- [106] Y. Roudi, E. Aurell, and J. Hertz, “Statistical physics of pairwise probability models”, *Front. Comput. Neurosci.* **3** (2009).
- [107] Y. Roudi, J. Tyrcha, and J. Hertz, “Ising model for neural data: Model quality and approximate methods for extracting functional connectivity”, *Phys. Rev. E* **79**, 051915 (2009).
- [108] T. Tanaka, “Mean-field theory of Boltzmann machine learning”, *Phys. Rev. E* **58**, 2302 (1998).
- [109] H. J. Kappen and F. B. Rodríguez, “Boltzmann machine learning using mean field theory and linear response correction”, in *Proc. 10th Int. Conf. Neural Inf. Process. Syst. NIPS’97* (Dec. 1997), pp. 280–286.
- [110] D. J. Thouless, P. W. Anderson, and R. G. Palmer, “Solution of ‘Solvable model of a spin glass’”, *Philos. Mag. J. Theor. Exp. Appl. Phys.* **35**, 593 (1977).
- [111] F. Ricci-Tersenghi, “The Bethe approximation for solving the inverse Ising problem: a comparison with other inference methods”, *J. Stat. Mech.* **2012**, P08015 (2012).
- [112] M. Mézard and T. Mora, “Constraint satisfaction problems and neural networks: A statistical physics perspective”, *Journal of Physiology-Paris, Neuromathematics of Vision* **103**, 107 (2009).
- [113] E. Marinari and V. V. Kerrebroeck, “Intrinsic limitations of the susceptibility propagation inverse inference for the mean field Ising spin glass”, *J. Stat. Mech.* **2010**, P02008 (2010).
- [114] E. Aurell, C. Ollion, and Y. Roudi, “Dynamics and performance of susceptibility propagation on synthetic data”, *Eur. Phys. J. B* **77**, 587 (2010).
- [115] V. Sessak and R. Monasson, “Small-correlation expansions for the inverse Ising problem”, *J. Phys. A: Math. Theor.* **42**, 055001 (2009).
- [116] S. Cocco and R. Monasson, “Adaptive Cluster Expansion for Inferring Boltzmann Machines with Noisy Data”, *Phys. Rev. Lett.* **106**, 090601 (2011).
- [117] H. C. Nguyen and J. Berg, “Mean-Field Theory for the Inverse Ising Problem at Low Temperatures”, *Phys. Rev. Lett.* **109**, 050602 (2012).
- [118] A. Y. Lokhov, M. Vuffray, S. Misra, and M. Chertkov, “Optimal structure and parameter learning of Ising models”, *Sci. Adv.*, **10**. 1126/sciadv.1700791 (2018).
- [119] M. Welling and Y. W. Teh, “Approximate inference in Boltzmann machines”, *Artificial Intelligence* **143**, 19 (2003).
- [120] P. Ravikumar, M. J. Wainwright, and J. D. Lafferty, “High-dimensional Ising model selection using l1-regularized logistic regression”, *Ann. Stat.* **38**, 1287 (2010).



- [121] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and É. Duchesnay, “Scikit-learn: Machine Learning in Python”, *J. Mach. Learn. Res.* **12**, 2825 (2011).
- [122] R. H. Byrd, P. Lu, J. Nocedal, and C. Zhu, “A Limited Memory Algorithm for Bound Constrained Optimization”, *SIAM J. Sci. Comput.* **16**, 1190 (1995).
- [123] C. Zhu, R. H. Byrd, P. Lu, and J. Nocedal, “Algorithm 778: L-BFGS-B: Fortran subroutines for large-scale bound-constrained optimization”, *ACM Trans. Math. Softw.* **23**, 550 (1997).
- [124] J. L. Morales and J. Nocedal, “Remark on “algorithm 778: L-BFGS-B: Fortran subroutines for large-scale bound constrained optimization””, *ACM Trans. Math. Softw.* **38**, 7:1 (2011).
