►Mathematical and Statistical Modeling of Global Warming
►Statistical Inference in Ecology

References and Further Reading

Baranov FI (1918) On the question of the biological basis of fisheries. Proc Inst Icht Invest 1(1):81–128
Beverton RJH, Holt SJ (1957) On the dynamics of exploited fish populations, vol 19. Marine Fisheries, Great Britain Ministry of Agriculture, Fisheries and Food
Beverton RJH, Holt SJ (1993) On the dynamics of exploited fish populations, vol 11. Chapman and Hall, London
Brandão A, Butterworth DS, Johnston SJ, Glazer JP (2004) Using a GLMM to estimate the somatic growth rate trend for male South African west coast rock lobster, Jasus lalandii. Fish Res 70(2–3):339–349, 2004
Collie JS, Sissenwine MP (1983) Estimating population size from relative abundance data measured with error. Can J Fish Aquat Sci 40:1871–1879
Conser RJ (1991) A delury model for scallops incorporating length-based selectivity of the recruiting year-class to the survey gear and partial recruitment to the commercial fishery. Northeast Regional Stock Assessment Workshop Report, Woods Hole, MA, Res. Doc. SAW12/2, Appendix to CRD-91-03, 18pp
de Valpine P, Hilborn R (2005) State-space likelihoods for nonlinear fisheries timeseries. Can J Fish Aquat Sci 62(9):1937–1952
Fournier DA (1996) AUTODIF. A C++ array language extension with automatic differentiation for use in nonlinear modeling and statistic. Otter Research, Nanaimo, BC, 1996
Fulton EA, Smith ADM, Punt AE (2005) Which ecological indicators can robustly detect effects of fishing? ICES J Marine Sci 62(3):540
Gudmundsson G (2005) Stochastic growth. Can J Fish Aquat Sci 62(8):1746–1755
Gulland JA (1965) Estimation of mortality rates. Annex to Arctic Fisheries Working Group Report. ICES (Int. Counc. Explor. Sea) Document C.M. D:3 (mimeo), 1965
Helgason T, Gislason H (1979) VPA-analysis with species interaction due to predation. ICES C.M. 1979/G:52
Millar RB (1992) Modelling environmental effects on growth of cod: fitting to growth increment data versus fitting to size-at-age data. ICES J Marine Sci 49(3):289
Myers RA, Cadigan NG (1995) Statistical analysis of catch-at-age data with correlated errors. Can J Fish Aquat Sci (Print) 52(6):1265–1273
Myers RA, Bowen KG, Barrowman NJ (1999) Maximum reproductive rate of fish at low population sizes. Can J Fish Aquat Sci 56(12):2404–2419
Patterson KR (1999) Evaluating uncertainty in harvest control law catches using Bayesian Markov chain Monte Carlo virtual population analysis with adaptive rejection sampling and including structural uncertainty. Can J Fish Aquat Sci 56(2):208–221
Patterson K, Cook R, Darby C, Gavaris S, Kell L, Lewy P, Mesnil B, Punt A, Restrepo V, Skagen DW, Stefansson G (2001) Estimating uncertainty in fish stock assessment and forecasting. Fish Fish 2(2):125–157
Pauly D, Christensen V, Walters C (2000) Ecopath, Ecosim, and Ecospace as tools for evaluating ecosystem impact of fisheries. ICES J Marine Sci 57(3):697
Richards LJ (1991) Use of contradictory data sources in stock assessments. Fish Res 11(3–4):225–238
Skaug HJ (2002) Automatic differentiation to facilitate maximum likelihood estimation in nonlinear random effects models. J Comput Gr Stat pp 458–470
Skaug HJ, Fournier DA (2006) Automatic approximation of the marginal likelihood in non-Gaussian hierarchical models. Comput Stat Data Anal 51(2):699–709
Stefansson G (1998) Comparing different information sources in a multispecies context. In Funk F, Quinn II TJ, Heifetz J, Ianelli JN, Powers JE, Schweigert JF, Sullivan PJ, Zhang CI (eds.), Fishery Stock Assessment Models: Proceedings of the international symposium; Anchorage 1997, 15th Lowell Wakefield Fisheries Symposium, pp 741–758
Stefansson G (2003) Issues in multispecies models. Natural Res Model 16(4):415–437
Stefansson G, Palsson OK (1998) A framework for multispecies modelling of boreal systems. Rev Fish Biol Fish 8:101–104
Taylor L, Stefansson G (1999) Growth and maturation of haddock (Melanogrammus aeglefinus) in icelandic waters. J Northwest Atlantic Fish Sci 25:101–114
Taylor L, Begley J, Kupca V, Stefansson G (2007) A simple implementation of the statistical modelling framework Gadget for cod in Icelandic waters. African J Marine Sci 29(2):223–245, AUG 2007. ISSN 1814-232X. doi: 10.2989/AJMS.2007.29.2.7.190
Trenkel VM (2008) A two-stage biomass random effects model for stock assessment without catches: what can be estimated using only biomass survey indices? Can J Fish Aquat Sci 65(6):1024–1035
Venables WN, Dichmont CM (2004) GLMs, GAMs and GLMMs: an overview of theory for applications in fisheries research. Fish Res 70(2–3):319–337

