

A Note on Subgeometric Rate Convergence for Ergodic Markov Chains in the Wasserstein Metric

Mokaedi V. Lekgari

Mathematics Department, University of Botswana, P/Bag 0022 Gaborone, Botswana

lekgarimv@mopipi.ub.bw

Keywords: Markov Chains; Ergodicity; Wasserstein metric.

Abstract. We investigate subgeometric rate ergodicity for Markov chains in the Wasserstein metric and show that the finiteness of the expectation $E_{(i,j)}[\sum_{k=0}^{\tau_{\Delta}-1} r(k)]$, where τ_{Δ} is the hitting time on the coupling set Δ and r is a subgeometric rate function, is equivalent to a sequence of Foster-Lyapunov drift conditions which imply subgeometric convergence in the Wasserstein distance. We give an example for a 'family of nested drift conditions'.

Introduction and Notations

We start with a brief review of ergodicity. Let $\mathbb{Z}_+ = \{0, 1, 2, \dots\}$, $\mathbb{N}_+ = \{1, 2, \dots\}$, and $\mathbb{R}_+ = [0, \infty)$. Let $(\Phi_n)_{n \in \mathbb{Z}_+}$ denote a Markov chain with transition kernel P on a *countably generated* state space denoted by $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$. $P^n(i, j) = P_i(\Phi_{n=j}) = E_i[\mathbf{1}_{\Phi_{n=j}}]$, where P_i and E_i respectively denote the probability and expectation of the chain under the condition that its initial state $\Phi_0 = i$, and $\mathbf{1}_A$ is the indicator function of set A . According to *Markov's theorem*, a Markov chain $(\Phi_n)_{n \in \mathbb{Z}_+}$ is ergodic if there's positive probability to pass from any state, say $i \in \mathcal{X}$ to any other state, say $\cdot \in \mathcal{X}$ in *one* step. That is, for states $i, \cdot \in \mathcal{X}$ then chain $(\Phi_n)_{n \in \mathbb{Z}_+}$ is ergodic if $P^1(i, \cdot) > 0$.

Also the chain $(\Phi_n)_{n \in \mathbb{Z}_+}$ is said to be (*ordinary*) ergodic if $\forall i, \cdot \in \mathcal{X}$ then

$$P^n(i, \cdot) \rightarrow \pi(\cdot) \text{ as } n \rightarrow \infty,$$

where the σ -finite measure π is the *invariant limit distribution* of the chain.

Chain $(\Phi_n)_{n \in \mathbb{Z}_+}$ is referred to as *geometrically ergodic* if there exists some measurable function $V : \mathcal{X} \rightarrow (0, \infty)$, and constants $\beta < 1$ and $M < \infty$ such that

$$\|P^n(i, \cdot) - \pi(\cdot)\| \leq MV(i)\beta^n, \quad \forall n \in \mathbb{N}_+,$$

where here and hereafter for the (signed) measure μ we define $\mu(f) = \int \mu(dj) f(j)$, and the norm $\|\mu\|$ is defined by $\sup_{|g| \leq f} |\mu(g)|$, whereas the total variation norm is defined similarly but with $f \equiv 1$.

Markov chain $(\Phi_n)_{n \in \mathbb{Z}_+}$ is *strongly ergodic* if

$$\lim_{n \rightarrow \infty} \sup_{i \in \mathcal{X}} \|P^n(i, \cdot) - \pi(\cdot)\| = 0.$$

Loosely speaking subgeometric ergodicity, which we define next, is a kind of convergence that's faster than ordinary ergodicity but slower than geometric ergodicity.

Let function $r \in \Lambda_0$ where Λ_0 is the family of measurable increasing functions $r : \mathbb{R}_+ \rightarrow [1, \infty)$ satisfying $\frac{\log r(t)}{t} \downarrow 0$ as $t \uparrow \infty$. Let Λ denote the class of positive functions $\bar{r} : \mathbb{R}_+ \rightarrow (0, \infty)$ such that for some $r \in \Lambda_0$ we have;

$$0 < \liminf_n \frac{\bar{r}(n)}{r(n)} \leq \limsup_n \frac{\bar{r}(n)}{r(n)} < \infty. \quad (1)$$

