

---

## Contents

|   |    |
|---|----|
| <b>Preface</b> .....  | V  |
| <b>Contributors</b> .....                                   | IX |
| <b>1 Introduction</b> .....                                 | 1  |
| 1.1 What Is a Hidden Markov Model? .....                    | 1  |
| 1.2 Beyond Hidden Markov Models .....                       | 4  |
| 1.3 Examples .....  | 6  |
| 1.3.1 Finite Hidden Markov Models .....                     | 6  |
| 1.3.2 Normal Hidden Markov Models .....                     | 13 |
| 1.3.3 Gaussian Linear State-Space Models .....              | 15 |
| 1.3.4 Conditionally Gaussian Linear State-Space Models .... | 17 |
| 1.3.5 General (Continuous) State-Space HMMs .....           | 24 |
| 1.3.6 Switching Processes with Markov Regime .....          | 29 |
| 1.4 Left-to-Right and Ergodic Hidden Markov Models .....    | 33 |
| <b>2 Main Definitions and Notations</b> .....               | 35 |
| 2.1 Markov Chains .....                                     | 35 |
| 2.1.1 Transition Kernels .....                              | 35 |
| 2.1.2 Homogeneous Markov Chains .....                       | 37 |
| 2.1.3 Non-homogeneous Markov Chains .....                   | 40 |
| 2.2 Hidden Markov Models .....                              | 42 |
| 2.2.1 Definitions and Notations .....                       | 42 |
| 2.2.2 Conditional Independence in Hidden Markov Models ...  | 44 |
| 2.2.3 Hierarchical Hidden Markov Models .....               | 46 |

---

## Part I State Inference

---

|          |   |     |
|----------|---|-----|
| <b>3</b> | <b>Filtering and Smoothing Recursions</b>                     | 51  |
| 3.1      | Basic Notations and Definitions                               | 53  |
| 3.1.1    | Likelihood  | 53  |
| 3.1.2    | Smoothing   | 54  |
| 3.1.3    | The Forward-Backward Decomposition                            | 56  |
| 3.1.4    | Implicit Conditioning (Please Read This Section!)             | 58  |
| 3.2      | Forward-Backward  | 59  |
| 3.2.1    | The Forward-Backward Recursions                               | 59  |
| 3.2.2    | Filtering and Normalized Recursion                            | 61  |
| 3.3      | Markovian Decompositions                                      | 66  |
| 3.3.1    | Forward Decomposition   | 66  |
| 3.3.2    | Backward Decomposition  | 70  |
| 3.4      | Complements   | 74  |
| <b>4</b> | <b>Advanced Topics in Smoothing</b>                           | 77  |
| 4.1      | Recursive Computation of Smoothed Functionals                 | 77  |
| 4.1.1    | Fixed Point Smoothing   | 78  |
| 4.1.2    | Recursive Smoothers for General Functionals                   | 79  |
| 4.1.3    | Comparison with Forward-Backward Smoothing                    | 82  |
| 4.2      | Filtering and Smoothing in More General Models                | 85  |
| 4.2.1    | Smoothing in Markov-switching Models                          | 86  |
| 4.2.2    | Smoothing in Partially Observed Markov Chains                 | 86  |
| 4.2.3    | Marginal Smoothing in Hierarchical HMMs                       | 87  |
| 4.3      | Forgetting of the Initial Condition                           | 89  |
| 4.3.1    | Total Variation   | 90  |
| 4.3.2    | Lipshitz Contraction for Transition Kernels                   | 95  |
| 4.3.3    | The Doeblin Condition and Uniform Ergodicity                  | 97  |
| 4.3.4    | Forgetting Properties   | 100 |
| 4.3.5    | Uniform Forgetting Under Strong Mixing Conditions             | 105 |
| 4.3.6    | Forgetting Under Alternative Conditions                       | 110 |
| <b>5</b> | <b>Applications of Smoothing</b>                              | 121 |
| 5.1      | Models with Finite State Space                                | 121 |
| 5.1.1    | Smoothing   | 122 |
| 5.1.2    | Maximum <i>a Posteriori</i> Sequence Estimation               | 125 |
| 5.2      | Gaussian Linear State-Space Models                            | 127 |
| 5.2.1    | Filtering and Backward Markovian Smoothing                    | 127 |
| 5.2.2    | Linear Prediction Interpretation                              | 131 |
| 5.2.3    | The Prediction and Filtering Recursions Revisited             | 137 |
| 5.2.4    | Disturbance Smoothing   | 143 |
| 5.2.5    | The Backward Recursion and the Two-Filter Formula             | 148 |
| 5.2.6    | Application to Marginal Filtering and Smoothing in<br>CGLSSMs | 155 |

