

# MONTE CARLO AND MARKOV CHAIN MONTE CARLO METHODS: A SHORT NOTE WITH KEY REFERENCES

BY HEDIBERT FREITAS LOPES

FEBRUARY 2021

## 1 Introduction

This note was primarily intended for my own use and for my MSc and PhD students and collaborators. It is by no means an exhaustive list and reveals my own research and interest bias, as well as my limited knowledge. Nonetheless, I believe it could become useful as an additional reference list for undergraduate and graduate course on Bayesian data analysis, computational statistics, Monte Carlo (MC) and Markov Chain Monte Carlo (MCMC) methods and the like. All comments and missing important references will be continuously added as I learn about them from potential readers. The most updated version of this note will be kept in my professional page at [hedibert.org](https://hedibert.org) and, as kindly suggested by Christian, at Cornell University's [arXiv.org](https://arxiv.org).

Speaking of Christian<sup>1</sup> and as will become obvious as you read through, this note is heavily based on Christian Robert and George Casella's (2011) *Statistical Science* paper "A Short History of MCMC: Subjective Recollections from Incomplete Data" (Volume 26, pages 102-115). Another useful review is Andrieu, de Freitas, Doucet and Jordan's (2003) *Machine Learning* paper "An Introduction to MCMC for Machine Learning" (Volume 50, pages 5-43). Two handbooks provide evidence of the breath and depth of MCMC research between the early 1990s to the early 2010s: *MCMC in Practice*, edited by Gilks, Richardson and Spiegelhalter (1996), and *Handbook of MCMC*, edited by Brooks, Gelman, Jones and Meng (2011). Additional review papers and/or books are listed below. Two textbooks on MC and MCMC methods for Bayesian inference are Robert and Casella (2004) *MC Statistical Methods* (12176 citations) and Gamerman and Lopes (2006) *MCMC: Stochastic Simulation for Bayesian Inference* (3860 citations); or 16000 citations combined.

Unfortunately, the vast and important contributions to the MCMC literature from 2000 to 2020 are omitted in this version. The omission include, but it not restricted to, reversible jump MCMC and dimension-varying methods, Hamiltonian MC and gradient-based algorithms, ancillarity-sufficiency interweaving and other adaptive schemes, as well as approximate Bayesian computation (ABC) and likelihood-free methods. Hopefully, I will add a few of them over the next couple of months. I also intend to include a bunch of references on sequential MC (SMC, aka particle filters) schemes.

The remainder of the notes is organized as follows. Sections 2 and 3 introduces the seminal MC articles from late 1940s to early 1950s, such as Metropolis and Ulam (1949) who introduce the Monte Carlo method and Metropolis *et al* (1953) that introduced the famous Metropolis algorithm. Hammersley and Handscomb (1964) wrote the first book on the Monte Carlo method. Section 4 introduces the very general Hastings algorithm, nowadays simply as Metropolis-Hastings algorithm. Section 5 lists the research developed in image and spatial analysis that eventually that would eventually become connected to the Metropolis-Hastings algorithm and the Gibbs sampler. Section 6 is all about the Gibbs sampler and the more general MCMC schemes for Bayesian inference. Finally, Sections 7 complement the above list with an additional set of review papers, textbooks and handbooks.

---

<sup>1</sup>Christian Robert's blog is a great source of book reviews, comments, and other peculiarities. I also recommend Andrew Gelman's blog and Dani Gamerman's blog (if you read Portuguese!) I highly recommend them.

