VALUE FUNCTION ITERATION USING MONTE CARLO

JENÖ PÁL[a] AND JOHN STACHURSKI[b]

Abstract to be written.

KEYWORDS: Value iteration, Monte Carlo, empirical process theory.

1. INTRODUCTION

A large number of economic decision problems are naturally expressed as stochastic dynamic programs (SDPs).[1] These problems are intrinsically high dimensional, and quantitative researchers have devoted a vast amount of computer time to solving them numerically over the past few decades. While computers are now faster and algorithms have improved, the need for richer and more detailed models means that computational constraints remain the bottleneck in many applications.[2]

SDPs are solved by trading off current rewards with future rewards. Future rewards are represented by a *value function* V, which returns the maximal expected discounted reward obtainable from each state. In the stationary, infinite horizon case that is the focus of this paper, the standard algorithm for calculating V is value function iteration (VFI), which involves iterating with an operator T (the *Bellman operator*) on some initial guess v_0, producing a sequence of functions $v_1 := Tv_0$, $v_2 := Tv_1 = T^2 v_0$, etc. The VFI algorithm is typically robust and globally convergent, in the sense that $v_k \to V$ as $k \to \infty$ for a large class of initial v_0.

For *stochastic* dynamic programs, VFI involves computing *expected* future values of different states and actions. In the general SDP treated here, these expectations take the form $\int w[F(x,a,u)]\phi(du)$, where $w = v_k = T^k v_0$ for some k, x

[a]Jenö's address

[b]Institute of Economic Research, Kyoto University, Sakyo-ku, Kyoto, Japan

[1]References?

[2]Numerical solution of SDPs is particularly challenging for in multidimensional settings (Bellman, 1957). The difficulty is exacerbated when the SDP must be solved many times at different parameterizations, as is the case with, for example, Bewley models, or estimated models where unknown parameters are contained in the primitives of the dynamic programming problem (citations?).

is the state variable, a is the current action or control taken by the agent, u is a possible realization of the shock, $F(x, a, u)$ is the next period state associated with x, a and u, and ϕ is the distribution of the shock process. For large SDPs, the integral is evaluated millions of times for each iteration of T.

Except for the discrete case, or for special functional forms, this integral cannot be evaluated exactly. Instead, a numerical integration procedure must be chosen. A common practice is to apply a discretization procedure to ϕ, replacing the integral with a sum (cite tauchen?). Another is to solve the integral using a more sophisticated numerical routine such as Gaussian quadrature, typically involving approximation of the function $u \mapsto w[F(x, a, u)]$ with polynomials (cite?). A third is Monte Carlo, which requires sampling draws $U_1, \ldots, U_n$ from ϕ using a random number generator, and computing $n^{-1} \sum_{i=1}^{n} w[F(x, a, U_i)]$. Convergence of this expression to the integral follows from the law of large numbers.[3]

The appropriateness of these different methods depends on the problem at hand. Discretization is suitable when the dimension of the integral is small, and the solution to the problem is not overly sensitive on the specification of the shock. Gaussian quadrature is efficient for integrating smooth functions in low dimensions. Monte Carlo is used primarily when u is multi-dimensional, since, unlike the other routines, the number of function calls does not increase exponentially with the dimension of the integral.

In this paper we consider VFI when the integral is evaluated using Monte Carlo. In effect, this involves replacing the true Bellman operator T with an approximation R, where the former evaluates integrals exactly—but cannot be implemented—while the latter approximates integrals using Monte Carlo. Employing the techniques of empirical process theory, we show that iterating with R generates a sequence of functions that converges, and that the limit point is close to V when the sample size is large. We also prove a version of the same result when the iterates of the Bellman operator must be approximated, as is necessary in continuous state spaces. In addition, we provide results on

[3]Monte Carlo may be necessary even when the ϕ is finite (i.e., supported on a finite set). In this case, the integral is a sum and can, at least in principle, be calcuated exactly. However, if the support is large (probably due to multiple dimensions), then exact computation of the sum may be infeasible or prohibitively slow. In that case, Monte Carlo can be used to approximate. In effect, Monte Carlo ignores those points in the support with low probability.

asmptotic rates of convergence.