Indeed (1) implies the equivalence of the class of functions Λ_0 with the class of functions Λ . Examples of functions in the class $r \in \Lambda$ is the rate $r(n) = \exp(sn^{1/(1+\alpha)})$, $\alpha > 0$, $s > 0$. Without loss to

generality we suppose that $r(0) = 1$ whenever $r \in \Lambda$. The properties of $r \in \Lambda_0$ which follow from (1) and are to be used frequently in this study are;

$$r(x + y) \leq r(x)r(y) \quad \forall x, y \in \mathbb{R}_+ \tag{2}$$

$$\frac{r(x + a)}{r(x)} \rightarrow 1 \quad \text{as } x \rightarrow \infty, \text{ for each } a \in \mathbb{R}_+. \tag{3}$$

Λ is referred to as the class of subgeometric rate functions(cf. [3]).

Let $r \in \Lambda$, then the ergodic chain Φ_n is said to be *subgeometrically ergodic* of order r in the f -norm, (or simply (f, r) -ergodic) if for the unique invariant distribution π of the process and $\forall i \in \mathcal{X}$, then

$$\lim_{n \rightarrow +\infty} r(n) \|P^n(i, \cdot) - \pi(\cdot)\|_f = 0, \tag{4}$$

where $\|\sigma\|_f = \sup_{|g| \leq f} |\sigma(g)|$ and $f : \mathcal{X} \rightarrow [1, \infty)$ is a measurable function. Also for subgeometric ergodic to hold it's necessary that there exist a deterministic sequence $\{V_n\}$ of functions $V_n : \mathcal{X} \rightarrow [1, \infty)$ which satisfy the Foster-Lyapunov drift condition:

$$PV_{n+1} \leq V_n - r(n)f + br(n)\mathbf{1}_C, \quad n \in \mathbb{Z}_+. \tag{5}$$

for a petite set $C \in \mathcal{B}(\mathcal{X})$ and a constant $b \in \mathbb{R}_+$ such that $\sup_C V_0 < \infty$. The Foster-Lyapunov drift conditions provide bounds on the return time to accessible sets thereby availing some control on the Markov process dynamics by focusing on the hitting times on a particular set.

Convergence in the *Wasserstein distance* is a very interesting research area through which [1] amongst other authors suggested a new technique for establishing subgeometric ergodicity. Following [1] we define the Wasserstein distance as follows. Let (\mathcal{X}, d) be a Polish space where d is a distance bounded by 1 and let $\mathcal{P}(\mathcal{X})$ denote the set of all probability measures on state space $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$. Let $\mu, \nu \in \mathcal{P}(\mathcal{X})$; λ is a coupling of μ and ν if λ is a probability on the product space $(\mathcal{X} \times \mathcal{X}, \mathcal{B}(\mathcal{X} \times \mathcal{X}))$, such that $\lambda(A \times \mathcal{X}) = \mu(A)$ and $\lambda(\mathcal{X} \times A) = \nu(A) \quad \forall A \in \mathcal{B}(\mathcal{X})$. We further let $\mathcal{C}(\mu, \nu)$ be set of all probability measures on $(\mathcal{X} \times \mathcal{X}, \mathcal{B}(\mathcal{X} \times \mathcal{X}))$ with marginals μ and ν , and Q be the coupling Markov kernel on $(\mathcal{X} \times \mathcal{X}, \mathcal{B}(\mathcal{X} \times \mathcal{X}))$ such that for every $i, j \in \mathcal{X}$, then $Q((i, j), \cdot)$ is a coupling of $P(i, \cdot)$ and $P(j, \cdot)$. The Wasserstein metric associated with the semimetric d on \mathcal{X} , between two probability measures μ and ν , is then given as

$$W_d(\mu, \nu) := \inf_{\gamma \in \mathcal{C}(\mu, \nu)} \int_{\mathcal{X} \times \mathcal{X}} d(i, j) d\gamma(i, j).$$

When d is the trivial metric $d_0(i, j) = \mathbf{1}_{i \neq j}$, then the associated Wasserstein metric is the total variation metric $W_{d_0}(\mu, \nu) = d_{TV}(\mu, \nu) := 2 \sup_{C \in \mathcal{B}(\mathcal{X})} |\mu(C) - \nu(C)|, \mu, \nu \in \mathcal{P}(\mathcal{X})$.