|          |  |     |
|----------|--|-----|
| <b>6</b> | <b>Monte Carlo Methods</b>                         | 161 |
| 6.1      | Basic Monte Carlo Methods                          | 161 |
| 6.1.1    | Monte Carlo Integration                            | 162 |
| 6.1.2    | Monte Carlo Simulation for HMM State Inference     | 163 |
| 6.2      | A Markov Chain Monte Carlo Primer                  | 166 |
| 6.2.1    | The Accept-Reject Algorithm                        | 166 |
| 6.2.2    | Markov Chain Monte Carlo                           | 170 |
| 6.2.3    | Metropolis-Hastings                                | 171 |
| 6.2.4    | Hybrid Algorithms                                  | 179 |
| 6.2.5    | Gibbs Sampling                                     | 180 |
| 6.2.6    | Stopping an MCMC Algorithm                         | 185 |
| 6.3      | Applications to Hidden Markov Models               | 186 |
| 6.3.1    | Generic Sampling Strategies                        | 186 |
| 6.3.2    | Gibbs Sampling in CGLSSMs                          | 194 |
| <b>7</b> | <b>Sequential Monte Carlo Methods</b>              | 209 |
| 7.1      | Importance Sampling and Resampling                 | 210 |
| 7.1.1    | Importance Sampling                                | 210 |
| 7.1.2    | Sampling Importance Resampling                     | 211 |
| 7.2      | Sequential Importance Sampling                     | 214 |
| 7.2.1    | Sequential Implementation for HMMs                 | 214 |
| 7.2.2    | Choice of the Instrumental Kernel                  | 218 |
| 7.3      | Sequential Importance Sampling with Resampling     | 231 |
| 7.3.1    | Weight Degeneracy                                  | 231 |
| 7.3.2    | Resampling   | 236 |
| 7.4      | Complements  | 242 |
| 7.4.1    | Implementation of Multinomial Resampling           | 242 |
| 7.4.2    | Alternatives to Multinomial Resampling             | 244 |
| <b>8</b> | <b>Advanced Topics in Sequential Monte Carlo</b>   | 251 |
| 8.1      | Alternatives to SISR                               | 251 |
| 8.1.1    | I.I.D. Sampling                                    | 253 |
| 8.1.2    | Two-Stage Sampling                                 | 256 |
| 8.1.3    | Interpretation with Auxiliary Variables            | 260 |
| 8.1.4    | Auxiliary Accept-Reject Sampling                   | 261 |
| 8.1.5    | Markov Chain Monte Carlo Auxiliary Sampling        | 263 |
| 8.2      | Sequential Monte Carlo in Hierarchical HMMs        | 264 |
| 8.2.1    | Sequential Importance Sampling and Global Sampling | 265 |
| 8.2.2    | Optimal Sampling                                   | 267 |
| 8.2.3    | Application to CGLSSMs                             | 274 |
| 8.3      | Particle Approximation of Smoothing Functionals    | 278 |