## 2 Late 1940s to early 1950s: Metropolis, Ulam and Los Alamos

1. Metropolis and Ulam (1949) [The MC method](#), *JASA*, 44, 335-341.
2. Von Neumann (1951) [Various techniques used in connection with random digits](#), *Applied Math Series, Notes by G. E. Forsythe, in National Bureau of Standards*, 12, 36-38.
3. Metropolis, Rosenbluth, Rosenbluth, Teller and Teller (1953) [Equation of State Calculations by Fast Computing Machines](#), *Journal of Chemical Physics*, 21(6), 1087-1092.
4. Kahn and Marshall (1953) [Methods of reducing sample size in MC computations](#), *Journal of Operations Research Society of America*, 1, 263-278.
5. Hammersley and Morton (1954) [Poor man's MC](#), *JRSS-B*, 16, 23-38.
6. Marshall (1954) [The Use of Multistage Sampling Schemes in MC Computations](#), RAND Corporation.
7. History
  - (a) Ulam (1987) [Stan Ulam: Vita's Excerpts from Adventures of a Mathematician](#), *Los Alamos Science: Special Issue, Stanislaw Ulam 1909-1984*, 15, 8-22.
  - (b) Metropolis (1987) [The Beginning of the MC Method](#), *Los Alamos Science: Special Issue, Stanislaw Ulam 1909-1984*, 15, 125-130.
  - (c) Eckhardt (1987) [Stan Ulam, John von Neumann, and the MC Method](#), *Los Alamos Science: Special Issue, Stanislaw Ulam 1909-1984*, 15, 131-141.
  - (d) Metropolis (1992) [The Age of Computing: A Personal Memoir](#), *Daedalus*, 121(1), 119-130.
  - (e) Hitchcock (2003) [History of the MH algorithms](#), *American Statistician*, 57, 254-257.
  - (f) Gubernatis (2005) [Marshall Rosenbluth and the Metropolis algorithm](#), *Phys. Plasmas*, 12, 057303.

## 3 The 1960s: MC rules!

1. Hammersley and Handscomb (1964) [MC Methods](#).
2. Kraft and Wensrich (1964) [MC Methods: A Bibliography Covering the Period 1949 to 1963](#). A total of about 750 articles listed.
3. Shreider (1966) [The MC Method: The Method of Statistical Trials](#)

## 4 Early 1970s: Hastings

1. Hastings (1970) [MC sampling methods using Markov chains and their application](#), *Biometrika*, 57, 97-109.
2. Peskun (1973) [Optimum MC sampling using Markov chains](#), *Biometrika*, 6, 607-612.
3. History
  - (a) Rosenthal (2005) [W.K. Hastings, Statistician and Developer of the Metropolis-Hastings Algorithm](#).
  - (b) Dunson and Johndrow (2020) [The Hastings algorithm at fifty](#), *Biometrika*, 107(1), 1-23.

## 5 Mid 1970s to mid 1980s: Images, lattices, MC integration and imputation

1. Geman and Geman (1984) [Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images](#). *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6(6), 721-741.
2. Besag's work
  - (a) Besag (1974) [Spatial Interaction and the Statistical Analysis of Lattice Systems \(with discussion\)](#), *JRSS-B*, 36(2), 192-236.

- (b) Hammersley (1974) Discussion of Mr Besag's paper, *JRSS-B*, 36, 230-231.
  - (c) Besag (1975) [Statistical analysis of non-lattice data](#), *The Statistician*, 24, 179-195.
  - (d) Besag (1986) [On the statistical analysis of dirty pictures](#), *JRSS-B*, 48, 259-302.
  - (e) Besag and Clifford (1989) [Generalized MC significance tests](#), *Biometrika*, 76, 633-642.
  - (f) Besag, York and Mollié (1991) [Bayesian image restoration, with two applications in spatial statistics \(with discussion\)](#), *Ann. Inst. Statist. Math.*, 43, 1-59.
3. Devroye (1986) [Non-Uniform Random Variate Generation](#).
  4. Ripley (1987) *Stochastic Simulation*. Section 4.7
  5. Tanner and Wong (1987) [The Calculation of Posterior Distributions by Data Augmentation](#), *JASA*, 82(398), 528-540.
  6. Pearl (1987) [Evidential reasoning using stochastic simulation of causal models](#), *Artificial Intelligence*, 32(2), 245-257.
  7. Escobar (1987) [Estimating the means of several normal populations by nonparametric estimation of the distribution of the means](#). PhD Thesis, Department of Statistics, Yale University. Supervisor: John A. Hartigan. Winner of the 1998 Leonard J. Savage Thesis Award.
  8. Qian and Titterton (1990) [Parameter estimation for hidden Gibbs chains](#), *Statist. Probab. Lett.*, 10, 49-58.
  9. Qian and Titterton (1991) [Estimation of Parameters in Hidden Markov Models](#), *Philosophical Transactions: Physical Sciences and Engineering*, 337(1647), 407-428.
  10. MC integration for Bayesian econometrics
    - (a) Kloek and van Dijk (1978) [Bayesian Estimates of Equation System Parameters: An Application of Integration by MC](#), *Econometrica* 46(1), 1-19.
    - (b) Kloek and van Dijk (1980) [Further experience in Bayesian analysis using MC integration](#), *JoE*, 14(3),
    - (c) Geweke (1988) [Antithetic Acceleration of MC integration in Bayesian Inference](#), *JoE*, 38, 73-90.
    - (d) Geweke (1989) [Bayesian Inference in Econometric Models Using MC integration](#), *Econometrica*, 57, 1317-1340.
  11. History
    - (a) More about [Julian Besag](#).