A set C is said to be small if there exists a constant $\epsilon > 0$ such that for all $i, j \in C$ then $\frac{1}{2}d_{TV}(P(i, \cdot), P(j, \cdot)) \leq 1 - \epsilon$. Set $C \in \mathcal{B}(\mathcal{X})$ is *petite* if there exist some non-trivial measure ν_a on $\mathcal{B}(\mathcal{X})$ and some probability distribution $a = \{a_n : n \in \mathbb{Z}_+\}$ such that

$$\sum_{n=1}^{\infty} a_n P^n(x, \cdot) \geq \nu_a(\cdot), \quad \forall x \in C. \tag{6}$$

Petite sets generalize small sets. The first hitting time on small set C delayed by a constant $\delta > 0$ is given by $\tau_C^\delta = \inf\{n \geq \delta : \Phi_n \in C\}$. We also have $\tau_C^+ = \inf\{n \geq J_1 : \Phi_n \in C\}$ as the first hitting time on the set C after the first jump J_1 of the process. We note that $\xi_C^+ = \xi_C$ if $\Phi_0 \notin C$. In the case when $\delta = 0$ we have $\tau_C^0 = \tau_C$. If C is a singleton consisting only of state i then we write τ_i^δ for τ_C^δ and equivalently τ_i^+ for τ_C^+ . It's worth noting that finite mean return times $E_i[\tau_i^+] < \infty$ guarantee ergodicity or the existence of stationary probability and the convergence $P^n(i, j) - \pi \rightarrow 0$

as $n \rightarrow \infty$. It's also known that subgeometric ergodic is equivalent to (f, r) -regularity. We define (f, r) -regularity as follows. Set $C \in \mathcal{X}$ is said to be (f, r) -regular if for all $i \in C$, a measurable function $f : \mathcal{X} \rightarrow [1, \infty)$, rate function r and $\forall B \in \mathcal{B}^+(\mathcal{X})$ then,

$$\sup_{i \in C} E_i \left[\sum_{k=0}^{\tau_B - 1} r(k) f(\Phi_k) \right] < \infty,$$

where the set $\mathcal{B}^+(\mathcal{X})$ is set of all *accessible* (or Ψ -irreducible) sets. By finding a suitable contracting metric d which may be different from the discrete metric, and a suitable Foster-Lyapunov function V with a ' d -small' sublevel set, [1] suggested a new technique for establishing subgeometric ergodicity. Then [2] extended the results of [1] by establishing sufficient conditions for the existence of the invariant distribution and subgeometric rates of convergence for chains that are not necessarily Ψ -irreducible. For the Polish space (\mathcal{X}, d_*) , the d -small set of [1] was extended by [2] to the (ℓ, ϵ, d) -coupling set (or simply coupling set) $\Delta \in \mathcal{X} \times \mathcal{X}$, where $\ell \in \mathbb{Z}_+$, $\epsilon \in (0, 1)$, and d is a distance on state space \mathcal{X} , topologically equivalent to d_* and bounded by 1.

Let $r \in \Lambda$, then we denote the sequence R as

$$R(n) = \sum_{k=0}^{n-1} r(k) \quad n \in \mathbb{N}_+, \quad R(0) = 1. \quad (7)$$

We show, in this paper through Proposition 1, that the sequence of drift inequalities proposed by [2] hold if and only if $R(\tau_\Delta) < \infty$. As an example, we explore a 'family of nested drift conditions' as proposed by [4] in both the discrete and continuous cases whose results we transfer to the convergence in the Wasserstein metric through Proposition 3.

Main Results

Lyapunov Drift Inequalities

In light of the definitions and notations given above, we state Assumption **A1** as follows:

A1. There exist a coupling set $\Delta \in \mathcal{B}(\mathcal{X} \times \mathcal{X})$ such that for a sequence $r \in \Lambda$ and $\forall i, j \in \mathcal{X}$,

$$\sup_{(i,j) \in \Delta} E_{(i,j)} [R(\tau_\Delta)] < \infty \quad (8)$$

According to Theorem 2.1(ii) of [5], as mentioned already, the Foster-Lyapunov drift conditions in (5) can also be used to define subgeometric rate ergodicity. Following this result [2] proposed a sequence of drift functions according to Assumption **A2** as follows.