Using Monte Carlo to solve SDPs is a common practice in the economic and financial literature. For example, Keane and Wolpin (1994) proposed and analyzed a method for solving and estimating discrete choice SDPs using Monte Carlo. Pakes and McGuire (2001) developed a stochastic algorithm for computing Markov perfect equilibria in models of industry dynamics. Longstaff and Schwartz (2001) introduced a popular stochastic algorithm for valuing American options. Den Haan and Marcet (1990) used simulation-based parameterized expectations to solve for decision rules in economic models with rational expectations.

These studies do not address the convergence problem treated in this paper. One study that does provide related convergence results is Rust's (1997) famous analysis of randomization for solving SDPs. Rust derives a number of major results on convergence and computational complexity. This paper extends and complements Rust's analysis, developing a simple and practical random Bellman operator suitable for a broad range of applications. Unlike Rust, we do not require that the choice set for the agent is finite, or that transition probabilities can be represented by a density for each state-action pair. Our state space is more general than the hypercube in $\mathbb{R}^n$. We also treat the function approximation problem inherent in continuous state SDPs, developing an algorithm that can be exactly implemented—modulo floating point arithmetic. (On the other hand, we do not add to the theoretical analysis of computational complexity that was Rust's main objective.)

Section 2 introduces background concepts and notation. Section 3 defines the model and outlines VFI. Section 4 introduces random VFI, and provides an initial convergence result, while section 5 does the same in the case where function approximation is present. Section 6 discusses rates of convergence. Section 7 concludes. Remaining proofs can be found in section 8.

2. PRELIMINARIES

We begin by introducing some notation. For topological space $\mathbb{T}$, the symbol $\mathscr{C}(\mathbb{T})$ denotes the collection of continuous, bounded, real-valued functions on $\mathbb{T}$, while $\|\cdot\|$ is the supremum norm on $\mathscr{C}(\mathbb{T})$. Operator $S\colon \mathscr{C}(\mathbb{T}) \to \mathscr{C}(\mathbb{T})$ is

called a *contraction of modulus* ρ if $\rho < 1$ and

$$\|Sv - Sw\| \leq \rho \|v - w\| \text{ for all pairs } v, w \in \mathscr{C}(\mathbb{T}) \tag{1}$$

S is called *nonexpansive* if (1) holds when $\rho = 1$. By Banach's contraction mapping theorem, every contraction S of modulus ρ on $\mathscr{C}(\mathbb{T})$ has a unique fixed point $W \in \mathscr{C}(\mathbb{T})$, and, moreover, $\|S^n w - W\| = O(\rho^n)$ for each $w \in \mathscr{C}(\mathbb{T})$. We will make use of the following elementary lemma.

LEMMA 2.1 *Let S and S' be operators from $\mathscr{C}(\mathbb{T})$ to itself.*

1. *If S is nonexpansive and S' is a contraction of modulus ρ, then the composition $S \circ S'$ is a contraction of modulus ρ.*
2. *If S and S' are both contractions of modulus ρ with fixed points W and W' respectively, then $\|W - W'\| \leq (1 - \rho)^{-1} \|SW' - W'\|$.*

Part 1 is trivial. For a proof of part 2, see, for example, Rust (1997, lemma 2.1).

In what follows, all random variables are defined on a common probability space $(\Omega, \mathscr{F}, \mathbf{P})$. If Y is a real-valued random variable on this space, then $\mathbf{E}Y$ denotes the expectation $\int Y(\omega)\mathbf{P}(d\omega)$ of Y. As usual, given real valued random variables $(Y_n)_{n\in\mathbb{N}}$ and positive real sequence $(\alpha_n)_{n\in\mathbb{N}}$, we write $Y_n = O_{\mathbf{P}}(\alpha_n^{-1})$ if, for any $\epsilon > 0$, there exists an $M \in \mathbb{N}$ such that $\mathbf{P}\{\alpha_n |Y_n| > M\} < \epsilon$ for all $n \in \mathbb{N}$.