## 6 Late 1980s to late 1990s: MCMC rules!

1. Gelfand and Smith (1990) [Sampling-Based Approaches to Calculating Marginal Densities](#), *JASA*, 85(410), 398-409. 8585 citations.
2. EM and data augmentation connections
  - (a) Gelfand and Smith (1990) [The Calculation of Posterior Distributions by Data Augmentation](#), *JASA*, 82(398), 528-540.
  - (b) Wei and Tanner (1990) [A MC implementation of the EM algorithm and the poor man's data augmentation algorithm](#), *JASA*, 85, 699-704.
  - (c) Meng and Rubin (1993) [Maximum likelihood estimation via the ECM algorithm: A general framework](#), *Biometrika*, 80, 267-278.
  - (d) Liu and Rubin (1994) [The ECME algorithm: A simple extension of EM and ECM with faster monotone convergence](#), *Biometrika*, 81, 633-648.
3. MCMC theory papers in the 1990s
  - (a) Müller (1991) [A generic approach to posterior integration and Gibbs sampling](#).
  - (b) Ritter and Tanner (1992) [Facilitating the Gibbs sampler: The Gibbs stopper and the Griddy-Gibbs sampler](#), *JASA*, 87(419), 861-868.
  - (c) Gilks and Wild (1992) [Adaptive Rejection Sampling for Gibbs Sampling](#), *Applied Statistics*, 41(2), 337-348.
  - (d) Neal (1993) [Probabilistic inference using MCMC](#).