A2. There exist

1. a sequence of measurable functions $\{\mathcal{V}_n\}_{n \in \mathbb{Z}_+}$, $\mathcal{V}_n : \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}_+$,
2. a set $\Delta \in \mathcal{B}(\mathcal{X} \times \mathcal{X})$, a constant $b \in \mathbb{R}_+$ and a sequence $r \in \Lambda$ such that $\forall i, j \in \mathcal{X}$ and for every coupling $\alpha \in \mathcal{C}(P(i, \cdot), P(j, \cdot))$;

$$\int_{\mathcal{X} \times \mathcal{X}} \mathcal{V}_{n+1}(z, t) d\alpha(z, t) \leq \mathcal{V}_n - r(n)f + br(n)\mathbf{1}_{(i,j) \in \Delta}, \quad n \in \mathbb{Z}_+. \quad (9)$$

Further, there exist measurable functions $(V_n)_{n \in \mathbb{Z}_+}$ such that $\forall i, j \in \mathcal{X}$ and any $n \in \mathbb{Z}_+$:

$$\mathcal{V}_n(i, j) \leq V_n(i) + V_n(j) \text{ and } PV_{n+1} \leq V_n + br(n). \quad (10)$$

and $\forall k \in \mathbb{Z}_+$

$$\sup_{(i,j) \in \Delta} \{P^k V_0(i) + P^k V_0(j)\} < +\infty \text{ and } \forall i \in \mathcal{X}, P^k V_0(i) < +\infty. \quad (11)$$

Proposition 1. $A1 \Leftrightarrow A2$.

Proof. 1. **A1** \Rightarrow **A2**

Let $r \in \Lambda_0$ and $\{\mathcal{V}_n\}, \{\mathcal{W}_n\} : \mathcal{X} \times \mathcal{X} \rightarrow [0, \infty)$ be sequences of functions defined for all $n \in \mathbb{Z}_+$ by

$$\mathcal{V}_n = E_{i,j} \left[\sum_{k=1}^{\tau_\Delta} r(n+k) \right] \mathbf{1}_{\Delta^c} + r(n),$$

$$\mathcal{W}_n = E_{i,j} \left[\sum_{k=1}^{\tau_\Delta} r(n+k) \right].$$

Then by the submultiplicative property (2) and Assumption **A1** we have $\mathcal{V}_0 \leq \mathcal{V}_n \leq r(n)\mathcal{V}_0 < \infty$ and $\mathcal{W}_0 \leq \mathcal{W}_n \leq r(n)\mathcal{W}_0 < \infty$. We also note that $\mathcal{V}_n = \mathcal{W}_n \mathbf{1}_{\Delta^c} + r(n)$ such that

$$\begin{aligned} \int_{\mathcal{X} \times \mathcal{X}} \mathcal{V}_{n+1}(z, t) d\alpha(z, t) &= \int_{\mathcal{X} \times \mathcal{X}} (\mathcal{W}_{n+1} \mathbf{1}_{\Delta^c}(z, t) + r(n+1)) d\alpha(z, t) \\ &\leq \int_{\mathcal{X} \times \mathcal{X}} (\mathcal{W}_{n+1} \mathbf{1}_{\Delta^c}(z, t) + \mathcal{W}_{n+1} \mathbf{1}_\Delta(z, t)) d\alpha(z, t) \\ &= \int_{\mathcal{X} \times \mathcal{X}} (\mathcal{W}_{n+1}(z, t)) d\alpha(z, t) \\ &\leq \mathcal{W}_n(i, j) \\ &= \mathcal{V}_n(i, j) - r(n) + \mathcal{W}_n \mathbf{1}_{\{(i,j) \in \Delta\}} \\ &\leq \mathcal{V}_n(i, j) - r(n) + r(n) \mathcal{W}_n \mathbf{1}_{\{(i,j) \in \Delta\}} \\ &\leq \mathcal{V}_n(i, j) - r(n) + r(n) \mathcal{W}_0 \mathbf{1}_{\{(i,j) \in \Delta\}} \\ &\leq \mathcal{V}_n(i, j) - r(n) + br(n) \mathbf{1}_\Delta(i, j), \end{aligned} \tag{12}$$

where we choose $b = \sup_\Delta \mathcal{W}_0$.