- [125] J. Sohl-Dickstein, P. B. Battaglino, and M. R. DeWeese, “New Method for Parameter Estimation in Probabilistic Models: Minimum Probability Flow”, *Phys. Rev. Lett.* **107**, 220601 (2011).
- [126] T. Ezaki, T. Watanabe, M. Ohzeki, and N. Masuda, “Energy landscape analysis of neuroimaging data”, *Phil. Trans. R. Soc. A* **375**, 20160287 (2017).
- [127] K. J. Friston, “Functional and effective connectivity in neuroimaging: A synthesis”, *Hum. Brain Mapp.* **2**, 56 (1994).
- [128] F. De Vico Fallani, J. Richiardi, M. Chavez, and S. Achard, “Graph analysis of functional brain networks: practical issues in translational neuroscience”, *Philos. Trans. R. Soc. B Biol. Sci.* **369**, 20130521 (2014).
- [129] O. David, D. Cosmelli, and K. J. Friston, “Evaluation of different measures of functional connectivity using a neural mass model”, *NeuroImage* **21**, 659 (2004).
- [130] K. Ansari-Asl, L. Senhadji, J.-J. Bellanger, and F. Wendling, “Quantitative evaluation of linear and nonlinear methods characterizing interdependencies between brain signals”, *Phys. Rev. E* **74**, 031916 (2006).
- [131] S. M. Smith, K. L. Miller, G. Salimi-Khorshidi, M. Webster, C. F. Beckmann, T. E. Nichols, J. D. Ramsey, and M. W. Woolrich, “Network modelling methods for FMRI”, *NeuroImage* **54**, 875 (2011).
- [132] A. J. Schwarz and J. McGonigle, “Negative edges and soft thresholding in complex network analysis of resting state functional connectivity data”, *NeuroImage* **55**, 1132 (2011).
- [133] S. Kajimura, N. Masuda, J. K. L. Lau, and K. Murayama, “Focused attention meditation changes the boundary and configuration of functional networks in the brain”, *Sci Rep* **10**, 18426 (2020).
- [134] G. Deco, V. K. Jirsa, and A. R. McIntosh, “Emerging concepts for the dynamical organization of resting-state activity in the brain”, *Nat Rev Neurosci* **12**, 43 (2011).

- [135] O. Korhonen, M. Zanin, and D. Papo, “Principles and open questions in functional brain network reconstruction”, *Hum. Brain Mapp.* **42**, 3680 (2021).
- [136] Z. Liang, J. King, and N. Zhang, “Anticorrelated resting-state functional connectivity in awake rat brain”, *NeuroImage* **59**, 1190 (2012).
- [137] D. Fraiman, P. Balenzuela, J. Foss, and D. R. Chialvo, “Ising-like dynamics in large-scale functional brain networks”, *Phys. Rev. E* **79**, 061922 (2009).
- [138] A. Decelle and F. Ricci-Tersenghi, “Pseudolikelihood Decimation Algorithm Improving the Inference of the Interaction Network in a General Class of Ising Models”, *Phys. Rev. Lett.* **112**, 070603 (2014).
- [139] J. Shlens, G. D. Field, J. L. Gauthier, M. I. Grivich, D. Petrusca, A. Sher, A. M. Litke, and E. J. Chichilnisky, “The Structure of Multi-Neuron Firing Patterns in Primate Retina”, *J. Neurosci.* **26**, 8254 (2006).
- [140] S. Cocco, S. Leibler, and R. Monasson, “Neuronal couplings between retinal ganglion cells inferred by efficient inverse statistical physics methods”, *Proc. Natl. Acad. Sci.* **106**, 14058 (2009).
- [141] I. E. Ohiorhenuan, F. Mechler, K. P. Purpura, A. M. Schmid, Q. Hu, and J. D. Victor, “Sparse coding and high-order correlations in fine-scale cortical networks”, *Nature* **466**, 617 (2010).
- [142] A. Tang, D. Jackson, J. Hobbs, W. Chen, J. L. Smith, H. Patel, A. Prieto, D. Petrusca, M. I. Grivich, A. Sher, P. Hottowy, W. Dabrowski, A. M. Litke, and J. M. Beggs, “A Maximum Entropy Model Applied to Spatial and Temporal Correlations from Cortical Networks In Vitro”, *Journal of Neuroscience* **28**, 505 (2008).