When applying the theory of empirical processes later in the paper, we will need the concept of outer expectation. Let X be a map from Ω into $\mathbb{R}$ that is not necessarily measurable. For such an X, the *outer expectation* of X is denoted by $\mathbf{E}^* X$, and defined by $\inf_Y \mathbf{E}Y$, where the infimum is over all random variables (i.e., $\mathscr{F}$-measurable maps) Y such that $X \leq Y$ and $\mathbf{E}Y$ exists. Note that if X is measurable, then $\mathbf{E}^* X = \mathbf{E}X$.

For a sequence of possibly nonmeasurable maps (U_n) from Ω into a metric space $(\mathbb{T}, d)$ and a $\mathbb{T}$-valued random variable U, we say that $U_n \to U$ holds $\mathbf{P}^*$-almost surely if there exists a measurable real-valued sequence Δ_n with $d(U_n, U) \leq \Delta_n$ and $\mathbf{P}\{\Delta_n \to 0\} = 1$. We say that U_n converges in distribution to U if $\mathbf{E}^* g(U_n) \to \mathbf{E}g(U)$ for every $g \in \mathscr{C}(\mathbb{T})$. For the former convergence we write $U_n \overset{a.s.^*}{\to} U$, while for the latter we write $U_n \overset{d^*}{\to} U$.

The continuous mapping theorem continues to hold in this setting:

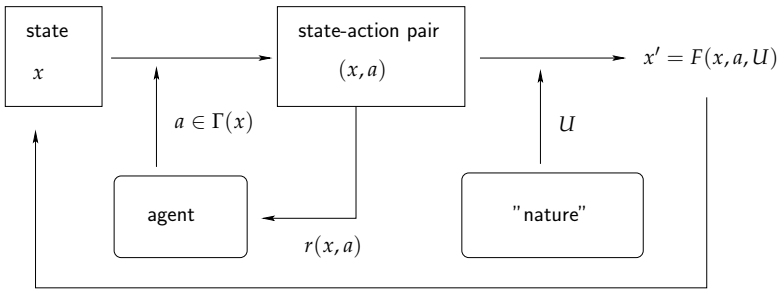

FIGURE 1.— The programming problem

LEMMA 2.2 *If $\mathbb{T}'$ is another metric space and $g\colon \mathbb{T} \to \mathbb{T}'$ is continuous, then*

$$U_n \xrightarrow{d^*} U \implies g(U_n) \xrightarrow{d^*} g(U)$$

Let $(X_n)_{n\geq 1}$ be a sequence of (not necessarily measurable) maps from Ω into $\mathbb{R}$. We write $X_n = O_{\mathbf{P}^*}(n^{-1/2})$ if there exists a sequence of real-valued random variables $(\Delta_n)_{n\geq 1}$ such that $X_n \leq \Delta_n$ for all n and $\Delta_n = O_{\mathbf{P}}(n^{-1/2})$.

3. SET UP

In this section we introduce a general stochastic dynamic programming problem and describe the value function iteration algorithm.

3.1. The Model

We consider a stochastic dynamic programming problem of the following form. A controller observes the state x of a given system, and responds with an action a from a feasible set $\Gamma(x)$ determined by the state. Given this state-action pair (x, a), the controller receives current reward $r(x, a)$, and the new state is determined as $x' = F(x, a, U)$, where U is a draw from a fixed distribution ϕ. The process now repeats. The controller's objective is to maximize expected discounted rewards (for some discount factor ρ). Figure 1 illustrates.

More formally, let $\mathbb{X}$ and $\mathbb{A}$ be metric spaces representing the state and action spaces respectively, and let Γ be a correspondence from $\mathbb{X}$ to $\mathbb{A}$. Let $\mathbb{G}$ be the

graph of Γ:

$$\mathbb{G} := \{(x, a) \in \mathbb{X} \times \mathbb{A} : a \in \Gamma(x)\}$$

so that $\mathbb{G}$ is the set of *feasible state-action pairs*. Let $(U_t)_{t \geq 1}$ be an IID sequence of random variables taking values in metric space $\mathbb{U}$, and having common distribution ϕ.[4] Let F be a map from $\mathbb{G} \times \mathbb{U}$ into $\mathbb{X}$.

