

---

Adrian Barbu • Song-Chun Zhu

# Monte Carlo Methods

---

# Contents

<b>1</b>	<b>Introduction to Monte Carlo Methods</b> .....	1
1.1	Introduction .....	1
1.2	Motivation and Objectives.....	2
1.3	Tasks in Monte Carlo Computing .....	3
1.3.1	Task 1: Sampling and Simulation .....	4
1.3.2	Task 2: Estimating Quantities by Monte Carlo Simulation .....	7
1.3.3	Task 3: Optimization and Bayesian Inference .....	9
1.3.4	Task 4: Learning and Model Estimation .....	11
1.3.5	Task 5: Visualizing the Landscape .....	12
	References .....	16
<b>2</b>	<b>Sequential Monte Carlo</b> .....	19
2.1	Introduction .....	19
2.2	Sampling a 1-Dimensional Density.....	19
2.3	Importance Sampling and Weighted Samples .....	20
2.4	Sequential Importance Sampling (SIS) .....	24
2.4.1	Application: The Number of Self-Avoiding Walks.....	25
2.4.2	Application: Particle Filtering for Tracking Objects in a Video.....	28
2.4.3	Summary of the SMC Framework .....	31
2.5	Application: Ray Tracing by SMC.....	33
2.5.1	Example: Glossy Highlights.....	34
2.6	Preserving Sample Diversity in Importance Sampling.....	35
2.6.1	Parzen Window Discussion.....	39
2.7	Monte Carlo Tree Search.....	41
2.7.1	Pure Monte Carlo Tree Search .....	42
2.7.2	AlphaGo .....	44
2.8	Exercises .....	46
	References .....	48

<b>3</b>	<b>Markov Chain Monte Carlo: The Basics</b> .....	49
3.1	Introduction .....	50
3.2	Markov Chain Basics .....	50
3.3	Topology of Transition Matrix: Communication and Period .....	53
3.4	The Perron-Frobenius Theorem .....	56
3.5	Convergence Measures .....	58
3.6	Markov Chains in Continuous or Heterogeneous State Spaces .....	61
3.7	Ergodicity Theorem .....	62
3.8	MCMC for Optimization by Simulated Annealing .....	62
3.8.1	Page Rank Example .....	65
3.9	Exercises .....	67
	References .....	70
<b>4</b>	<b>Metropolis Methods and Variants</b> .....	71
4.1	Introduction .....	71
4.2	The Metropolis-Hastings Algorithm .....	72
4.2.1	The Original Metropolis-Hastings Algorithm .....	72
4.2.2	Another Version of the Metropolis-Hastings Algorithm .....	74
4.2.3	Other Acceptance Probability Designs .....	74
4.2.4	Key Issues in Metropolis Design .....	75
4.3	The Independence Metropolis Sampler .....	75
4.3.1	The Eigenstructure of the IMS .....	77
4.3.2	General First Hitting Time for Finite Spaces .....	78
4.3.3	Hitting Time Analysis for the IMS .....	78
4.4	Reversible Jumps and Trans-Dimensional MCMC .....	80
4.4.1	Reversible Jumps .....	80
4.4.2	Toy Example: 1D Range Image Segmentation .....	81
4.5	Application: Counting People .....	85
4.5.1	Marked Point Process Model .....	85
4.5.2	Inference by MCMC .....	86
4.5.3	Results .....	87
4.6	Application: Furniture Arrangement .....	88
4.7	Application: Scene Synthesis .....	90
4.8	Exercises .....	95
	References .....	96
<b>5</b>	<b>Gibbs Sampler and Its Variants</b> .....	97
5.1	Introduction .....	97
5.2	Gibbs Sampler .....	99
5.2.1	A Major Problem with the Gibbs Sampler .....	101