2. **A2** \Rightarrow **A1**

Analogous to Proposition 11.3.3 in [7] we get from **A2** that for some constant $c < \infty$

$$E_{i,j}[R(\tau_\Delta)] \leq \begin{cases} \mathcal{V}_0(i, j) & , (i, j) \in \Delta^c, \\ r(0) + cQ\mathcal{V}_0(i, j) & , (i, j) \in \mathcal{X} \times \mathcal{X} \end{cases}$$

Then by Eq. 8, Eq. 9 and $\sup_\Delta \mathcal{V}_0 < \infty$ we get $E_{i,j}[R(\tau_\Delta)] < \infty$. ■

Family of nested drift conditions

The phenomenon of ergodicity as given in Proposition 1 is not altogether new as is clear from the following Proposition which deals with a family of nested drift conditions for subgeometrically ergodic general state space Markov processes analogous to one proposed by [4].

Proposition 2. *Suppose that there are functions $\mathcal{V}_k, \mathcal{W}_k : \mathcal{X} \times \mathcal{X} \rightarrow [1, \infty)$, where $k \in \mathbb{Z}_+$, a coupling set $\Delta \in \mathcal{B}(\mathcal{X} \times \mathcal{X})$ such that for any initial state $(i, j) \in \mathcal{X} \times \mathcal{X}$ of the chain, we have*

$$E_{i,j}[\mathcal{V}_{k+n+1}(i+1, j+1) | \mathcal{F}_n] \leq \mathcal{V}_k(i, j) - E_{i,j}[R(\tau_\Delta)] + \mathcal{W}_k(i, j) \mathbf{1}_{(i,j) \in \Delta} \tag{13}$$

then the chain Φ_n is subgeometrically ergodic.

Proposition 2 follows from Proposition 3 which is the extension of Proposition 3.1 in [4] on family of nested drift conditions. As is evident in the Propositions that follow we note that the results of Proposition 2 stay the same if we replace $R(\tau_\Delta)$ with $R_m(\tau_\Delta)$, where $R_m(n) = \sum_{k=0}^{n-1} r(m+k)$ for $m \geq 0$, $n \geq 1$ with $R(0) = 1$. The results also stay the same if we replace the measurable functions \mathcal{V}_k and \mathcal{W}_k with V_k and W_k respectively as is the case for convergence in the f -norm.

Proposition 3. *Let the chain $(\Phi_n)_{n \in \mathbb{Z}_+}$ be irreducible and aperiodic. Further suppose that there are functions $f, V_k, W_k : \mathcal{X} \rightarrow [1, \infty)$, where $k \in \mathbb{Z}_+$, with $\sup_C V_k < \infty$, $\sup_C W_k < \infty$ and a small set C such that for a non-decreasing sequence of stopping times $\{\mathcal{T}_n : n \in \mathbb{Z}_+\}$ and any $\Phi_{\mathcal{T}_n} \in \mathcal{X}$, we have*

$$E_{\Phi_{\mathcal{T}_n}} [V_{k+\mathcal{T}_{n+1}}(\Phi_{\mathcal{T}_{n+1}}) | \mathcal{F}_n] \leq V_k(\Phi_{\mathcal{T}_n}) - E_{\Phi_{\mathcal{T}_n}} \left[\sum_{l=\mathcal{T}_n}^{\mathcal{T}_{n+1}-1} f(\Phi_l) r(k+l) \right] + W_k(\Phi_{\mathcal{T}_n}) \mathbf{1}_{\{\Phi_{\mathcal{T}_n} \in C\}} \quad (14)$$

then the chain Φ_n is (f, r) -ergodic.

Proof. We let \mathcal{T}_n be some random stopping time with \mathcal{F}_n as the σ -algebra of events generated by \mathcal{T}_n . Then by Dynkin's inequality we get

$$E_{\Phi_{\mathcal{T}_n}} \left[\sum_{l=1}^{\mathcal{T}_n-1} f(\Phi_l) r(k+l) \right] \leq V_k(\Phi_{\mathcal{T}_n}) + W_k(\Phi_{\mathcal{T}_n}) \mathbf{1}_{\{\Phi_{\mathcal{T}_n} \in C\}} \quad (15)$$

which implies that

$$\begin{aligned} E \left[\sum_{n=0}^{\tau_C-1} r(n) f(n) | \mathcal{F}_n \right] &\leq V_0(i) + W_0(i) E \left[\sum_{n=0}^{\tau_C-1} \mathbf{1}_{\{i \in C\}} \right] \\ &\leq c(V_0(i) + W_0(i)), \quad c \in [0, \infty) \\ &< \infty. \end{aligned} \quad (16)$$

by assuming that $\sup_C V_k < \infty$, $\sup_C W_k < \infty$. Hence the chain is (f, r) -ergodic. \blacksquare