We assume throughout that $\mathbb{X}$ and $\mathbb{A}$ are compact, and that Γ is continuous and compact-valued. We assume further that the reward function $r \colon \mathbb{G} \to \mathbb{R}$ is everywhere continuous, as is

$$\mathbb{G} \ni (x, a) \mapsto F(x, a, u) \in \mathbb{X}$$

for all $u \in \mathbb{U}$. For fixed $(x, a) \in \mathbb{G}$, the map $u \mapsto F(x, a, u)$ is required to be measurable.[5]

The model we have described is a standard SDP. Macroeconomists have traditionally used a more specialized formulation, with correlated shocks. For example, consider the reduced form macroeconomic model considered in Santos and Vigo-Aguiar (1998). The state space is a product space $K \times Z \subset \mathbb{R}^m \times \mathbb{R}^n$, where $k \in K$ is a vector of endogenous variables and $z \in Z$ is a vector of exogenous variables. Technology is summarized by a feasible set $\Theta \subset K \times Z \times K$. The exogenous process $(z_t)_{t \geq 0}$ evolves according to $z_{t+1} = g(z_t, \epsilon_{t+1})$, where $(\epsilon_t)_{t \geq 1}$ is IID. Instanteneous rewards are given by $v(k, z, k')$.

This formulation is a special case of our SDP. To see this, for the state take $x = (k, z) \in K \times Z$, and for the action take $a = k' \in K$. The feasible correspondence is $\Gamma(x) = \Gamma(k, z) = \{k' \in K : (k, z, k') \in \Theta\}$. The shock is $u = \epsilon$, and the transition function is $F(x, a, u) = F(k, z, k', \epsilon) = (k', g(z, \epsilon)) \in K \times Z$. The reward function is $r(x, a) = r(k, z, k') = v(k, z, k')$.

3.2. *Value Function Iteration*

A *feasible policy* is a Borel measurable map $\sigma \colon \mathbb{X} \to \mathbb{A}$ such that $\sigma(x) \in \Gamma(x)$ for all $x \in \mathbb{X}$. Let Σ be the set of all such policies. Each feasible policy $\sigma \in \Sigma$ and

[4]The distribution ϕ is a Borel probability measure on $\mathbb{U}$ satisfying $\mathbf{P}\{U_t \in B\} = \phi(B)$ for all $t \geq 0$ and all Borel sets $B \subset \mathbb{U}$.

[5]When stated without qualification, measurability refers to Borel measurability.

$x \in \mathbb{X}$ defines a Markov process on $\mathbb{X}$ given by

$$X_{t+1} = F(X_t, \sigma(X_t), U_{t+1}) \quad \text{with} \quad X_0 = x \tag{2}$$

From this process, we let $V_\sigma \colon \mathbb{X} \to \mathbb{R}$ be defined by

$$V_\sigma(x) = \mathbf{E}\left\{\sum_{t=0}^{\infty} \rho^t r(X_t, \sigma(X_t))\right\} = \sum_{t=0}^{\infty} \rho^t \, \mathbf{E}\, r(X_t, \sigma(X_t)) \qquad (x \in \mathbb{X})$$

Let $T \colon \mathscr{C}(\mathbb{X}) \to \mathscr{C}(\mathbb{X})$ be the Bellman operator, defined by

$$Tv(x) := \max_{a \in \Gamma(x)} \left\{ r(x,a) + \rho \int v[F(x,a,u)]\phi(du) \right\} \qquad (x \in \mathbb{X}) \tag{3}$$

For $v \in \mathscr{C}(\mathbb{X})$, a policy $\sigma \in \Sigma$ is called *v-greedy* if, for all $x \in \mathbb{X}$,

$$\sigma(x) \in \operatorname{argmax}_{a \in \Gamma(x)} \left\{ r(x,a) + \rho \int v[F(x,a,u)]\phi(du) \right\}$$

The *value function* V is defined pointwise on $\mathbb{X}$ by $V(x) = \sup_{\sigma \in \Sigma} V_\sigma(x)$. A policy $\sigma \in \Sigma$ is called *optimal* if $V_\sigma = V$. The following results are standard.

