ELSEVIER

Statistics & Probability Letters 60 (2002) 289–295

STATISTICS & PROBABILITY LETTERS

www.elsevier.com/locate/stapro

Establishing geometric drift via the Laplace transform of symmetric measures

Niels Richard Hansen*, Ernst Hansen

Department of Statistics and Operations Research, University of Copenhagen, Universitetsparken 5, Copenhagen 2100, Denmark

Received February 2001; received in revised form July 2002

Abstract

A new kind of drift function based on the Laplace transform of rotation symmetric measures on $\mathbb{R}^k$ is introduced and applied to the class of the so-called *affine* Markov chains. An example is given where this approach provides a better criterion for geometric drift than standard drift functions.

Keywords: Markov chains; Affine Markov chains; Drift functions; Geometric drift; Geometric ergodicity

1. Introduction

Let $(E, \mathbb{E})$ be the state space and let P be a Markov kernel on E. P is called V-geometrically ergodic if $V : E \to [1, \infty[$ is a measurable function and if there exist constants R and $\rho < 1$ such that

$$\|P^n(x, \cdot) - \pi\|_V \leqslant RV(x)\rho^n,$$

where π is the (necessarily unique) invariant probability measure. The V-norm is defined in Meyn and Tweedie (1993). To verify geometric ergodicity one can use the so-called drift criterion. Hence if P is φ-irreducible, aperiodic and has V-geometric drift towards a small set with drift function $V : E \to [1, \infty[$, i.e. if there exist constants b and $\beta < 1$ and a small set C such that

$$PV(x) := \int V(y)P(x, \mathrm{d}y) \leqslant \beta V(x) + b1_C(x), \tag{1}$$

then P is V-geometrically ergodic (Meyn and Tweedie, 1993, Theorem 15.0.1).

* Corresponding author.
E-mail address: richard@stat.ku.dk (N.R. Hansen).

PII: S0167-7152(02)00288-2

The verification of irreducibility and aperiodicity is sometimes trivial but can also be very difficult indeed. Results on this issue can be found in, for instance, Meyn and Tweedie (1993), Mokkadem (1990) or Doukhan (1994). We will not discuss how to verify irreducibility or aperiodicity but instead focus on a sufficient criterion for the drift inequality (1) to hold.

The main result is Theorem 2, which states that for $E=\mathbb{R}^k$ geometric drift towards a compact set can be achieved if the Markov chain on average tends to move towards the center of the space when it is outside some compact set. We need several assumptions to prove this result. First of all we place the result in the frame of what we call affine Markov chains. Furthermore, we need the existence of some exponential moments of the transition probabilities and it is also assumed that the conditional variance is bounded. Though the affine Markov chain structure is convenient and the existence of exponential moments is necessary for the approach given in this paper, it is the bounded conditional variance that is believed to be essential for a result like Theorem 2 to hold. Finally we illustrate in Example 4 how Theorem 2 gives stronger results than what can be achieved by another standard drift function.

2. Affine Markov chains

We will consider a special class of Markov chains on $\mathbb{R}^k$, which have a nice and interpretable structure. On $\mathbb{R}^k$ we denote the usual inner product by $\langle\cdot,\cdot\rangle$ and the related 2-norm by $|x|^2=\langle x,x\rangle$. Also let $M(k)$ denote the set of $k\times k$-matrices.

Definition 1. A Markov chain $(X_n)_{n\in\mathbb{N}_0}$ on $\mathbb{R}^k$ is affine if there exist measurable maps $\xi:\mathbb{R}^k\to\mathbb{R}^k$ and $\Phi:\mathbb{R}^k\to M(k)$ and a sequence of iid stochastic variables $(W_n)_{n\in\mathbb{N}}$ independent of X_0 such that

$$X_{n+1}=\xi(X_n)+\Phi(X_n)W_{n+1} \tag{2}$$

for $n\in\mathbb{N}_0$.

