ASSESSMENT OF MCMC CONVERGENCE: A TIME SERIES AND DYNAMICAL SYSTEMS APPROACH

Rodney C. Wolff

School of Mathematical Sciences
Queensland University of Technology
GPO Box 2434
Brisbane QLD 4001, AUSTRALIA
email: r.wolff@qut.edu.au

Darfiana Nur, Kerrie L. Mengersen

Department of Mathematics
University of Newcastle
University Drive
Callaghan NSW 2308, AUSTRALIA

ABSTRACT

Important in the application of Markov chain Monte Carlo (MCMC) methods is the determination that a search run has converged. Given that such searches typically take place in high-dimensional spaces, there are many pitfalls and difficulties in making such assessments. In the present paper, we discuss the use of phase randomisation as tool in the MCMC context, provide some details of its distributional properties for time series which enable its use as a convergence diagnostic, and contrast its performance with a selection of other widely used diagnostics. Some brief comments on analytical results, obtained via Edgeworth expansion, are also made.

1. INTRODUCTION

MCMC methods support the application of Bayesian statistical methods through permitting complex distributions to be evaluated (specifically, by handling theoretically intractable integrals of high-dimensional probability density functions). Given the numerical and geometrical complexity of MCMC methods, assessment of convergence is a non-trivial task. Diagnostics for convergence are required in practical settings, and thus need to be accessible, accurate and fast.

In the theory of time series resampling, the method of phase randomisation has been used to generate so-called *surrogate* time series with the same first- and second-order properties as the original: see Theiler *et al.* (1992) and Timmer (1998), as well as Davison and Hinkley (1997) who use the term *phase scrambling*, and Braun and Kulperger (1997) who use the term *Fourier bootstrap*. In essence, one takes the discrete Fourier transform of a time series, replaces the phase with a new phase randomly chosen from the interval $(0, 2\pi)$, and back-transforms to obtain a new time series. Second-order properties are maintained by virtue of retaining the original amplitudes at their original locations in the original spectral estimate. If the original time series has an asymmetric marginal distribution then appropriate adjustments can be made in accordance with the so-called *rescaling methods* of Davison and Hinkley (1997), otherwise the standard algorithm suffices.

The algorithms as are follows. Denote the original series (of length n) as the array $x[t]$ with ranks r_t among the original unordered series.

Standard Algorithm

1. Compute the Discrete Fourier Transform $z[t] = DFT\,(x[t])$.
2. Randomise the phases; that is, randomly choose $\phi[t]$ from the uniform distribution of $(0, 2\pi)$, and put $z'[t] = z[t]\exp(i\phi[t])$.
3. Symmetrise the phases such that $Re\,(z''[t]) = Re\,(z'[t] + z'[n+1-t])\,/2$ and also $Im\,(z''[t]) = Im\,(z'[t] - z'[n+1-t])\,/2$.
4. Invert, putting $x'[t] = DFT^{-1}\,(z''[t])$.
5. The resulting series $x'[t]$ is the surrogate.

Rescaling Algorithm

1. Let $y_t = \Phi^{-1}\{r_t/(n+1)\}$, where Φ is the empirical distribution function of the original unordered series.
2. Apply the Standard Algorithm to $y_1, \ldots, y_n$, giving $Y_1^*, \ldots, Y_n^*$ (see above).
3. Set the surrogate series to be $X_t^* = x_{(r_t')}$, where r_t' is the rank of Y_t^* among $Y_1^*, \ldots, Y_n^*$.

One can use surrogate time series to test a null hypothesis that the original series arises from a linear, stochastic, Gaussian stationary process. (Note that the rejection of this hypothesis covers a wide range of alternatives.) If a statistic from the original series is denoted as V_0 and the corresponding statistic from the j'th surrogate is denoted as V_j, with $E(V_j) = \mu_V$ and $\operatorname{var}(V_J) = \sigma_V^2$, then one may use as the test statistic $|V_0 - \mu_V| / \sigma_V$, and calibrate against a Normal distribution, if appropriate. Timmer (1998) has illustrated this using the correlation dimension as the underlying statistic in the context of cyclostationary processes, demonstrating power to reject the null hypothesis in the presence of non-stationarity.