Markov Chain Monte Carlo

Siddhartha Chib
Harry C. Hartkopf Professor of Econometrics and Statistics
Washington University in St. Louis, St. Louis, MO, USA

Introduction

Suppose that π is a probability measure on the probability space $(S, \mathcal{A})$, h is a measurable function from $S \to \mathrm{R}$, and one is interested in the calculation of the expectation

$$\bar{h} = \int h d\pi$$

assuming that the integral exists. In many problems, especially when the sample space S is multivariate or when the normalizing constant of π is not easily calculable, finding the value of this integral is not feasible either by numerial methods of integration (such as the method of quadrature) or by classical Monte Carlo methods (such as the method of rejection sampling). In such instances, it is usually possible to find $\bar{h}$ by Markov chain Monte Carlo, or MCMC for short, a method that stems from Metropolis et al. (1953)

in connection with work related to the hydrogen bomb project. It found early and wide use in computational statistical mechanics and quantum field theory where it was used to sample the coordinates of a point in phase space. Applications and developments of this method in statistics, in particular for problems arising in ►Bayesian statistics, can be traced to Hastings (1970), Geman and Geman (1984), Tanner and Wong (1987) and Gelfand and Smith (1990).

The idea behind MCMC is to generate a sequence of draws $\{\boldsymbol{\psi}^{(g)}, g \geq 0\}$ that follow a Markov chain (see ►Markov Chains) with the property that the unique invariant distribution of this Markov chain is the target distribution π. Then, after ignoring the first n_0 draws to remove the effect of the initial value $\boldsymbol{\psi}^{(0)}$, the sample

$$\{\boldsymbol{\psi}^{(n_0+1)}, ..., \boldsymbol{\psi}^{(n_0+M)}\}$$

for M large, is taken as an approximate sample from π and $\bar{h}$ estimated by the sample average

$$M^{-1}\sum_{g=1}^{M} h(\boldsymbol{\psi}^{(n_0+g)})$$

Laws of large numbers for Markov chains show that

$$M^{-1}\sum_{g=1}^{M} h(\boldsymbol{\psi}^{(n_0+g)}) \rightarrow \int h d\pi$$

as the simulation sample size M goes to infinity (Tierney 1994; Chib and Greenberg 1995; Chen et al. 2000; Liu 2001; Robert and Casella 2004).

A key reason for the interest in MCMC methods is that, somewhat surprisingly, it is straightforward to construct one or more Markov chains whose limiting invariant distribution is the desired target distribution. A leading method is the Metropolis–Hasting (M-H) method.

Metropolis–Hastings method

In the Metropolis–Hastings method, as the Hastings (1970) extension of the Metropolis et al. (1953) method is called, the Markov chain simulation is constructed by a recursive two step process.

Let $\pi(\boldsymbol{\psi})$ be a probabiliy measure that is dominated by a sigma-finite measure μ. Let the density of π with respect to μ be denoted by $p(\cdot)$. Let $q(\boldsymbol{\psi}, \boldsymbol{\psi}^{\dagger})$ denote a conditional density for $\boldsymbol{\psi}^{\dagger}$ given $\boldsymbol{\psi}$ with respect to μ. This density $q(\boldsymbol{\psi}, \cdot)$ is referred to as the proposal or candidate generating density. Then, the Markov chain in the M-H algorithm is constructed in two steps as follows.