The sequence $(W_n)_{n\in\mathbb{N}}$ is called the innovation sequence and the distribution of W_1 (and hence of all the W_n's), the innovation distribution. If v is the innovation distribution we can identify the transition probabilities as the Markov kernel P given by

$$P(x,\cdot)=A(x)(v), \tag{3}$$

where $A(x):\mathbb{R}^k\to\mathbb{R}^k$ is the affine map $A(x)(y)=\xi(x)+\Phi(x)y$. Thus, if we define an *affine Markov kernel* to be a kernel of form (3) for some ξ, Φ and a probability measure v, we have that the transition probabilities for an affine Markov chain are given by an affine Markov kernel.

If v is normalized, i.e. satisfies $\int xv(\mathrm{d}x)=0$ and $\int xx^{\mathrm{T}}v(\mathrm{d}x)=I$, which is often assumed, it is clear that $E(X_{n+1}|X_n=x)=\xi(x)$ and we also find that $V(X_{n+1}|X_n=x)=\Phi(x)\Phi(x)^{\mathrm{T}}$. This gives the obvious interpretation of ξ and Φ as conditional mean and scale functions. In this paper we will, however, only need to assume that v is centered, i.e. $\int xv(\mathrm{d}x)=0$.

Several classes of Markov chains, some of them being more general than the affine class introduced here, have been considered in the literature. For instance, chapter 7 in Meyn and Tweedie (1993) deals with nonlinear state space models in general, Tjøstheim (1990) considers affine Markov chains with constant Φ and Tong (1990) considers a class of Markov chains similar to but more general than the affine Markov chains, where the dynamic of the Markov chain is separated into a deterministic term and a noise term. The affine class discussed here is easy to handle and interpret and includes many of the Markov chains being used in applications, like AR-models (Meyn and Tweedie, 1993), ARCH-models (Hansen and Rahbek, 1998; Engle and Kroner, 1995) and SETAR-models (Tong, 1990; Meyn and Tweedie, 1993).

2.1. Geometric drift for affine Markov chains

If we can verify φ-irreducibility and aperiodicity, then we can show geometric ergodicity by showing that the drift inequality is fulfilled for some drift function. One example of a simple but useful drift function is $V(x)=\exp(s|x|)$ for some suitable $s>0$. To use this drift function, we need the innovation distribution to have light tails in the sense that

$$\int \exp(s|x|)\nu(\mathrm{d}x) < \infty \tag{4}$$

for some $s>0$. If we assume that Φ is a bounded map and fix $s>0$ small enough, then with

$$\delta(s)=\sup_x \int \exp(s|\Phi(x)y|)\nu(\mathrm{d}y)\in[1,\infty],$$

we get that

$$PV(x)=\int \exp(s|\xi(x)+\Phi(x)y|)\nu(\mathrm{d}y)\leqslant V(x)\delta(s)\exp(s(|\xi(x)|-|x|)),$$

hence if ξ is bounded on the compact sets this shows that we have geometric drift with drift function $V(x)=\exp(s|x|)$ if

$$\limsup_{|x|\to\infty} |\xi(x)|-|x| < -\frac{1}{s}\log(\delta(s)). \tag{5}$$

The constant on the right-hand side is negative and as shown in Example 4 it may be strictly bounded away from 0.

Another class of widely used drift functions is $1+|x|^s$ for $s>0$—especially, the case $s=2$ is easy to use. These drift functions tend to give weaker results but are applicable to a wider class of Markov chains, for instance, the ARCH-models with unbounded Φ (Hansen and Rahbek, 1998). One could also try to use a drift function like $s\mapsto\exp(s|x|^2)$ instead, but then we need the innovation distribution to have even faster decaying tails. See also Theorem 16.3.1 in Meyn and Tweedie (1993) for the general possibility of using $\exp(sV(x))$ as a drift function for some function V.

3. Laplace transforms as drift functions

In this section we show that by choosing another drift function, we can substitute the (in general strictly negative) constant in (5) with 0 without assuming more than the integrability condition (4).