5.3	Gibbs Sampler Generalizations .....	102
5.3.1	Hit-and-Run .....	103
5.3.2	Generalized Gibbs Sampler .....	103
5.3.3	Generalized Hit-and-Run .....	104
5.3.4	Sampling with Auxiliary Variables.....	105
5.3.5	Simulated Tempering .....	105
5.3.6	Slice Sampling .....	106
5.3.7	Data Augmentation .....	107
5.3.8	Metropolized Gibbs Sampler .....	108
5.4	Data Association and Data Augmentation .....	110
5.5	Julesz Ensemble and MCMC Sampling of Texture .....	112
5.5.1	The Julesz Ensemble: A Mathematical Definition of Texture.....	112
5.5.2	The Gibbs Ensemble and Ensemble Equivalence .....	115
5.5.3	Sampling the Julesz Ensemble .....	116
5.5.4	Experiment: Sampling the Julesz Ensemble .....	117
5.6	Exercises .....	119
	References .....	120
<b>6</b>	<b>Cluster Sampling Methods .....</b>	<b>123</b>
6.1	Introduction .....	123
6.2	Potts Model and Swendsen-Wang .....	124
6.3	Interpretations of the SW Algorithm .....	127
6.3.1	Interpretation 1: Metropolis-Hastings Perspective .....	128
6.3.2	Interpretation 2: Data Augmentation .....	131
6.4	Some Theoretical Results .....	135
6.5	Swendsen-Wang Cuts for Arbitrary Probabilities .....	137
6.5.1	Step 1: Data-Driven Clustering .....	138
6.5.2	Step 2: Color Flipping .....	139
6.5.3	Step 3: Accepting the Flip .....	140
6.5.4	Complexity Analysis.....	142
6.6	Variants of the Cluster Sampling Method .....	142
6.6.1	Cluster Gibbs Sampling: The “Hit-and-Run” Perspective .....	143
6.6.2	The Multiple Flipping Scheme .....	144
6.7	Application: Image Segmentation .....	145
6.8	Multigrid and Multi-level SW-cut .....	148
6.8.1	SW-Cuts at Multigrid .....	150
6.8.2	SW-cuts at Multi-level .....	152
6.9	Subspace Clustering .....	153
6.9.1	Subspace Clustering by Swendsen-Wang Cuts.....	155
6.9.2	Application: Sparse Motion Segmentation .....	158

6.10	C4: Clustering Cooperative and Competitive Constraints .....	163
6.10.1	Overview of the $C^+$ Algorithm .....	165
6.10.2	Graphs, Coupling, and Clustering .....	166
6.10.3	$C^+$ Algorithm on Flat Graphs .....	172
6.10.4	Experiments on Flat Graphs .....	175
6.10.5	Checkerboard Ising Model .....	176
6.10.6	$C^+$ on Hierarchical Graphs .....	181
6.10.7	Experiments on Hierarchical $C^+$ .....	183
6.11	Exercises .....	184
	References .....	186
<b>7</b>	<b>Convergence Analysis of MCMC</b> .....	<b>189</b>
7.1	Introduction .....	189
7.2	Key Convergence Topics .....	189
7.3	Practical Methods for Monitoring .....	191
7.4	Coupling Methods for Card Shuffling .....	193
7.4.1	Shuffling to the Top .....	193
7.4.2	Riffle Shuffling .....	194
7.5	Geometric Bounds, Bottleneck and Conductance .....	196
7.5.1	Geometric Convergence .....	196
7.6	Peskun's Ordering and Ergodicity Theorem.....	200
7.7	Path Coupling and Exact Sampling .....	201
7.7.1	Coupling From the Past .....	202
7.7.2	Application: Sampling the Ising Model.....	203
7.8	Exercises .....	205
	References .....	209
<b>8</b>	<b>Data Driven Markov Chain Monte Carlo</b> .....	<b>211</b>
8.1	Introduction .....	211
8.2	Issues with Segmentation and Introduction to DDMCMC .....	211
8.3	Simple Illustration of the DDMCMC.....	213
8.3.1	Designing MCMC: The Basic Issues .....	216
8.3.2	Computing Proposal Probabilities in the Atomic Spaces: Atomic Particles.....	217
8.3.3	Computing Proposal Probabilities in Object Spaces: Object Particles .....	219
8.3.4	Computing Multiple, Distinct Solutions: Scene Particles.....	220
8.3.5	The $\Psi$ -World Experiment .....	221
8.4	Problem Formulation and Image Models .....	223
8.4.1	Bayesian Formulation for Segmentation .....	223
8.4.2	The Prior Probability .....	224
8.4.3	The Likelihood for Grey Level Images .....	224