2. PHASE RANDOMISATION AND STATIONARITY

Second order properties, and some marginal shape properties, are known to be preserved under phase randomisation, the latter if the scaling method is used. We examine here the effect of phase randomisation on higher order moments and cumulants of a time series, in particular, to determine if conditions on linearity and stationarity are related to preservation of higher order properties under phase randomisation. In particular, for a time series $\{X_t\}$ with marginal mean μ, we treat higher central moments, of the form $E\{(X_t)\}^r$; higher order cumulants, of the form $E\left\{\prod_{j=1}^{r}(X_{t+k+j} - \mu)\right\}$; and higher order cross cumulants, of the lagged product form $E\{(X_t - \mu)^r (X_{t+k} - \mu)^r\}$. In each of these forms, $r = 3, 4, \ldots$, $k = 1, 2, \ldots$, and standard estimates were used in simulations.

Numerical experiments were based on some classical linear and non-linear time series models, including linear autoregression (AR), random walk (RW), bilinear stationary (BS), bilinear non-stationary (BN), GARCH stationary (GS), GARCH non-stationary (GN), threshold autoregression stationary (TS) and threshold autoregression non-stationary (TN). See Tong (1990) for a detailed discussion on the form and properties of these models.

Timeplots obtained from the numerical experiments showed broad agreement with the original data sets, and can be qualitatively compared as in the following table (using the rescaling method).

Model	AR	RW	BS	BN
Note	same	more symm.	same	same

Model	GS	GN	TS	TN
Note	larger vals.	same	same	same

When comparing the stationary with non-stationary models, it was sometimes possible to distinguish between them on the basis of higher order moments: the standard method produced zero values for odd moments; however, the rescaling method produced small values for the third moment for stationary series yet the same value as in the original series for non-stationary models. Thus, third order moments appear to have a reasonably good discriminatory ability for stationarity, and hence for convergence of MCMC procedures.

The behaviour of the higher order cumulants of the surrogates can be summarised according to the following table. The non-stationary models are shown in the last three rows of the table.

Model	Original	Standard	Rescaling
AR	small	odd near zero	all zero
BS	near zero	odd near zero	near zero
GS	near zero	near zero	near zero
TS	odd nr zero	odd near zero	small
RW	3rd, 4th small	odd smaller	small
BN	large	odd near zero	smaller
GN	large	smaller	smaller
TN	large	smaller	large

In addition, the distribution of higher order cumulants can be revealing in questions of stationarity, as the following table indicates. Modes refer to the number of modes of the empirical density function of the cumulants of surrogate series. (The standard method showed the same results as the rescaling method for all models.)

Model	Rescaling	Mode
AR	odd, even unimodal symm.	near zero
BS	unimodal	near zero
GS	unimodal, tails	near zero
TS	unimodal, tails	near zero
RW	multimodal	non-zero
BN	unimodal, tail	non-zero
GN	multimodal, tails	non-zero
TN	multimodal	non-zero

To summarise the results of these tables, we can comment as follows. In the case of the standard algorithm (i) higher order moments and cumulants are preserved for linear, Gaussian, stationary processes; (ii) higher moments are preserved for non-linear stationary processes, but not so for some higher order cumulants; (iii) second and cross-cumulants are not preserved for moderate and large lags if the process is linear non-stationary; and (iv) higher cumulants are not preserved for non-linear non-stationary processes. In the case of the rescaling algorithm (i) the method is inappropriate for linear, Gaussian, stationary processes as second order cumulants are not preserved; (ii) higher order cumulants are not preserved for non-linear stationary processes; (iii) higher order cumulants and cross-cumulants are not preserved for linear non-stationary processes; (iv) higher cumulants are substantially different from the originals for non-linear non-stationary models; and (v) smoothing densities of higher order cumulants are multimodal, or at least unimodal with heavy tails, for non-stationary processes, while remaining unimodal for stationary processes.