Step 1 Sample a proposal value $\boldsymbol{\psi}^{\dagger}$ from $q(\boldsymbol{\psi}^{(g)}, \boldsymbol{\psi})$ and calculate the quantity (the *acceptance probability* or the *probability of move*)

$$\alpha(\boldsymbol{\psi}, \boldsymbol{\psi}^{\dagger}) = \begin{cases} \min\left[\frac{p(\boldsymbol{\psi}^{\dagger})q(\boldsymbol{\psi}^{\dagger}, \boldsymbol{\psi})}{p(\boldsymbol{\psi})q(\boldsymbol{\psi}, \boldsymbol{\psi}^{\dagger})}, 1\right] & \text{if } p(\boldsymbol{\psi})q(\boldsymbol{\psi}, \boldsymbol{\psi}^{\dagger}) > 0; \\ 1 & \text{otherwise}. \end{cases}$$

Step 2 Set

$$\boldsymbol{\psi}^{(g+1)} = \begin{cases} \boldsymbol{\psi}^{\dagger} & \text{with prob } \alpha(\boldsymbol{\psi}^{(g)}, \boldsymbol{\psi}^{\dagger}) \\ \boldsymbol{\psi}^{(g)} & \text{with prob } 1 - \alpha(\boldsymbol{\psi}^{(g)}, \boldsymbol{\psi}^{\dagger}) \end{cases}$$

If the proposal value is rejected then the next sampled value is taken to be the current value which means that when a rejection occurs the current value is repeated and the chain stays at the current value. Given the new value, the same two step process is repeated and the whole process iterated a large number of times.

Given the form of the acceptance probability $\alpha(\boldsymbol{\psi}, \boldsymbol{\psi}')$ it is clear that the M-H algorithm does not require knowledge of the normalizing constant of $p(\cdot)$. Furthermore, if the proposal density satisfies the symmetry condition $q(\boldsymbol{\psi}, \boldsymbol{\psi}') = q(\boldsymbol{\psi}', \boldsymbol{\psi})$, the acceptance probability reduces to $p(\boldsymbol{\psi}')/p(\boldsymbol{\psi})$; hence, if $p(\boldsymbol{\psi}') \geq p(\boldsymbol{\psi})$, the chain moves to $\boldsymbol{\psi}'$, otherwise it moves to $\boldsymbol{\psi}'$ with probability given by $p(\boldsymbol{\psi}')/p(\boldsymbol{\psi})$. The latter is the algorithm originally proposed by Metropolis et al. (1953).

A full expository discussion of this algorithm, along with a derivation of the method from the logic of reversibility, is provided by Chib and Greenberg (1995).

The M-H method delivers variates from π under quite general conditions. A weak requirement for a law of large numbers for sample averages based on the M-H output involve positivity and continuity of $q(\boldsymbol{\psi}, \boldsymbol{\psi}')$ for $(\boldsymbol{\psi}, \boldsymbol{\psi}')$ and connectedness of the support of the target distribution. In addition, if π is bounded then conditions for ergodicity, required to establish the central limit theorem (see ►Central Limit Theorems), are satisfied (Tierney 1994).

It is important that the proposal density be chosen to ensure that the chain makes large moves through the support of the invariant distribution without staying at one place for many iterations. Generally, the empirical behavior of the M-H output is monitored by the autocorrelation time of each component of $\boldsymbol{\psi}$ defined as

$$\left\{1 + 2\sum_{s=1}^{M} \rho_{ks}\right\},$$

where ρ_{ks} is the sample autocorrelation at lag s for the kth component of $\boldsymbol{\psi}$, and by the acceptance rate which is the proportion of times a move is made as the sampling proceeds. Because independence sampling produces an autocorrelation time that is theoretically equal to one, one tries to tune the M-H algorithm to get values close to one, if possible.

Different proposal densities give rise to specific versions of the M-H algorithm, each with the correct invariant distribution π. One family of candidate-generating densities is given by $q(\psi, \psi') = q(\psi' - \psi)$. The candidate ψ' is thus drawn according to the process $\psi' = \psi + z$, where z follows the distribution q, and is referred to as the random walk M-H chain. The random walk M-H chain is perhaps the simplest version of the M-H algorithm and is quite popular in applications. One has to be careful, however, in setting the variance of z because if it is too large it is possible that the chain may remain stuck at a particular value for many iterations while if it is too small the chain will tend to make small moves and move inefficiently through the support of the target distribution. Hastings (1970) considers a second family of candidate-generating densities that are given by the form $q(\psi, \psi') = q(\psi')$. Proposal values are thus drawn independently of the current location ψ.