- [143] E. Ganmor, R. Segev, and E. Schneidman, “Sparse low-order interaction network underlies a highly correlated and learnable neural population code”, *Proc Natl Acad Sci U S A* **108**, 9679 (2011).
- [144] G. Tkačik, O. Marre, T. Mora, D. Amodei, M. J. Berry II, and W. Bialek, “The simplest maximum entropy model for collective behavior in a neural network”, *J. Stat. Mech.* **2013**, P03011 (2013).
- [145] T. Watanabe, S. Hirose, H. Wada, Y. Imai, T. Machida, I. Shirouzu, S. Konishi, Y. Miyashita, and N. Masuda, “A pairwise maximum entropy model accurately describes resting-state human brain networks”, *Nat Commun* **4**, 1370 (2013).
- [146] T. Watanabe, S. Hirose, H. Wada, Y. Imai, T. Machida, I. Shirouzu, S. Konishi, Y. Miyashita, and N. Masuda, “Energy landscapes of resting-state brain networks”, *Front. Neuroinform.* **8**, 10.3389/fninf.2014.00012 (2014).
- [147] T. Watanabe, N. Masuda, F. Megumi, R. Kanai, and G. Rees, “Energy landscape and dynamics of brain activity during human bistable perception”, *Nat Commun* **5**, 4765 (2014).
- [148] E. Granot-Atedgi, G. Tkačik, R. Segev, and E. Schneidman, “Stimulus-dependent Maximum Entropy Models of Neural Population Codes”, *PLOS Computational Biology* **9**, e1002922 (2013).

- [149] C. Donner, K. Obermayer, and H. Shimazaki, “Approximate Inference for Time-Varying Interactions and Macroscopic Dynamics of Neural Populations”, *PLoS Comput Biol* **13**, edited by M. Bethge, [e1005309](#) (2017).
- [150] K. A. Dill and J. L. MacCallum, “The Protein-Folding Problem, 50 Years On”, *Science* **338**, 1042 (2012).
- [151] M. Weigt, R. A. White, H. Szurmant, J. A. Hoch, and T. Hwa, “Identification of direct residue contacts in protein-protein interaction by message passing”, *Proceedings of the National Academy of Sciences* **106**, 67 (2009).
- [152] M. Weigt, R. A. White, H. Szurmant, J. A. Hoch, and T. Hwa, “Identification of direct residue contacts in protein-protein interaction by message passing”, *Proc. Natl. Acad. Sci.* **106**, 67 (2009).
- [153] F. Morcos, A. Pagnani, B. Lunt, A. Bertolino, D. S. Marks, C. Sander, R. Zecchina, J. N. Onuchic, T. Hwa, and M. Weigt, “Direct-coupling analysis of residue coevolution captures native contacts across many protein families”, *Proc. Natl. Acad. Sci.* **108**, E1293 (2011).
- [154] F. Morcos, A. Pagnani, B. Lunt, A. Bertolino, D. S. Marks, C. Sander, R. Zecchina, J. N. Onuchic, T. Hwa, and M. Weigt, “Direct-coupling analysis of residue coevolution captures native contacts across many protein families”, *Proceedings of the National Academy of Sciences* **108**, E1293 (2011).
- [155] F. Y. Wu, “The Potts model”, *Rev. Mod. Phys.* **54**, 235 (1982).
- [156] M. Ekeberg, C. Lövkvist, Y. Lan, M. Weigt, and E. Aurell, “Improved contact prediction in proteins: Using pseudolikelihoods to infer Potts models”, *Phys. Rev. E* **87**, 012707 (2013).
- [157] A. Cavagna, A. Cimarelli, I. Giardina, G. Parisi, R. Santagati, F. Stefanini, and M. Viale, “Scale-free correlations in starling flocks”, *Proceedings of the National Academy of Sciences* **107**, 11865 (2010).