It is on the above basis that convergence (i.e., stationarity) can be concluded from a run of an MCMC algorithm. Nur *et al.* (2001) give further details of the above methodology. In that paper, the methods were applied to some well-known data sets, and was found to reject convergence where some other less dynamically-driven methods concluded convergence of chains.

3. PHASE RANDOMISATION AS AN MCMC CONVERGENCE DIAGNOSTIC: AN EXAMPLE

There is a variety of tests for convergence of MCMC algorithms. Raftery and Lewis (1996) reduce the output of a chain to a two-state Markov chain and apply analytically explicit results to the modified output. Clearly, this is a form of discretisation and there is the possibility that important information about the original process may be lost. Heidelberger and Welch (1983) adopt a spectral analysis approach, as does Geweke (1992). These and other algorithms are available in the software package CODA (Best *et al.*, 1995).

We briefly describe an analysis of a widely-used 'benchmark' data set, and compare the relative performance of the existing methods with the present method.

The example concerns mortality rates in 12 hospitals performing cardiac surgery on babies: see Spiegelhalter *et al.* (1994). The authors proposed a random effects model for the number of deaths, r_j, in hospital j, with true unknown mortality probability p_j, as follows: $r_j \sim \text{Binomial}(p_j, n_j)$ $(j = 1, \ldots, 12)$, $\log p_j = b_j$, $b_j \sim N(\mu, \tau)$, $\tau = 1/\sigma^2$, $\mu \sim N\left(0, 10^{-6}\right)$, $\tau \sim \Gamma\left(10^{-3}, 10^{-3}\right)$. The analysis was restricted to a short run of 200 epochs. The timeplot of the MCMC run appeared to be similar to a bilinear stationary time series, based on the simulations we described in the previous section. Smoothing densities of the higher order cumulant estimates were plainly unimodal around zero, and standard quantile plots ascertained Normality of the surrogates' cumulants (supported strongly by the Shapiro-Wilks test). We can thus conclude that the MCMC algorithm has converged. This is supported by the *diag* assessment in BUGS, by Raftery and Lewis' test, and by Heidelberger and Welch's test. However, Geweke's test fails for this example because of the very short run, although it passes if a considerably longer run is used.

A detailed discussion of this analysis, along those of two other data sets, is given by D Nur, KL Mengersen and RC Wolff in an as yet unpublished manuscript. It indicates that phase randomisation performs at least as well as other existing methods in the assessment of MCMC convergence and, moreover, it is more informative about higher order statistical structures which in turn can classify stationarity and linearity. Their work also suggests that higher order cumulants from surrogate time series appear to be asymptotically Normally distributed, thus providing a route to robust formal testing of convergence (stationarity) hypotheses, and calibration thereof. There also appears to be evidence that the Metropolis-Hastings algorithm results in a Markov chain which is geometrically ergodic to the average when the target density is log-concave in the tails.

4. THEORETICAL ISSUES FOR PHASE RANDOMISATION

To give the above methodology a firm theoretical basis, it is required to prove that third (and higher order) cumulants of a stochastic process can be bootstrapped with accuracy $o\left(n^{-1/2}\right)$. Results of Götze and Hipp (1983) can be employed to verify this.

Let $\{\varepsilon_t\}$ be independent and identically distributed (iid) random variables. Generalising the Wold Decomposition Theorem for stationary processes, we write $X_t = \mu + \sum b_j \varepsilon_{t-j} + \sum\sum b_{kj} \varepsilon_{t-j} \varepsilon_{t-k} + \ldots$, and clearly X_t is non-linear if any of the higher order coefficients are non-zero.

We consider the formal Edgeworth expansion of order $s-2$ of the third cumulant of X_t, as follows. Define $Y_{jkt} = X_t X_{t-j} X_{t-k} - \sigma_{kj}$, where σ_{kj} is the theoretical third cumulant of X_t. Let Y_t denote the matrix form of Y_{jkt}. Götze and Hipp (1994) obtain valid formal Edgeworth expansions for sums of weakly dependent random vectors, with error of approximation $o\left(n^{-(s-2)/2}\right)$ if the moments of order $s+1$ are bounded, a conditional Cramer condition holds, and the random vectors can be approximated by other random vectors which satisy a strong mixing condition and a Markov-type condition. We extend their result, as follows.