Let κ be a rotation symmetric probability measure on $\mathbb{R}^k$ with compact support and let $V_\kappa : \mathbb{R}^k \to \mathbb{R}$ be the Laplace transform of κ, i.e.

$$V_\kappa(x) = \int_{\mathbb{R}^k} \exp(\langle y, x\rangle)\kappa(\mathrm{d}y). \tag{6}$$

If we fix a unit vector e and for each $x \in \mathbb{R}^k$ find a rotation O_x such that $O_x x = |x|e$, then because κ is rotation symmetric we get that

$$V_\kappa(x) = \int_{\mathbb{R}^k} \exp(|x|\langle y, e\rangle)\kappa(\mathrm{d}y) = \int_{\mathbb{R}} \exp(|x|t)\mu(\mathrm{d}t)$$

with μ being the transformation of κ under the map $y \mapsto \langle y, e\rangle$.

For the result in this paper it is sufficient to consider a special family of rotation symmetric measures ω_s for $s > 0$ with ω_s being the probability measure concentrated on the sphere of radius s and invariant under rotations. This is the so-called surface measure or Lebesgue measure on the sphere normalized to a probability measure. In this case, we write V_s for V_{ω_s} and with $\omega = \omega_1$ and S^{k-1} denoting the unit sphere in $\mathbb{R}^k$ we have that

$$V_s(x) = \int_{\mathbb{R}^k} \exp(\langle y, x\rangle)\omega_s(\mathrm{d}y) = \int_{S^{k-1}} \exp(s\langle y, x\rangle)\omega(\mathrm{d}y) \tag{7}$$

$$= \int_{-1}^{1} \exp(s|x|t)\mu(\mathrm{d}t). \tag{8}$$

Note that in the one-dimensional case $V_s(x) = \cosh(sx)$. For the application of V_s as a drift function, the fact that it has the form of a Laplace transform as given by (7) and not just the integral form given by (8) will play an important role.

Consider an affine Markov kernel P with innovation distribution v. We will assume that v satisfies (4) for some $s_0 > 0$ and we define the map

$$\psi(s, y) = \int \exp(s\langle y, x\rangle)v(\mathrm{d}x)$$

for $(s, y) \in [0, s_0] \times B(0, 1)$, with $B(0, 1) = \{x \in \mathbb{R}^k \mid |x| \leqslant 1\}$ being the closed unit ball. By decreasing s_0 we can increase $B(0, 1)$ to be any closed ball, and hence we can assume that ψ is defined on $[0, s_0] \times B(0, d)$ for some suitable large d and sufficiently small s_0 if necessary.

Theorem 2. *Let P be an affine Markov kernel with v satisfying (4). Assume that v is centered, that Φ is bounded, that ξ is bounded on the compact sets and that*

$$\limsup_{|x|\to\infty} |\xi(x)| - |x| < 0 \tag{9}$$

holds, then P has geometric drift towards a compact set with drift function $V_{\tilde{s}}$ for some suitable $\tilde{s} > 0$.

Proof. By Tonelli and the integral transformation theorem we get

$$
\begin{aligned}
PV_s(x) &= \int V_s(z)P(x,\mathrm{d}z)\\
&= \int\int_{S^{k-1}} \exp(s\langle y,z\rangle)\omega(\mathrm{d}y)P(x,\mathrm{d}z)\\
&= \int_{S^{k-1}}\int \exp(s\langle y,\xi(x)+\Phi(x)z\rangle)v(\mathrm{d}z)\omega(\mathrm{d}y)\\
&= \int_{S^{k-1}} \exp(s\langle y,\xi(x)\rangle)\int \exp(s\langle \Phi(x)^{\mathrm{T}}y,z\rangle)v(\mathrm{d}z)\omega(\mathrm{d}y)\\
&= \int_{S^{k-1}} \exp(s\langle y,\xi(x)\rangle)\psi(s,\Phi(x)^{\mathrm{T}}y)\omega(\mathrm{d}y)
\end{aligned}
$$

for $s>0$ sufficiently small.