THEOREM 3.1 *Under our assumptions,*

1. *T is a contraction of modulus ρ on $\mathscr{C}(\mathbb{X})$, and V is the unique fixed point;*
2. *a policy $\sigma \in \Sigma$ is optimal if and only if it is V-greedy; and*
3. *at least one such policy exists.*

From part 2 of the theorem, if V is known then an optimal policy can be calculated in a relatively straightforward way. In computing V, the most common technique is *value function iteration* (VFI). The procedure is as follows:

Algorithm 1: VFI

1 fix $v \in \mathscr{C}(\mathbb{X})$;
2 compute $T^k v$ iteratively, where T^k is the k-th iterate of T ;
3 compute a $T^k v$-greedy policy σ

In view of theorem 3.1, we have $\|T^k v - V\| = O(\rho^k)$. Using this fact and optimality of V-greedy policies, one can show that the $T^k v$-greedy policy σ produced by the algorithm is approximately optimal when k is sufficiently large (see, e.g., Puterman, 1994, theorem 6.3.1, or Stachurski, 2009, theorem 10.2.1). The appropriate k is usually chosen according to some stopping criterion that depends on the deviation between successive iterates of the Bellman operator.

4. RANDOM VFI

As discussed in the introduction, we consider the effect of replacing the exact integral in (3) with a Monte Carlo approximation. The most straightforward implementation of this idea is to generate a sample

$$U_1, \ldots, U_n \overset{\text{IID}}{\sim} \phi \tag{4}$$

using Monte Carlo (i.e., a random number generator), and then iterate with the random Bellman operator R_n defined by

$$R_n v(x) := \max_{a \in \Gamma(x)} \left\{ r(x,a) + \rho \frac{1}{n} \sum_{i=1}^{n} v[F(x,a,U_i)] \right\} \quad (x \in \mathbb{X}) \tag{5}$$

Given its dependence on the sample (4), the operator R_n is clearly random. A realization of $\omega \in \Omega$ determines a particular realization $(U_i(\omega))_{i=1}^n$ of the sample (4), which in turn defines a realization $R_n(\omega)$ of R_n. Each realization $R_n(\omega)$ is an operator from $\mathscr{C}(\mathbb{X})$ to itself.

The complete procedure for random value function iteration is as follows:

Algorithm 2: Random VFI

1 generate the sample $(U_1, \ldots, U_n) \overset{\text{IID}}{\sim} \phi$ in (4) ;
2 fix $v \in \mathscr{C}(\mathbb{X})$;
3 compute $R_n^k v$ iteratively, where R_n^k is the k-th iterate of R_n ;
4 compute a $R_n^k v$-greedy policy σ ;

The error introduced by using random VFI is due to the deviation between $\lim_{k\to\infty} R_n^k v$ and $V = \lim_{k\to\infty} T^k v$. Without further analysis, one cannot rule out the possibility that this deviation is large. Even if R_n and T are similar as maps—as they should be for large n—the error $\|R_n^k v - T^k v\|$ is compounded at each iteration. Moreover, $\lim_{k\to\infty} R_n^k v$ may fail to exist, or only exist for some v or some realizations of uncertainly.

This turns out not to be the case. To show this, we begin with the following result (proved in the appendix):

LEMMA 4.1 *The operator $R_n(\omega)$ is a contraction of modulus ρ on $\mathscr{C}(\mathbb{X})$ for all $n \in \mathbb{N}$ and all $\omega \in \Omega$.*

As a consequence of lemma 4.1, for any $\omega \in \Omega$, there exists a unique $V_n(\omega) \in \mathscr{C}(\mathbb{X})$ such that $R_n(\omega)V_n(\omega) = V_n(\omega)$. We refer to V_n as a random function, although $\omega \mapsto V_n(\omega)$ may not be Borel measurable as mapping from Ω to $\mathscr{C}(\mathbb{X})$. Our first aim is to show that $\|V_n - V\| \to 0$ with probability one as $n \to \infty$.[6] Since Borel measurability of $\omega \mapsto V_n(\omega)$ is problematic, we use the concept of $\mathbf{P}^*$-almost sure convergence.