8.4.4	Model Calibration .....	226
8.4.5	Image Models for Color .....	228
8.5	Anatomy of Solution Space .....	229
8.6	Exploring the Solution Space by Ergodic Markov Chains .....	229
8.6.1	Five Markov Chain Dynamics.....	230
8.6.2	The Bottlenecks .....	232
8.7	Data-Driven Methods.....	233
8.7.1	Method I: Clustering in Atomic Spaces .....	233
8.7.2	Method II: Edge Detection .....	237
8.8	Computing Importance Proposal Probabilities .....	237
8.9	Computing Multiple Distinct Solutions .....	242
8.9.1	Motivation and a Mathematical Principle.....	242
8.9.2	A $K$ -Adventurers Algorithm for Multiple Solutions .....	243
8.10	Image Segmentation Experiments .....	244
8.11	Application: Image Parsing .....	247
8.11.1	Bottom-Up and Top-Down Processing.....	250
8.11.2	Generative and Discriminative Methods .....	251
8.11.3	Markov Chain Kernels and Sub-Kernels .....	252
8.11.4	DDMCMC and Proposal Probabilities .....	254
8.11.5	The Markov Chain Kernels .....	264
8.11.6	Image Parsing Experiments .....	272
8.12	Exercises .....	276
	References .....	277
<b>9</b>	<b>Hamiltonian and Langevin Monte Carlo .....</b>	<b>281</b>
9.1	Introduction .....	281
9.2	Hamiltonian Mechanics .....	282
9.2.1	Hamilton's Equations .....	282
9.2.2	A Simple Model of HMC.....	283
9.3	Properties of Hamiltonian Mechanics .....	284
9.3.1	Conservation of Energy.....	285
9.3.2	Reversibility .....	286
9.3.3	Symplectic Structure and Volume Preservation .....	286
9.4	The Leapfrog Discretization of Hamilton's Equations.....	288
9.4.1	Euler's Method .....	288
9.4.2	Modified Euler's Method .....	288
9.4.3	The Leapfrog Integrator .....	289
9.4.4	Properties of the Leapfrog Integrator .....	290
9.5	Hamiltonian Monte Carlo and Langevin Monte Carlo.....	292
9.5.1	Formulation of HMC .....	292
9.5.2	The HMC Algorithm.....	293
9.5.3	The LMC Algorithm .....	295

9.5.4	Tuning HMC .....	298
9.5.5	Proof of Detailed Balance for HMC .....	299
9.6	Riemann Manifold HMC .....	301
9.6.1	Linear Transformations in HMC .....	301
9.6.2	RMHMC Dynamics .....	305
9.6.3	RMHMC Algorithm and Variants .....	307
9.6.4	Covariance Functions in RMHMC .....	308
9.7	HMC in Practice .....	310
9.7.1	Simulated Experiments on Constrained Normal Distributions .....	310
9.7.2	Sampling Logistic Regression Coefficients with RMHMC .....	313
9.7.3	Sampling Image Densities with LMC: FRAME, GRADE and DeepFRAME .....	318
9.8	Exercises .....	323
	References .....	325
<b>10</b>	<b>Learning with Stochastic Gradient</b> .....	<b>327</b>
10.1	Introduction .....	327
10.2	Stochastic Gradient: Motivation and Properties .....	327
10.2.1	Motivating Cases .....	328
10.2.2	Robbins-Monro Theorem .....	331
10.2.3	Stochastic Gradient Descent and the Langevin Equation .....	333
10.3	Parameter Estimation for Markov Random Field (MRF) Models .....	336
10.3.1	Learning a FRAME Model with Stochastic Gradient .....	337
10.3.2	Alternate Methods of Learning for FRAME .....	339
10.3.3	Four Variants of the FRAME Algorithm .....	341
10.3.4	Experiments .....	345
10.4	Learning Image Models with Neural Networks .....	349
10.4.1	Contrastive Divergence and Persistent Contrastive Divergence .....	350
10.4.2	Learning a Potential Energy for Images with Deep Networks: DeepFRAME .....	351
10.4.3	Generator Networks and Alternating Backward Propagation .....	355
10.4.4	Cooperative Energy and Generator Models .....	360
10.5	Exercises .....	364
	References .....	365
<b>11</b>	<b>Mapping the Energy Landscape</b> .....	<b>367</b>
11.1	Introduction .....	367
11.2	Landscape Examples, Structures, and Tasks .....	367
11.2.1	Energy-Based Partitions of the State Space .....	371
11.2.2	Constructing a Disconnectivity Graph .....	372

---

11.2.3	ELM Example in 2D .....	374
11.2.4	Characterizing the Difficulty (or Complexity) of Learning Tasks .....	375
11.3	Generalized Wang-Landau Algorithm .....	377
11.3.1	Barrier Estimation for GWL Mapping .....	379
11.3.2	Volume Estimation with GWL .....	380
11.3.3	GWL Convergence Analysis .....	382
11.4	GWL Experiments .....	383
11.4.1	GWL Mappings of Gaussian Mixture Models.....	383
11.4.2	GWL Mapping of Grammar Models .....	389
11.5	Mapping the Landscape with Attraction-Diffusion .....	396
11.5.1	Metastability and a Macroscopic Partition.....	396
11.5.2	Introduction to Attraction-Diffusion .....	398
11.5.3	Attraction-Diffusion and the Ising Model .....	400
11.5.4	Attraction-Diffusion ELM Algorithm.....	402
11.5.5	Tuning ADELM.....	405
11.5.6	Barrier Estimation with AD .....	405
11.6	Mapping the SK Spin Glass Model with GWL and ADELM .....	407
11.7	Mapping Image Spaces with Attraction-Diffusion .....	410
11.7.1	Structure of Image Galaxies .....	410
11.7.2	Experiments .....	413
11.8	Exercises .....	417
	References .....	418
<b>Index</b> .....		<b>421</b>