Let $K=\{\Phi(x)y\in\mathbb{R}^k \mid x\in\mathbb{R}^k \text{ and } y\in S^{k-1}\}$ be the image of the map $(x,y)\mapsto\Phi(x)y$. This set is bounded since Φ is bounded and therefore K is contained in a compact, convex box $B_r=\{x\in\mathbb{R}^k \mid |x|_\infty\leqslant r\}$ for some $r>0$, with $|x|_\infty=\max_{i=1,\dots,k}|x_i|$ being the max-norm. Assume now that s_0 is chosen sufficiently small such that ψ is defined on $[0,s_0]\times B(0,d)$ with $B_r\subseteq B(0,d)$.

The box B_r has exactly 2^k extreme points, which we will denote by $z_1,\dots,z_{2^k}$, and the map

$$z\mapsto\psi(s,z)$$

for fixed $s\leqslant s_0$ is convex, hence it attains its maximal value over a compact, convex set in one of the extreme points. With

$$\tilde{\psi}(s)=\max_{j=1,\dots,2^k}\psi(s,z_j),$$

we therefore get that $\psi(s,z)\leqslant\tilde{\psi}(s)$ for all $z\in B_r\supseteq K$. This gives us

$$PV_s(x)\leqslant\tilde{\psi}(s)\int_{S^{k-1}}\exp(s\langle y,\xi(x)\rangle)\omega(\mathrm{d}y)=\tilde{\psi}(s)\int_{-1}^{1}\exp(s|\xi(x)|t)\mu(\mathrm{d}t).$$

Now assume that R is given such that for $|x|\geqslant R$ we have $|\xi(x)|\leqslant-\alpha+|x|$ for some $\alpha>0$. Put

$$\varphi_j(s)=\psi(s,z_j)\exp(-s\alpha)=\int\exp(s(\langle z_j,x\rangle-\alpha))v(\mathrm{d}x)$$

for $j=1,\dots,2^k$, then φ_j is convex and differentiable at 0 with

$$\varphi_j'(0)=\int\langle z_j,x\rangle v(\mathrm{d}x)-\alpha=-\alpha,$$

since v is centered, and $\varphi_j(0)=1$, it follows that we can find $s_j>0$ such that $\varphi_j(s)<1$ for all $s\leqslant s_j$. With $\tilde{s}=\min_{j=1,\dots,2^k}\{s_j\}>0$, we get that

$$c=\tilde{\psi}(\tilde{s})\exp(-\tilde{s}\alpha)<1.$$

The map $u \mapsto \int_{-1}^{1} \exp(ut)\mu(\mathrm{d}t)$ is increasing for $u \geqslant 0$, and therefore

$$\int_{-1}^{1} \exp(\tilde{s}|\xi(x)|t)\mu(\mathrm{d}t) \leqslant \int_{-1}^{1} \exp(-\tilde{s}\alpha t)\exp(\tilde{s}|x|t)\mu(\mathrm{d}t)$$

for $|x| \geqslant R$ and since $t \mapsto \exp(-\tilde{s}\alpha t)$ is continuous on $[-1,1]$, it follows from Lemma 3 that

$$\frac{\int_{-1}^{1} \exp(-\tilde{s}\alpha t)\exp(\tilde{s}|x|t)\mu(\mathrm{d}t)}{\int_{-1}^{1} \exp(\tilde{s}|x|t)\mu(\mathrm{d}t)} \to \exp(-\tilde{s}\alpha)$$

for $|x| \to \infty$. This shows that

$$\frac{PV_{\tilde{s}}(x)}{V_{\tilde{s}}(x)} \leqslant \tilde{\psi}(\tilde{s}) \frac{\int_{-1}^{1} \exp(-\tilde{s}\alpha t)\exp(\tilde{s}|x|t)\mu(\mathrm{d}t)}{\int_{-1}^{1} \exp(\tilde{s}|x|t)\mu(\mathrm{d}t)} \to \tilde{\psi}(\tilde{s})\exp(-\tilde{s}\alpha) = c < 1$$

for $|x| \to \infty$ and it follows that for fixed $\beta \in (c,1)$, we can choose an $R_0 \geqslant R$, such that for $|x| \geqslant R_0$ we have

$$\frac{PV_{\tilde{s}}(x)}{V_{\tilde{s}}(x)} \leqslant \beta.$$

Since ξ is bounded on compact sets, we have

$$PV_{\tilde{s}}(x) \leqslant \beta V_{\tilde{s}}(x) + b1_{B(0,R_0)}(x)$$

with $b = \sup_{|x| \leqslant R_0} PV_{\tilde{s}}(x) < \infty$. □

The following lemma was used in the theorem. The proof is elementary and is omitted.