|          |   |     |
|----------|---|-----|
| <b>9</b> | <b>Analysis of Sequential Monte Carlo Methods</b> | 287 |
| 9.1      | Importance Sampling                               | 287 |
| 9.1.1    | Unnormalized Importance Sampling                  | 287 |
| 9.1.2    | Deviation Inequalities                            | 291 |
| 9.1.3    | Self-normalized Importance Sampling Estimator     | 293 |
| 9.2      | Sampling Importance Resampling                    | 295 |
| 9.2.1    | The Algorithm                                     | 295 |
| 9.2.2    | Definitions and Notations                         | 297 |
| 9.2.3    | Weighting and Resampling                          | 300 |
| 9.2.4    | Application to the Single-Stage SIR Algorithm     | 307 |
| 9.3      | Single-Step Analysis of SMC Methods               | 311 |
| 9.3.1    | Mutation Step                                     | 311 |
| 9.3.2    | Description of Algorithms                         | 315 |
| 9.3.3    | Analysis of the Mutation/Selection Algorithm      | 319 |
| 9.3.4    | Analysis of the Selection/Mutation Algorithm      | 320 |
| 9.4      | Sequential Monte Carlo Methods                    | 321 |
| 9.4.1    | SISR  | 321 |
| 9.4.2    | I.I.D. Sampling                                   | 324 |
| 9.5      | Complements                                       | 333 |
| 9.5.1    | Weak Limits Theorems for Triangular Array         | 333 |
| 9.5.2    | Bibliographic Notes                               | 342 |

---

## Part II Parameter Inference

---

|           |   |     |
|-----------|---|-----|
| <b>10</b> | <b>Maximum Likelihood Inference, Part I:</b>      |     |
|           | <b>Optimization Through Exact Smoothing</b>       | 347 |
| 10.1      | Likelihood Optimization in Incomplete Data Models | 347 |
| 10.1.1    | Problem Statement and Notations                   | 348 |
| 10.1.2    | The Expectation-Maximization Algorithm            | 349 |
| 10.1.3    | Gradient-based Methods                            | 353 |
| 10.1.4    | Pros and Cons of Gradient-based Methods           | 358 |
| 10.2      | Application to HMMs                               | 359 |
| 10.2.1    | Hidden Markov Models as Missing Data Models       | 359 |
| 10.2.2    | EM in HMMs  | 360 |
| 10.2.3    | Computing Derivatives                             | 362 |
| 10.2.4    | Connection with the Sensitivity Equation Approach | 364 |
| 10.3      | The Example of Normal Hidden Markov Models        | 367 |
| 10.3.1    | EM Parameter Update Formulas                      | 367 |
| 10.3.2    | Estimation of the Initial Distribution            | 370 |
| 10.3.3    | Recursive Implementation of E-Step                | 371 |
| 10.3.4    | Computation of the Score and Observed Information | 374 |
| 10.4      | The Example of Gaussian Linear State-Space Models | 384 |
| 10.4.1    | The Intermediate Quantity of EM                   | 385 |
| 10.4.2    | Recursive Implementation                          | 387 |