- [158] W. Bialek, A. Cavagna, I. Giardina, T. Mora, E. Silvestri, M. Viale, and A. M. Walczak, “Statistical mechanics for natural flocks of birds”, *Proceedings of the National Academy of Sciences* **109**, 4786 (2012).
- [159] W. Bialek, A. Cavagna, I. Giardina, T. Mora, O. Pohl, E. Silvestri, M. Viale, and A. M. Walczak, “Social interactions dominate speed control in poising natural flocks near criticality”, *Proc. Natl. Acad. Sci.* **111**, 7212 (2014).
- [160] T. Bury, “Market structure explained by pairwise interactions”, *Physica A: Statistical Mechanics and its Applications* **392**, 1375 (2013).
- [161] M. A. Valle, G. A. Ruz, and S. Rica, “Market basket analysis by solving the inverse Ising problem: Discovering pairwise interaction strengths among products”, *Physica A: Statistical Mechanics and its Applications* **524**, 36 (2019).
- [162] T. Bury, “A statistical physics perspective on criticality in financial markets”, *J. Stat. Mech.* **2013**, 10.1088/1742-5468/2013/11/P11004 (2013).

- [163] S. S. Borysov, Y. Roudi, and A. V. Balatsky, “U.S. stock market interaction network as learned by the Boltzmann machine”, *Eur. Phys. J. B* **88**, 321 (2015).
- [164] M. A. Valle, J. F. Lavin, and N. S. Magner, “Equity Market Description under High and Low Volatility Regimes Using Maximum Entropy Pairwise Distribution”, *Entropy* **23** (2021).
- [165] E. Tagliazucchi, P. Balenzuela, D. Fraiman, and D. R. Chialvo, “Criticality in Large-Scale Brain fMRI Dynamics Unveiled by a Novel Point Process Analysis”, *Front. Physiol.* **3**, 10.3389/fphys.2012.00015 (2012).
- [166] N. Bertschinger and T. Natschläger, “Real-Time Computation at the Edge of Chaos in Recurrent Neural Networks”, *Neural Computation* **16**, 1413 (2004).
- [167] R. Legenstein and W. Maass, “Edge of chaos and prediction of computational performance for neural circuit models”, *Neural Networks* **20**, 323 (2007).
- [168] J. Wilting and V. Priesemann, “25 years of criticality in neuroscience — established results, open controversies, novel concepts”, *Current Opinion in Neurobiology, Computational Neuroscience* **58**, 105 (2019).
- [169] K. Heiney, O. Huse Ramstad, V. Fiskum, N. Christiansen, A. Sandvig, S. Nichele, and I. Sandvig, “Criticality, Connectivity, and Neural Disorder: A Multifaceted Approach to Neural Computation”, *Front. Comput. Neurosci.* **15** (2021).
- [170] S. Haykin, J. C. Principe, T. J. Sejnowski, J. McWhirter, T. G. Dietterich, M. I. Jordan, and S. Haykin, *New Directions in Statistical Signal Processing: From Systems to Brains* (MIT Press, Cambridge, UNITED STATES, 2006).
- [171] J. M. Beggs and D. Plenz, “Neuronal Avalanches in Neocortical Circuits”, *J. Neurosci.* **23**, 11167 (2003).
- [172] J. M. Beggs, “The criticality hypothesis: how local cortical networks might optimize information processing”, *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* **366**, 329 (2008).
- [173] J. M. Beggs, “Neuronal Avalanches Are Diverse and Precise Activity Patterns That Are Stable for Many Hours in Cortical Slice Cultures”, *Journal of Neuroscience* **24**, 5216 (2004).
- [174] D. Plenz, “Neuronal avalanches and coherence potentials”, *Eur. Phys. J. Spec. Top.* **205**, 259 (2012).
- [175] S. Yu, H. Yang, O. Shriki, and D. Plenz, “Universal organization of resting brain activity at the thermodynamic critical point”, *Front. Syst. Neurosci.* **7**, 10.3389/fnsys.2013.00042 (2013).