Assume the following.

(A1) Let $\{\varepsilon_t\}$ be an iid sequence such that $E\left(\varepsilon_t\right) = 0$, $E\left(\varepsilon_t^2\right) = 1$, $E\left(\varepsilon_t^{3q(s+1)}\right) < \infty$, for some $s \geq 3$, $q \geq 1$.

(A2) For linear processes, $\sum_{r=m}^{\infty} |b_r| \leq c \exp(-\alpha m)$, $\alpha > 0$, for all m sufficiently large.

(A3) Let f denote a strongly contracting and continuous differentiable function, and let ε_t have density satisfying $E |f(\varepsilon_1, \ldots, \varepsilon_d)| < \infty$, f being positive and continuous.

(A4) $\Gamma = \lim_{n\to\infty} \left(n^{-1/2} \sum_{t=1}^{n} Y_t\right)$ exists and is positive definite. Denote the quantity under the limit as S_n.

Suppose that $|f(x)| \leq M(1 + |x|^{s_0})$ for every vector x. If the assumptions as set out in Götze and Hipp (1994) hold, then there exists $\delta > 0$ not depending on f and M, and, for any $k > 0$, the exists a constant $C = C(M) > 0$ not depending on f, such that

$$\left| Ef(S_n) - \int f d\psi_{n,s} \right| \leq Cw\left(f, n^{-k}\right) + o\left(n^{-(\delta-2)/2}\right),$$

where ψ is a functional of signed measures relating to the determinant of Γ, the term $o(.)$ depends on f through M only, and w is a supremum operator on a Lipschitz condition for f constraining y to be less than n^{-k} in norm.

Under conditions (A1) through (A4), the result holds for X_t, and the required Edgeworth expansion can be obtained.

In an as yet unpublished manuscript in preparation by D Nur, RC Wolff and KL Mengersen, the conditions for this theorem are being confirmed.

REFERENCES

Best, N, Cowles, MK and Vines, L (1995) CODA Manual v0.30. MRC Biostatistics Unit, Cambridge.

Braun, WJ and Kulperger, RJ (1997) Properties of a Fourier bootstrap method for time series. *Communications in Statistics – Theory and Methods* **26**, 1329-1336.

Davison, AC and Kinkley, DV (1997) *Bootstrap Methods and their Applications.* Cambridge University Press.

Geweke, J (1992) Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments. In *Bayesian Statistics* **4** (JM Bernardo, J Berger, AP Dawid and AFM Smith, eds). Oxford University Press, 169-193.

Götze, F and Hipp, C (1994) Asymptotic expansions for sums of weakly dependent random vectors. *Zeitschrift Wahrscheinlichkeitstheorie Verwandte Gebiete* **64**, 211-239.

Heidelberger, P and Welch, PD (1983) A spectral method for confidence interval generation and run-length control in simulation. *Communication ACM* **24**, 233-245.

Nur, D, Wolff, R C and Mengersen, KL (2001) Phase randomisation: numerical study of higher cumulants behaviour. *Computational Statistics and Data Analysis*, in press.

Raftery, A and Lewis, S (1992) Implementing MCMC. In *MCMC in Practice* (WR Gilks, ST Richardson and DJ Spiegelhalter, eds). Chapman and Hall: London, 115-130.

Spiegelhalter, DJ, Thomas, A, Best, NG and Gilks, WR (1994) BUGS: Bayesian Inference Using the Gibbs Sampler v40.4. Technical Report, MRC Biostatistics Unit, University of Cambridge.

Theiler, J, Eubank, S, Longtin, A, Galdrikian, B and Farmer, JD (1992) Testing for nonlinearity in time series: the method of surrogate data. *Physica D* **58**, 77-94.

Timmer, J (1998) Power of surrogate data testing with respect to non-stationarity. *Physical Review E* **58**, 5153-5156.

Tong, H (1990) *Non-linear Time Series: a Dynamical System Approach.* Oxford University Press.