- (e) Tierney (1994) **Markov chains for exploring posterior distributions (with discussion)**, *Ann. Statist.*, 22, 1701-1786. Convergence of ergodic averages + CLTs.
  - (f) Chan and Geyer (1994) **Discussion of "Markov chains for exploring posterior distribution"**, *Annals of Statistics*, 22, 1747-1758. Geometric ergodicity.
  - (g) Geyer (1995) **Conditioning in MCMC**, *JCGS*, 4(2), 148-154.
  - (h) Liu, Wong and Kong (1994) **Covariance structure of the Gibbs sampler with applications to the comparisons of estimators and augmentation schemes**, *Biometrika*, 81, 27-40. Rao-Blackwellization.
  - (i) Liu, Wong and Kong (1995) **Covariance structure and convergence rate of the Gibbs sampler with various scans**, *JRSS-B*, 57, 157-169. Rao-Blackwellization.
  - (j) Gilks, Best and Tan (1995) **Adaptive Rejection Metropolis Sampling Within Gibbs Sampling**, *Applied Statistics*, 44(4), 455-472.
  - (k) Rosenthal (1995) **Minorization conditions and convergence rates for MCMC**, *JASA*, 90, 558-566. Exact rates of convergence.
  - (l) Green (1995) **Reversible Jump MCMC Computation and Bayesian Model Determination**, *Biometrika*, 82, 711-732.
  - (m) Casella and Robert (1996) **Rao-Blackwellisation of sampling schemes**, *Biometrika*, 83(1), 81-94. More on Rao-Blackwellization.
  - (n) Mergensen and Tweedie (1996) **Rates of convergence of the Hastings and Metropolis algorithms**, *Annals of Statistics*, 24, 101-121. Speed of convergence.
  - (o) Neal (1996) **Sampling from multimodal distributions using tempered transitions**, *Statistics and Computing*, 6(4), 353-366.
  - (p) Grenander and Miller (1994) **Representation of Knowledge in Complex Systems (with discussion)**, *JRSS-B*, 56, 549-603.
  - (q) Cowles and Carlin (1996) **MCMC Convergence Diagnostics: A Comparative Review**, *JASA*, 91, 883-904. Comparing convergence diagnostics.
  - (r) Roberts, Gelman and Gilks (1997) **Weak convergence and optimal scaling of random walk Metropolis algorithms**, *Annals of Applied Probability*, 7, 110-120.
  - (s) Roberts and Sahu (1997) **Updating schemes, correlation structure, blocking and parameterization for the Gibbs sampler**, *JRSS-B*, 59, 291-317.
  - (t) Brooks and Roberts (1998) **Assessing Convergence of MCMC Algorithms**, *Statistics and Computing*, 8, 319-335.
  - (u) Roberts and Rosenthal (1999) **Convergence of slice sampler Markov chains**, *JRSS-B*, 61, 643-660.
  - (v) Hobert and Casela (1996) **The effect of improper priors on Gibbs sampling in hierarchical linear mixed models**, *JASA*, 91, 1461-1473. Full conditionals might not be enough.
4. MCMC application papers in the 1990s
- (a) Gelfand, Hills, Racine-Poon, Smith (1990) **Illustration of Bayesian inference in normal data models using Gibbs sampling**, *JASA*, 85, 972-982.
  - (b) Zeger and Karim (1991) **Generalized linear models with random effects; a Gibbs sampling approach**, *JASA*, 86, 79-86.
  - (c) Raftery and Banfield (1991) Discussion of Besag, York and Mollié (1991): **Stopping the Gibbs sampler, the use of morphology, and other issues in spatial statistics**, *Ann. Inst. Statist. Math.*, 43, 32-43.
  - (d) Gelfand, Smith and Lee (1992) **Bayesian analysis of constrained parameter and truncated data problems using Gibbs sampling**, *JASA*, 87, 523-532.
  - (e) Smith and Gelfand (1992) **Bayesian statistics without tears: A sampling-resampling perspective**. *Amer. Statist.*, 46, 84-88.
  - (f) Lange, Carlin and Gelfand (1992) **Hierarchical Bayes models for the progression of HIV infection using longitudinal CD4 T-cell numbers**, *JASA*, 87, 615-626.
  - (g) Carlin, Gelfand and Smith (1992) **Hierarchical Bayesian analysis of change point problems**, *Appl. Statist.*, 41, 389-405.
  - (h) George and McCulloch (1993) **Variable selection via Gibbs sampling**, *JASA*, 88, 881-889.
  - (i) Besag and Green (1993) **Spatial Statistics and Bayesian Computation (with discussion)**, *JRSS-B*, 55, 25-37.
  - (j) Albert and Chib (1993) **Bayesian analysis of binary and polychotomous response data**, *JASA*, 88, 669-679.
  - (k) Wang, Rutledge and Gionola (1993) **Marginal inferences about variance-components in a mixed linear model using Gibbs sampling**, *Gen. Sel. Evol.*, 25, 41-62.

- (l) Wakefield, Smith, Racine-Poon and Gelfand (1994) [Bayesian analysis of linear and non-linear population models using the Gibbs sampler](#), *Appl. Statist.*, 43, 201-222.
  - (m) Diebolt and Robert (1994) [Estimation of finite mixture distributions through Bayesian sampling](#), *JRSS-B*, 56(2), 363-375.
  - (n) Besag, Green, Higdon and Mengersen (1995) [Bayesian Computation and Stochastic Systems \(with discussion\)](#), *Statistical Science*, 10, 3-66.
  - (o) Latent variables - Damien, Wakefield and Walker (1999) [Gibbs sampling for Bayesian non-conjugate and hierarchical models by using auxiliary variables](#), *JRSS-B*, 61, 331-344.
  - (p) Neal (2003) [Slice sampling](#), *Annals of Statistics*, 31(3), 705-767. [Check his MCMC research page](#).
5. History: meetings around 1990