- [176] D. Marinazzo, M. Pellicoro, G. Wu, L. Angelini, J. M. Cortés, and S. Stramaglia, “Information Transfer and Criticality in the Ising Model on the Human Connectome”, *PLoS ONE* **9**, edited by R. Lambiotte, e93616 (2014).

- [177] S. Friederich, “Fine-Tuning”, in *The Stanford Encyclopedia of Philosophy*, edited by E. N. Zalta, Summer 2022 (Metaphysics Research Lab, Stanford University, 2022).
- [178] B. S. Ryden, *Introduction to cosmology* (Addison-Wesley, San Francisco, 2003).
- [179] H. Hinrichsen, “Non-equilibrium critical phenomena and phase transitions into absorbing states”, *Adv. Phys.* **49**, 815 (2000).
- [180] P. Bak, C. Tang, and K. Wiesenfeld, “Self-organized criticality: An explanation of the 1/f Noise”, *Phys. Rev. Lett.* **59**, 381 (1987).
- [181] V. Frette, K. Christensen, A. Malthé-Sørensen, J. Feder, T. Jøssang, and P. Meakin, “Avalanche dynamics in a pile of rice”, *Nature* **379**, 49 (1996).
- [182] B. Gutenberg and C. F. Richter, “Magnitude and Energy of Earthquakes”, *Ann. Geophys.* **9**, 1 (1956).
- [183] R. B. Griffiths, “Nonanalytic Behavior Above the Critical Point in a Random Ising Ferromagnet”, *Phys. Rev. Lett.* **23**, 17 (1969).
- [184] A. J. Bray, “Nature of the Griffiths phase”, *Phys. Rev. Lett.* **59**, 586 (1987).
- [185] T. Vojta, “Rare region effects at classical, quantum and nonequilibrium phase transitions”, *J. Phys. A: Math. Gen.* **39**, R143 (2006).
- [186] H. Takano and S. Miyashita, “Relaxation of the Spin Autocorrelation Function in the Ising Spin Glass”, *J. Phys. Soc. Jpn.* **64**, 423 (1995).
- [187] C. Chatelain, “Griffiths phase and critical behavior of the two-dimensional Potts models with long-range correlated disorder”, *Phys. Rev. E* **89**, 032105 (2014).
- [188] P. Moretti and M. A. Muñoz, “Griffiths phases and the stretching of criticality in brain networks”, *Nat Commun* **4**, 2521 (2013).
- [189] M. A. Muñoz, R. Juhász, C. Castellano, and G. Ódor, “Griffiths Phases on Complex Networks”, *Phys. Rev. Lett.* **105**, 128701 (2010).
- [190] G. Tkačik, T. Mora, O. Marre, D. Amodei, S. E. Palmer, M. J. Berry, and W. Bialek, “Thermodynamics and signatures of criticality in a network of neurons”, *Proc Natl Acad Sci USA* **112**, 11508 (2015).
- [191] T. Mora, S. Deny, and O. Marre, “Dynamical Criticality in the Collective Activity of a Population of Retinal Neurons”, *Phys. Rev. Lett.* **114**, 078105 (2015).
- [192] M. G. Kitzbichler, M. L. Smith, S. R. Christensen, and E. Bullmore, “Broadband Criticality of Human Brain Network Synchronization”, *PLoS Comput Biol* **5**, edited by T. Behrens, e1000314 (2009).
- [193] C. Haldeman and J. M. Beggs, “Critical Branching Captures Activity in Living Neural Networks and Maximizes the Number of Metastable States”, *Phys. Rev. Lett.* **94**, 058101 (2005).
- [194] J. M. Beggs and N. Timme, “Being Critical of Criticality in the Brain”, *Front. Physiol.* **3**, 10.3389/fphys.2012.00163 (2012).

- [195] N. Friedman, S. Ito, B. A. W. Brinkman, M. Shimono, R. E. L. DeVile, K. A. Dahmen, J. M. Beggs, and T. C. Butler, “Universal Critical Dynamics in High Resolution Neuronal Avalanche Data”, [Phys. Rev. Lett. \*\*108\*\*, 208102 \(2012\)](#).