|           |  |     |
|-----------|--|-----|
| 10.5      | Complements .....  | 389 |
| 10.5.1    | Global Convergence of the EM Algorithm .....   | 389 |
| 10.5.2    | Rate of Convergence of EM .....  | 392 |
| 10.5.3    | Generalized EM Algorithms .....  | 393 |
| 10.5.4    | Bibliographic Notes .....  | 394 |
| <b>11</b> | <b>Maximum Likelihood Inference, Part II:</b>  |     |
|           | <b>Monte Carlo Optimization</b> .....  | 397 |
| 11.1      | Methods and Algorithms .....   | 398 |
| 11.1.1    | Monte Carlo EM .....   | 398 |
| 11.1.2    | Simulation Schedules .....   | 403 |
| 11.1.3    | Gradient-based Algorithms .....  | 408 |
| 11.1.4    | Interlude: Stochastic Approximation and the<br>Robbins-Monro Approach .....            | 411 |
| 11.1.5    | Stochastic Gradient Algorithms .....   | 412 |
| 11.1.6    | Stochastic Approximation EM .....  | 414 |
| 11.1.7    | Stochastic EM .....  | 416 |
| 11.2      | Analysis of the MCEM Algorithm .....   | 419 |
| 11.2.1    | Convergence of Perturbed Dynamical Systems .....                                       | 420 |
| 11.2.2    | Convergence of the MCEM Algorithm .....  | 423 |
| 11.2.3    | Rate of Convergence of MCEM .....  | 426 |
| 11.3      | Analysis of Stochastic Approximation Algorithms .....                                  | 429 |
| 11.3.1    | Basic Results for Stochastic Approximation<br>Algorithms .....                         | 429 |
| 11.3.2    | Convergence of the Stochastic Gradient Algorithm .....                                 | 431 |
| 11.3.3    | Rate of Convergence of the Stochastic Gradient<br>Algorithm .....                      | 432 |
| 11.3.4    | Convergence of the SAEM Algorithm .....  | 433 |
| 11.4      | Complements .....  | 435 |
| <b>12</b> | <b>Statistical Properties of the Maximum Likelihood</b>                                |     |
|           | <b>Estimator</b> .....   | 441 |
| 12.1      | A Primer on MLE Asymptotics .....  | 442 |
| 12.2      | Stationary Approximations .....  | 443 |
| 12.3      | Consistency .....  | 446 |
| 12.3.1    | Construction of the Stationary Conditional<br>Log-likelihood .....                     | 446 |
| 12.3.2    | The Contrast Function and Its Properties .....   | 448 |
| 12.4      | Identifiability .....  | 450 |
| 12.4.1    | Equivalence of Parameters .....  | 451 |
| 12.4.2    | Identifiability of Mixture Densities .....   | 454 |
| 12.4.3    | Application of Mixture Identifiability to Hidden<br>Markov Models .....                | 455 |
| 12.5      | Asymptotic Normality of the Score and Convergence of the<br>Observed Information ..... | 457 |

|           |  |            |
|-----------|--|------------|
| 12.5.1    | The Score Function and Invoking the Fisher Identity . . .        | 457        |
| 12.5.2    | Construction of the Stationary Conditional Score . . . . .       | 459        |
| 12.5.3    | Weak Convergence of the Normalized Score . . . . .               | 464        |
| 12.5.4    | Convergence of the Normalized Observed Information . .           | 465        |
| 12.5.5    | Asymptotics of the Maximum Likelihood Estimator . . .            | 465        |
| 12.6      | Applications to Likelihood-based Tests . . . . .                 | 466        |
| 12.7      | Complements . . . . .  | 468        |
| <b>13</b> | <b>Fully Bayesian Approaches . . . . .</b>                       | <b>471</b> |
| 13.1      | Parameter Estimation . . . . .                                   | 471        |
| 13.1.1    | Bayesian Inference . . . . .                                     | 471        |
| 13.1.2    | Prior Distributions for HMMs . . . . .                           | 475        |
| 13.1.3    | Non-identifiability and Label Switching . . . . .                | 478        |
| 13.1.4    | MCMC Methods for Bayesian Inference . . . . .                    | 481        |
| 13.2      | Reversible Jump Methods . . . . .                                | 488        |
| 13.2.1    | Variable Dimension Models . . . . .                              | 488        |
| 13.2.2    | Green's Reversible Jump Algorithm . . . . .                      | 490        |
| 13.2.3    | Alternative Sampler Designs . . . . .                            | 498        |
| 13.2.4    | Alternatives to Reversible Jump MCMC . . . . .                   | 500        |
| 13.3      | Multiple Imputations Methods and Maximum <i>a Posteriori</i> . . | 501        |
| 13.3.1    | Simulated Annealing . . . . .                                    | 502        |
| 13.3.2    | The SAME Algorithm . . . . .                                     | 503        |