- (a) Adrian Smith (1986) 10-lecture summer course on hierarchical models at Bowling Green.
- (b) Adrian Smith (1989) talk at Bayesian workshop in Sherbrooke, Québec.
- (c) MCMC meeting (1991) at Ohio State University organized by Gelfand, Goel and Smith (February 15th-17th). Several presentations became highly recognized papers a few years down<sup>2</sup>:
  - i. Albert and Chib (1993) [Bayesian analysis of binary and polychotomous response data](#), *JASA*, 88, 669-679.
  - ii. Gelman and Rubin (1992) [Inference from iterative simulation using multiple sequences \(with discussion\)](#), *Statistical Science*, 7, 457-511.
  - iii. Geyer (1992) [Practical MC Markov chain \(with discussion\)](#), *Statistical Science*, 7, 473-511.
  - iv. Gilks (1992) Derivative-free adaptive rejection sampling for Gibbs sampling. In *Bayesian Statistics*, 4 (Bernardo, Berger, Dawid and Smith, eds.) 641-649.
  - v. Liu, Wong and Kong (1994) [Covariance structure of the Gibbs sampler with applications to the comparisons of estimators and augmentation schemes](#), *Biometrika*, 81, 27-40.
  - vi. Liu, Wong and Kong (1995) [Covariance structure and convergence rate of the Gibbs sampler with various scans](#), *JRSS-B*, 57, 157-169.
  - vii. Tierney (1994) [Markov chains for exploring posterior distributions \(with discussion\)](#), *Ann. Statist.*, 22, 1701-1786.
- (d) Royal Statistical Society May 1992 meeting - JRSS-B (1993) Volume 1 with 49 pages of discussion. “*Afternoon of the 11 Bayesians*”:
  - i. Smith and Roberts (1993) [Bayesian Computation Via the Gibbs Sampler and Related MCMC Methods](#), pages 3-23.
  - ii. Besag and Green (1993) [Spatial Statistics and Bayesian Computation](#), pages 25-37.
  - iii. Gilks, Clayton, Spiegelhalter, Best and McNeil (1993) [Modelling Complexity: Applications of Gibbs Sampling in Medicine](#), pages 39-52.
  - iv. [Discussion on the Meeting on the Gibbs Sampler and Other MCMC Methods](#), pages 53-102.<sup>3</sup>

## 6. Interviews

- (a) Carlin and Herring (2015) [A Conversation with Alan Gelfand](#), *Statistical Science*, 30(3), 413-422.
- (b) Dellaportas and Stephens (2020) [Interview with Professor Adrian FM Smith](#), *International Statistical Review*, 88(2), 265-279.

## 7 More reviews and books

### 7.1 Additional reviews

1. Müller (1991) [Monte Carlo integration in general dynamic models](#), *Contemporary Mathematics*, 115, 145-164.
2. Müller (1992) [Posterior integration in dynamic models](#), *Computing Science and Statistics*, 24, 318-324.

<sup>2</sup>A few other luminaries in the meeting were Arnold Zellner, Mike West, Rob McCulloch, Nick Polson, Brad Carlin, Peter Müller, John Wakefield, Ruey Tsay, Adrian Raftery and Michael Newton.

<sup>3</sup>Discussants were Clifford, Jennison, Wakefield, Phillips, Frigessi, Gray, Lawson, Foster, Ramgopal, Arslan, Constable, Kent, Wolff, Harding, Middleton, Diggle, Aykrod, Berzuini, Brewer, Aitken, Celeux, Diebolt, Critchley, Diaconis, Rosenthal, Robert, Gelfand, Lee, Gelman, Rubin, Geman, Geweke, Geyer, Gigli, Givens, Goodall, Dona-Lasinio, Grieve, Han, Higdon, Kolassa, Tanner, Kooperberg, Lewis, Lin, Thompson, Litton, Buck, Liu, Liu, Mardia, Marriot, Moller, Raftery, Shephard, Sinha, Sokal, Titterington, Wilson, York and Madigan.