- [196] T. Bellay, A. Klaus, S. Seshadri, and D. Plenz, “Irregular spiking of pyramidal neurons organizes as scale-invariant neuronal avalanches in the awake state”, [eLife \*\*4\*\*, edited by F. K. Skinner, e07224 \(2015\)](#).
- [197] S. Zapperi, K. B. Lauritsen, and H. E. Stanley, “Self-Organized Branching Processes: Mean-Field Theory for Avalanches”, [Phys. Rev. Lett. \*\*75\*\*, 4071 \(1995\)](#).
- [198] A. Levina, J. M. Herrmann, and T. Geisel, “Dynamical synapses causing self-organized criticality in neural networks”, [Nature Phys \*\*3\*\*, 857 \(2007\)](#).
- [199] V. Priesemann, M. Valderrama, M. Wibral, and M. Le Van Quyen, “Neuronal Avalanches Differ from Wakefulness to Deep Sleep – Evidence from Intracranial Depth Recordings in Humans”, [PLoS Comput Biol \*\*9\*\*, edited by O. Sporns, e1002985 \(2013\)](#).
- [200] T. Mora, A. M. Walczak, W. Bialek, and C. G. Callan, “Maximum entropy models for antibody diversity”, [Proceedings of the National Academy of Sciences \*\*107\*\*, 5405 \(2010\)](#).
- [201] G. J. Stephens, T. Mora, G. Tkačik, and W. Bialek, “Statistical Thermodynamics of Natural Images”, [Phys. Rev. Lett. \*\*110\*\*, 018701 \(2013\)](#).
- [202] A. Montanari and J. Pereira, “Which graphical models are difficult to learn?”, in *Adv. Neural Inf. Process. Syst. Vol. 22* (2009).
- [203] V. Ngampruetikorn, V. Sachdeva, J. Torrence, J. Humplik, D. J. Schwab, and S. E. Palmer, “Inferring couplings in networks across order-disorder phase transitions”, [Phys. Rev. Res. \*\*4\*\*, 023240 \(2022\)](#).
- [204] I. Mastromatteo and M. Marsili, “On the criticality of inferred models”, [J. Stat. Mech. \*\*2011\*\*, P10012 \(2011\)](#).
- [205] T. Aspelmeier, A. Billoire, E. Marinari, and M. A. Moore, “Finite-size corrections in the Sherrington–Kirkpatrick model”, [J. Phys. A: Math. Theor. \*\*41\*\*, 324008 \(2008\)](#).
- [206] M. P. Allen and D. J. Tildesley, *Computer Simulation of Liquids*, Second (Oxford University Press, Oxford, United Kingdom, Aug. 2017).
- [207] W. Janke, “Monte Carlo Simulations in Statistical Physics - From Basic Principles to Advanced Applications”, in *Order, Disorder and Criticality* (WORLD SCIENTIFIC, June 2012), pp. 93–166.
- [208] A. Albert and J. A. Anderson, “On the Existence of Maximum Likelihood Estimates in Logistic Regression Models”, [Biometrika \*\*71\*\*, 1 \(1984\)](#).
- [209] H. Gould and J. Tobochnik, “Overcoming Critical Slowing Down”, [Comput. Phys. \*\*3\*\*, 82 \(1989\)](#).
- [210] M. Marsili, I. Mastromatteo, and Y. Roudi, “On sampling and modeling complex systems”, [J. Stat. Mech. \*\*2013\*\*, P09003 \(2013\)](#).



- [211] I. Kosmidis, “Bias in parametric estimation: reduction and useful side-effects”, *WIREs Comput. Stat.* **6**, 185 (2014).
- [212] C. Zorn, “A Solution to Separation in Binary Response Models”, *Polit. Anal.* **13**, 157 (2005).
- [213] G. Heinze and M. Schemper, “A solution to the problem of separation in logistic regression”, *Stat. Med.* **21**, 2409 (2002).