Multiple-Block M-H

In applications when the dimension of ψ is large it is usually necessary to construct the Markov chain simulation by first grouping the variables ψ into smaller blocks. Suppose that two blocks are adequate and that ψ is written as (ψ_1, ψ_2), with $\psi_k \in \Omega_k \subseteq \Re^{d_k}$. In that case the M-H algorithm requires the specification of two proposal densities,

$$q_1\left(\psi_1, \psi_1^{\dagger} | \psi_2\right) \; ; \; q_2\left(\psi_2, \psi_2^{\dagger} | \psi_1\right),$$

one for each block ψ_k, where the proposal density q_k may depend on the current value of the remaining block. Also, define

$$\alpha\left(\psi_1, \psi_1^{\dagger} | \psi_2\right) = \min\left\{\frac{p\left(\psi_1^{\dagger}, \psi_2\right) q_1\left(\psi_1^{\dagger}, \psi_1 | \psi_2\right)}{p(\psi_1, \psi_2) q_1\left(\psi_1, \psi_1^{\dagger} | \psi_2\right)}, 1\right\}$$

and

$$\alpha\left(\psi_2, \psi_2^{\dagger} | \psi_1\right) = \min\left\{\frac{p\left(\psi_1, \psi_2^{\dagger}\right) q_2\left(\psi_2^{\dagger}, \psi_2 | \psi_1\right)}{p(\psi_1, \psi_2) q_2\left(\psi_2, \psi_2^{\dagger} | \psi_1\right)}, 1\right\},$$

as the probability of move for block ψ_k conditioned on the other block. Then, one cycle of the algorithm is completed by updating each block using a M-H step with the above probability of move, given the most current value of the other block.

Gibbs Sampling

A special case of the multiple-block M-H method is the Gibbs sampling method which was introduced by Geman and Geman (1984) in the context of image-processing and broadened for use in Bayesian problems by Gelfand and Smith (1990). To describe this algorithm, suppose that the parameters are grouped into two blocks (ψ_1, ψ_2) and each block is sampled according to the full conditional distribution of block ψ_k,

$$p(\psi_1 | \psi_2) \; ; \; p(\psi_2 | \psi_1)$$

defined as the conditional distribution under π of ψ_k given the other block. In parallel with the multiple-block M-H algorithm, the most current value of the other block is used in sampling the full conditional distribution. Derivation of these full conditional distributions is usually quite simple since, by ►Bayes' theorem, each full conditional is proportional to $p(\psi_1, \psi_2)$, the joint distribution of the two blocks. In addition, the introduction of latent or auxiliary variables can sometimes simplify the calculation and sampling of the full conditional distributions. Albert and Chib (1993) develop such an approach for the Bayesian analysis of categorical response data.

Concluding Remarks

Some of the recent theoretical work on MCMC methods is related to the question of the rates of convergence (Cai 2000; Fort et al. 2003; Jarner and Tweedie 2003; Douc et al. 2007) and in the development of adaptive MCMC methods (Atchade and Rosenthal; Andrieu and Moulines 2005; 2006).

The importance of MCMC methods in statistics and in particular Bayesian statistics cannot be overstated. The remarkable growth of Bayesian thinking over the last 20 years was made possible largely by the innovative use of MCMC methods. Software programs such as WINBUGS and the various MCMC packages in R have contributed to the use of MCMC methods in applications across the sciences and social sciences (Congdon 2006) and these applications are likely to continue unabated.

About the Author

Siddhartha Chib is the Harry Hartkopf Professor of Econometrics and Statistics at the Olin Business School, Washington University in St. Louis. He is a Fellow of the American Statistical Association and the Director of the NBER-NSF Seminar in Bayesian Inference in Econometrics and Statistics. Professor Chib has made several contributions in the areas of binary, categorical and censored response models, the Metropolis-Hastings algorithm and MCMC methods, the estimation of the marginal likelihood and Bayes factors, and in the treatment of hidden Markov and change-point models, and stochastic volatility and diffusion models. He has served as an Associate Editor

of the *Journal of the American Statistical Association* (Theory and Methods), *Journal of Econometrics*, the *Journal of Business and Economics Statistics*, and others. Currently he is an Associate Editor of the *Journal of Computational and Graphical Statistics*, and *Statistics and Computing*.