3. Smith and Gelfand (1992) [Bayesian Statistics without Tears: A Sampling-Resampling Perspective](#), *The American Statistician*, 46(2), 84-88.
4. Casella and George (1992) [Explaining the Gibbs Sampler](#), *The American Statistician*, 46(3), 167-174.
5. Chib and Greenberg (1995) [Understanding the Metropolis-Hastings Algorithm](#), *The American Statistician*, 49(4), 327-335.
6. Chib and Greenberg (1996) [MCMC Simulation Methods in Econometrics](#), *Econometric Theory*, 12(3), 409-431.
7. Titterton (1997) [Introduction to Gelfand and Smith \(1990\) Sampling-Based Approaches to Calculating Marginal Densities](#). In *Breakthroughs in Statistics, Volume III*, Pages 519-550.
8. Kass, Carlin, Gelman and Neal (1998) [MCMC in Practice: A Roundtable Discussion](#), *The American Statistician*, 52(2), 93-100.
9. Dongarra and Sullivan (2000) [Guest editors' introduction: The top 10 algorithms](#), *Computing in Science and Engineering*, 2, 22-23.
10. Beichl and Sullivan (2000) [The Metropolis algorithm](#), *Computing in Science and Engineering*, 2, 65-69.
11. Cappé and Robert (2000) [MCMC: 10 Years and Still Running!](#) *JASA*, 95(452), 1282-1286.
12. Gelfand (2000) [Gibbs Sampling](#), *JASA*, 95(452), 1300-1304.
13. Müller (2001) [Markov Chain Monte Carlo Methods](#), in *International Encyclopedia of the Social & Behavioral Sciences*. pages 9236-9240.
14. Chib (2001) [MCMC Methods: Computation and Inference](#). In *Handbook of Econometrics*, Volume 5, 2001, 3569-3649 (chapter 57).
15. Dellaportas and Roberts (2003) [An Introduction to MCMC](#). In *Spatial Statistics and Computational Methods*, 1-41 (chapter 1).
16. Andrieu, Doucet and Robert (2004) [Computational advances for and from Bayesian analysis](#), *Statistical Science*, 19(1), 118-127.
17. Berg (2005) [Introduction to MCMC simulations and their statistical analysis](#). In *MCMC: Innovations and Applications* (Editors Kendall, Liang and Wang).
18. Chib (2011) [Introduction to Simulation and MCMC Methods](#). In *Handbook of Bayesian Econometrics*, chapter 5.
19. Robert (2016) [The Metropolis-Hastings algorithm](#).
20. Sharma (2017) [Markov Chain Monte Carlo Methods for Bayesian Data Analysis in Astronomy](#), *Annual Review of Astronomy and Astrophysics*, 55, 213-259.
21. Robert and Wu (2020) [MCMC Methods, a survey with some frequent misunderstandings](#).
22. Martin, Frazier and Robert (2020) [Computing Bayes: Bayesian Computation from 1763 to the 21st Century](#).

## 7.2 Textbooks and handbooks

1. Tanner (1996) *Tools for Statistical Inference: Methods for the Exploration of Posterior Distributions and Likelihood Functions*.
2. Chen, Shao and Ibrahim (2000) *Monte Carlo Methods in Bayesian Computation*.
3. Gentle (2003) *Random Number Generation and MC Methods (2nd edition)*.
4. Liu (2004) *MC Strategies in Scientific Computing (2nd edition)*.
5. Kendall, Liang and Wang (2005) *MCMC: Innovations and Applications*.
6. Marin and Robert (2007) *Bayesian Core: A Practical Approach to Computational Bayesian Statistics*.
7. Rizzo (2008) *Statistical Computing with R*.
8. Rubinstein and Kroese (2008) *Simulation and the MC Method*.
9. Albert (2009) *Bayesian Computation with R*.

10. Robert and Casella (2010) *Introducing MC Methods with R + solutions*.
11. Bolstad (2010) *Understanding Computational Bayesian Statistics*.
12. Liang, Liu and Carroll (2010) *Advanced MCMC Methods*.
13. Geweke, Koop and Van Dijk (2011) *The Oxford Handbook of Bayesian Econometrics*.
14. Givens and Hoeting (2012) *Computational Statistics (2nd edition)*.
15. Marin and Robert (2014) *Bayesian Essentials with R + solutions*.
16. Turkman, Paulino and Mueller (2019) *Computational Bayesian Statistics: An Introduction*.
17. McElreath (2020) *Statistical Rethinking: A Bayesian course with Examples in R and STAN (2nd edition)*.