Proposition 4. *Let the chain $(\Phi_t)_{t \in \mathbb{R}_+}$ be irreducible. Further suppose that there are functions $f, V_k, W_k : \mathcal{X} \rightarrow [1, \infty)$ where $k \in \mathbb{Z}_+$, some constant $\varepsilon > 0$, a small set C such that for a non-decreasing sequence of stopping times $\{\mathcal{T}_n : n \in \mathbb{Z}_+\}$ and any $\Phi_{\mathcal{T}_n} \in \mathcal{X}$, we have*

$$E_{\Phi_{\mathcal{T}_n}} [V_{k+\mathcal{T}_{n+1}}(\Phi_{\mathcal{T}_{n+1}}) | \mathcal{F}_n] \leq V_k(\Phi_{\mathcal{T}_n}) - E_{\Phi_{\mathcal{T}_n}} \left[\int_0^\varepsilon f(\Phi_s) r(k+s) ds \right] + W_k(\Phi_{\mathcal{T}_n}) \mathbf{1}_{\{\Phi_{\mathcal{T}_n} \in C\}} \quad (17)$$

then the chain Φ_t is (f, r) -ergodic.

Proof. Note that this Proposition is the continuous counterpart to Proposition 3. For the term

$$E_{\Phi_{\mathcal{T}_n}} \left[\int_0^\varepsilon f(\Phi_s) r(k+s) ds \right],$$

where $\varepsilon > 0$, by the submultiplicative property (2) we have

$$E_{\Phi_{\mathcal{T}_n}} \left[\int_0^\varepsilon f(\Phi_s) r(k+s) ds \right] \leq r(k) E_{\Phi_{\mathcal{T}_n}} \left[\int_0^\varepsilon f(\Phi_s) r(s) ds \right] < \infty \quad (18)$$

because $r \in \Lambda$ is finite for all $k \in \mathbb{Z}_+$ and $E_{\Phi_{\mathcal{T}_n}} \left[\int_0^\varepsilon f(\Phi_s) r(s) ds \right] < \infty$ by proof of Theorem 6 in [6], hence we conclude that the chain Φ_t is (f, r) -ergodic. \blacksquare

References

- [1] O. Butkovsky, Subgeometric rates of convergence of Markov processes in the Wasserstein metric, *Ann. Appl. Probab.* 24(2) (2014) 526–552.
- [2] A. Durmus, G. Fort, E. Moulines, New conditions for subgeometric rates of convergence in the Wasserstein distance for Markov chains, Unpublished paper, 2014. Available on: <https://hal.archives-ouvertes.fr/hal-00948661v1/document>.
- [3] E. Nummelin, P. Tuominen, The rate of convergence in Orey’s theorem for Harris recurrent Markov chains with applications to renewal theory, *Stochastic Process Appl.* 15 (1983) 295–311.
- [4] S.B. Connor, G. Fort, State-dependent Foster-Lyapunov criteria for subgeometric convergence of Markov chains, *Stochastic Processes and their Applications.* 119 (2009) 4176–4193.
- [5] S.P. Meyn, R.L. Tweedie, State-dependent criteria for convergence of Markov chains, *Ann. Appl. Probab.* (1994) 149–168.
- [6] M.V. Lekkari, Subgeometric Ergodicity Analysis of Continuous-time Markov Chains Under Random-time State-dependent Lyapunov Drift Conditions, *J. Prob. Stat.* 2014 (2014), Article ID 274535.
- [7] S.P. Meyn, R.L. Tweedie, *Markov Chains and Stochastic Stability*, Springer, 1993.
- [8] R.L. Tweedie, Criteria for rates of convergence of Markov chains, in J.F.C. Kingman , G.E.H. Reuter(Eds.), *Probability Statistics and Analysis*, in: London Mathematical Society Lecture Note Series, Cambridge University Press, 1983, pp. 227–250.