Cross References

►Bayesian Reliability Modeling
►Bayesian Statistics
►Bootstrap Methods
►Markov Chains
►Model Selection
►Model-Based Geostatistics
►Monte Carlo Methods in Statistics
►Non-Uniform Random Variate Generations
►Rubin Causal Model
►Small Area Estimation
►Social Network Analysis
►Statistics: An Overview

References and Further Reading

Albert JH, Chib S (1993) Bayesian analysis of binary and polychotomous response data. J Am Stat Assoc 88:669–679
Andrieu C, Moulines E (2006) On the ergodicity properties of some adaptive MCMC algorithms. Ann Appl Probab 16:1462–1505
Atchade YF, Rosenthal JS (2005) On adaptive Markov Chain Monte Carlo algorithms. Bernoulli 11:815–828
Cai HY (2000) Exact bound for the convergence of Metropolis chains. Stoch Anal Appl 18:63–71
Chen MH, Shao QM, Ibrahim JG (2000) Monte Carlo methods in Bayesian computation. Springer, New York
Chib S, Greenberg E (1995) Understanding the Metropolis-Hastings algorithm. Am Stat 49(4):327–335
Congdon P (2006) Bayesian statistical modelling, 2nd edn. Wiley, Chichester
Douc R, Moulines E, Soulier P (2007) Computable convergence rates for subgeometric ergodic Markov chains. Bernoulli 13:831–848
Fort G, Moulines E, Roberts GO, Rosenthal JS (2003) On the geometric ergodicity of hybrid samplers. J Appl Probab 40:123–146
Gelfand AE, Smith AFM (1990) Sampling-based approaches to calculating marginal densities. J Am Stat Assoc 85:398–409
Geman S, Geman D (1984) Stochastic relaxation, Gibbs distribution and the Bayesian restoration of images. IEEE Trans PAMI 6: 721–741
Hastings WK (1970) Monte-Carlo sampling methods using Markov chains and their applications. Biometrika 57:97–109
Jarner SF, Tweedie RL (2003) Necessary conditions for geometric and polynomial ergodicity of random-walk-type markov chains. Bernoulli 9:559–578
Liu JS (2001) Monte Carlo strategies in scientific computing. Springer, New York
Metropolis N, Rosenbluth AW, Rosenbluth MN, Teller AH, Teller E (1953) Equations of state calculations by fast computing machines. J Chem Phys 21:1087–1092
Robert CP, Casella G (2004) Monte Carlo statistical methods, 2nd edn. Springer, New York
Tanner MA, Wong WH (1987) The calculation of posterior distributions by data augmentation. J Am Stat Assoc 82:528–550 (with discussion)
Tierney L (1994) Markov-chains for exploring posterior distributions. Ann Stat 22:1701–1728

Markov Chains

Arnoldo Frigessi[1,2], Bernd Heidergott[3]
[1]Director
Norwegian Centre for Research-Based Innovation
"Statistics for Innovation," Oslo, Norway
[2]Professor
University of Oslo & Norwegian Computing Centre, Oslo, Norway
[3]Associate Professor
Vrije Universiteit, Amsterdam, The Netherlands

Introduction

Markov chains, which comprise Markov chains and ►Markov processes, have been successfully applied in areas as divers as biology, finance, manufacturing, telecommunications, physics and transport planning, and even for experts it is impossible to have an overview on the full richness of Markovian theory. Roughly speaking, Markov chains are used for modeling how a system moves from one state to another at each time point. Transitions are random and governed by a conditional probability distribution which assigns a probability to the move into a new state, given the current state of the system. This dependence represents the memory of the system. A basic example of a Markov chain is the so-called random walk defined as follows. Let $X_t \in N$, for $t \in \mathbb{N}$, be a sequence of random variables with initial value $X_0 = 0$. Furthermore assume that $P(X_{t+1} = X_t+1|X_t \geq 1) = p = 1-P(X_{t+1} = X_t-1|X_t \geq 1)$. The sequence $X = \{X_t : t \in \mathbb{N}\}$ is an example of a Markov chain (for a detailed definition see below) and the aspects of X one is usually interested in in Markov chain theory is (i) whether X returns to 0 in a finite number of steps (this holds for $0 \leq p \leq 1/2$), (ii) the expected number of steps until the chain returns to 0 (which is finite for $0 \leq p < 1/2$), and (iii) the limiting behavior of X_t.

In the following we present some realistic examples. A useful model in modeling infectious diseases assumes that there are four possible states: Susceptible (S), Infected (I), Immune (A), Dead (R). Possible transitions are from S to I, S or R; from I to A or R; from A to A or R; from R to R only. The transitions probabilities, from S to I, S to R