---

## Part III Background and Complements

---

|           |   |            |
|-----------|---|------------|
| <b>14</b> | <b>Elements of Markov Chain Theory . . . . .</b>      | <b>513</b> |
| 14.1      | Chains on Countable State Spaces . . . . .            | 513        |
| 14.1.1    | Irreducibility . . . . .                              | 513        |
| 14.1.2    | Recurrence and Transience . . . . .                   | 514        |
| 14.1.3    | Invariant Measures and Stationarity . . . . .         | 517        |
| 14.1.4    | Ergodicity . . . . .                                  | 519        |
| 14.2      | Chains on General State Spaces . . . . .              | 520        |
| 14.2.1    | Irreducibility . . . . .                              | 521        |
| 14.2.2    | Recurrence and Transience . . . . .                   | 523        |
| 14.2.3    | Invariant Measures and Stationarity . . . . .         | 534        |
| 14.2.4    | Ergodicity . . . . .                                  | 541        |
| 14.2.5    | Geometric Ergodicity and Foster-Lyapunov Conditions . | 548        |
| 14.2.6    | Limit Theorems . . . . .                              | 552        |
| 14.3      | Applications to Hidden Markov Models . . . . .        | 556        |
| 14.3.1    | Phi-irreducibility . . . . .                          | 557        |
| 14.3.2    | Atoms and Small Sets . . . . .                        | 558        |
| 14.3.3    | Recurrence and Positive Recurrence . . . . .          | 560        |

|  |     |
|--|-----|
| <b>15 An Information-Theoretic Perspective on Order</b>  |     |
| <b>Estimation</b> .....  | 565 |
| 15.1 Model Order Identification: What Is It About? .....   | 566 |
| 15.2 Order Estimation in Perspective .....   | 567 |
| 15.3 Order Estimation and Composite Hypothesis Testing .....                                       | 569 |
| 15.4 Code-based Identification .....   | 571 |
| 15.4.1 Definitions .....   | 571 |
| 15.4.2 Information Divergence Rates .....  | 574 |
| 15.5 MDL Order Estimators in Bayesian Settings .....   | 576 |
| 15.6 Strongly Consistent Penalized Maximum Likelihood<br>Estimators for HMM Order Estimation ..... | 577 |
| 15.7 Efficiency Issues .....   | 580 |
| 15.7.1 Variations on Stein's Lemma .....   | 581 |
| 15.7.2 Achieving Optimal Error Exponents .....   | 584 |
| 15.8 Consistency of the BIC Estimator in the Markov Order<br>Estimation Problem .....              | 587 |
| 15.8.1 Some Martingale Tools .....   | 589 |
| 15.8.2 The Martingale Approach .....   | 591 |
| 15.8.3 The Union Bound Meets Martingale Inequalities .....   | 592 |
| 15.9 Complements .....   | 600 |

---

## Part IV Appendices

---

|   |     |
|---|-----|
| <b>A Conditioning</b> .....                                 | 605 |
| A.1 Probability and Topology Terminology and Notation ..... | 605 |
| A.2 Conditional Expectation .....                           | 606 |
| A.3 Conditional Distribution .....                          | 611 |
| A.4 Conditional Independence .....                          | 614 |
| <b>B Linear Prediction</b> .....                            | 617 |
| B.1 Hilbert Spaces .....                                    | 617 |
| B.2 The Projection Theorem .....                            | 619 |
| <b>C Notations</b> .....                                    | 621 |
| C.1 Mathematical .....                                      | 621 |
| C.2 Probability .....                                       | 622 |
| C.3 Hidden Markov Models .....                              | 622 |
| C.4 Sequential Monte Carlo .....                            | 624 |
| <b>References</b> .....                                     | 625 |
| <b>Index</b> .....  | 645 |

Inference in Hidden Markov Models

Cappé, O.; Moulines, E.; Ryden, T.

2005, XVII, 653 p., Hardcover

ISBN: 978-0-387-40264-2