- [214] M. A. Mansournia, A. Geroldinger, S. Greenland, and G. Heinze, “Separation in Logistic Regression: Causes, Consequences, and Control”, *American Journal of Epidemiology* **187**, 864 (2018).
- [215] R. G. Miller, “The Jackknife—A Review”, *Biometrika* **61**, 1 (1974).
- [216] I. Kosmidis and D. Firth, “Bias reduction in exponential family non-linear models”, *Biometrika* **96**, 793 (2009).
- [217] I. Kosmidis and D. Firth, “A generic algorithm for reducing bias in parametric estimation”, *Electron. J. Stat.* **4**, 1097 (2010).
- [218] P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, S. J. van der Walt, M. Brett, J. Wilson, K. J. Millman, N. Mayorov, A. R. J. Nelson, E. Jones, R. Kern, E. Larson, C. J. Carey, Í. Polat, Y. Feng, E. W. Moore, J. VanderPlas, D. Laxalde, J. Perktold, R. Cimrman, I. Henriksen, E. A. Quintero, C. R. Harris, A. M. Archibald, A. H. Ribeiro, F. Pedregosa, and P. van Mulbregt, “SciPy 1.0: fundamental algorithms for scientific computing in Python”, *Nat Methods* **17**, 261 (2020).
- [219] J. A. Nelder and M. R., “A Simplex Method for Function Minimization”, *Comput. J.* **7**, 308 (1965).
- [220] R. A. Baer, “Mindfulness training as a clinical intervention: A conceptual and empirical review”, *Clin. Psychol. Sci. Pract.* **10**, 125 (2003).
- [221] J. S. Damoiseaux, S. A. R. B. Rombouts, F. Barkhof, P. Scheltens, C. J. Stam, S. M. Smith, and C. F. Beckmann, “Consistent resting-state networks across healthy subjects”, *Proc. Natl. Acad. Sci.* **103**, 13848 (2006).
- [222] S. M. Smith, P. T. Fox, K. L. Miller, D. C. Glahn, P. M. Fox, C. E. Mackay, N. Filippini, K. E. Watkins, R. Toro, A. R. Laird, and C. F. Beckmann, “Correspondence of the brain’s functional architecture during activation and rest”, *Proc. Natl. Acad. Sci.* **106**, 13040 (2009).
- [223] M. Raemaekers, W. Schellekens, N. Petridou, and N. F. Ramsey, “Knowing left from right: asymmetric functional connectivity during resting state”, *Brain Struct Funct* **223**, 1909 (2018).
- [224] C. Dansereau, Y. Benhajali, C. Risterucci, E. M. Pich, P. Orban, D. Arnold, and P. Bellec, “Statistical power and prediction accuracy in multisite resting-state fMRI connectivity”, *NeuroImage* **149**, 220 (2017).
- [225] K. N. Kay, T. Naselaris, R. J. Prenger, and J. L. Gallant, “Identifying natural images from human brain activity”, *Nature* **452**, 352 (2008).



- [226] A. Clauset, C. R. Shalizi, and M. E. J. Newman, “Power-Law Distributions in Empirical Data”, *SIAM Rev.* **51**, 661 (2009).
- [227] Y. Virkar and A. Clauset, “Power-Law Distributions in Binned Empirical Data”, *Ann. Appl. Stat.* **8**, 89 (2014).
- [228] J. Alstott, E. Bullmore, and D. Plenz, “Powerlaw: A Python Package for Analysis of Heavy-Tailed Distributions”, [PLOS ONE 9, e85777 \(2014\)](#).
- [229] J. M. Kosterlitz, “The critical properties of the two-dimensional xy model”, [J. Phys. C: Solid State Phys. 7, 1046 \(1974\)](#).
- [230] M. Breakspear, S. Heitmann, and A. Daffertshofer, “Generative Models of Cortical Oscillations: Neurobiological Implications of the Kuramoto Model”, *Front. Hum. Neurosci.* **4** (2010).
- [231] H. Choi and S. Mihalas, “Synchronization dependent on spatial structures of a mesoscopic whole-brain network”, [PLOS Computational Biology 15, e1006978 \(2019\)](#).