Lemma 3. *If μ is a probability measure on $[-1,1]$ and v_h is the probability measure defined by*

$$v_h(B) = \frac{1}{\int \exp(ht)\mu(\mathrm{d}t)} \int_B \exp(ht)\mu(\mathrm{d}t) \tag{10}$$

for $h \in \mathbb{R}$, then if $\mu([q,1]) > 0$ for all $q < 1$, v_h converges weakly to the Dirac measure δ_1 with mass at the point 1 for $h \to \infty$.

Example 4 (SETAR-model). Consider the affine Markov chain on $\mathbb{R}^k$ with $\xi(x) = x - \varepsilon n(x)$ for $\varepsilon \in \mathbb{R}$ and $n(x) = x/|x|$. Let $\Phi(x) = \sigma I$ for some $\sigma > 0$, and assume that the innovation distribution v satisfies (4). In the one-dimensional case, this Markov chain is a special case of the SETAR-model well known to be ergodic for $\varepsilon > 0$ (Meyn and Tweedie, 1993).

First we try to use the drift function $V(x) = \exp(s|x|)$. By Jensens inequality we find, with notation as in Section 2.1, that

$$\frac{1}{s}\log\delta(s) \geqslant \sigma \int |y| v(\mathrm{d}y).$$

Put $\gamma = \sigma \int |y| v(\mathrm{d}y)$, which is strictly positive if v is not concentrated at 0, then since $1/s \log\delta(s) \to \gamma$ for $s \to 0$ it follows that (5) is fulfilled for some $s > 0$ if and only if

$$\limsup_{|x|\to\infty} |\xi(x)| - |x| = -\varepsilon < -\gamma.$$

Thus, using the standard exponential drift function the best result obtainable is that the SETAR-model has geometric drift towards a compact set if $\varepsilon > \gamma$.

However, using Theorem 2, we can now see that the Markov chain has geometric drift towards a compact set if just $\varepsilon > 0$. The result is easy to generalize to other models like the general SETAR-models.

4. Concluding remarks

Under the assumption of sufficiently light-tailed distributions, we have presented an approach to show geometric drift and thereby geometric ergodicity for the class of multidimensional affine Markov chains. The result obtained is useful and appealing to the intuition, and when applying the result to SETAR-models we avoid any artificial restrictions on the parameters. In the one-dimensional case, the result is easily established using a drift function like $V(x) = \cosh(sx)$, and we have been able to generalized this to the multidimensional case by using a kind of generalized hyperbolic cosine.

Acknowledgements

We thank the anonymous referee for several useful comments.

References

Doukhan, P., 1994. Mixing, Properties and Examples. Springer, Berlin.

Engle, R.F., Kroner, K.F., 1995. Multivariate simultaneous generalized ARCH. Econometric Theory 11, 122–150.

Hansen, E., Rahbek, A., 1998. Stationarity and asymptotics of multivariate ARCH time series with an application to robustness of cointegration analysis. Preprint 12, Department of Statistics and Operations Research, University of Copenhagen.

Meyn, S.P., Tweedie, R.L., 1993. Markov Chains and Stochastic Stability. Springer, Berlin.

Mokkadem, A., 1990. Propriétés de mélange des processus autorégressifs polynomiaux. Ann. Inst. H. Poincaré 26 (4), 219–260.

Tjøstheim, D., 1990. Non-linear time series and Markov chains. Adv. Appl. Probab. 22 (2), 587–611.

Tong, H., 1990. Non-linear Time Series. A Dynamical System Approach. Oxford University Press, Oxford.