THEOREM 4.1 $\|V_n - V\| \overset{a.s.^*}{\to} 0$ *as* $n \to \infty$.

PROOF: By lemma 4.1, the operator R_n is a contraction of modulus ρ for all n (with probability one). Applying lemma 2.1, we get

$$\|V_n - V\| \leq \frac{1}{1-\rho}\|R_n V - V\| \qquad (n \in \mathbb{N}) \tag{6}$$

Hence, to prove theorem 4.1, it is sufficient to prove that $\|R_n V - V\|$ converges to zero with probability one. To bound $\|R_n V - V\|$, we make use of the following: If $g, g' \in \mathscr{C}(\mathbb{Y})$ for some compact metric space $\mathbb{Y}$, then

$$|\max g - \max g'| \leq \max |g - g'| =: \|g - g'\| \tag{7}$$

So let us consider the deviation $\|R_n V - V\| = \|R_n V - TV\|$. Using (7), we obtain

$$|R_n V(x) - TV(x)| \leq \rho \max_{a \in \Gamma(x)} \left| \frac{1}{n}\sum_{i=1}^{n} V[F(x,a,U_i)] - \int V[F(x,a,u)]\phi(du) \right| \tag{8}$$

where $x \in \mathbb{X}$ is arbitrary. To simplify notation, let $y = (x,a)$ denote a typical element of $\mathbb{G}$, and define

$$h_y(u) :=: h_{(x,a)}(u) :=: V[F(x,a,u)] :=: V[F(y,u)] \tag{9}$$

Taking the supremum of (8) over $x \in \mathbb{X}$, we now have

$$\begin{aligned} \|R_n V - TV\| &\leq \rho \max_{(x,a) \in \mathbb{G}} \left| \frac{1}{n}\sum_{i=1}^{n} V[F(x,a,U_i)] - \int V[F(x,a,u)]\phi(du) \right| \\ &= \rho \max_{y \in \mathbb{G}} \left| \frac{1}{n}\sum_{i=1}^{n} h_y(U_i) - \int h_y(u)\phi(du) \right| \end{aligned} \tag{10}$$

[6] The relative optimality of the $R_n^k v$-greedy policy σ computed by the random VFI algorithm depends on the deviation between $R_n^k v$ and V. Using the triangle inequality, we can bound the latter by $\|R_n^k v - V_n\| + \|V_n - V\|$. By lemma 4.1, the first term is $O(\rho^k)$ in k. Convergence of V_n to V is less clear. Although the operators R_n and T should be "similar" for large n, their differences may be compounded through iteration, and the fixed points V_n and V are the *limits* of this iterative process. Hence the focus on the deviation $\|V_n - V\|$.

To bound (10), we now require some standard definitions from empirical process theory.[7] To this end, Let $\mathscr{H}$ be a family of bounded measurable functions mapping $\mathbb{U}$ into $\mathbb{R}$. For $h \in \mathscr{H}$, we use the notation

$$\phi_n(h) := \frac{1}{n}\sum_{i=1}^{n} h(U_i) \quad \text{and} \quad \phi(h) := \int h d\phi \tag{11}$$

By the scalar law of large numbers we have $\phi_n(h) \to \phi(h)$ with probability one for every $h \in \mathscr{H}$. Our interest, however, is in convergence of the term

$$\sup_{h\in\mathscr{H}} |\phi_n(h) - \phi(h)| \tag{12}$$

The class of functions $\mathscr{H}$ is called *ϕ-Glivenko-Cantelli* if (12) converges to zero $\mathbf{P}^*$-almost surely as $n \to \infty$.[8] In our case, we are interested in convergence of

$$\sup_{y\in\mathbb{G}} \left| \frac{1}{n}\sum_{i=1}^{n} h_y(U_i) - \int h_y(u)\phi(du) \right| = \sup_{y\in\mathbb{G}} |\phi_n(h_y) - \phi(h_y)|$$

in (10), so the relevant class of functions is $\{h_y\}_{y\in\mathbb{G}}$. Since $\mathbb{G}$ is a compact metric space and $\mathbb{G} \ni y \mapsto h_y(u) \in \mathbb{R}$ is continuous for every $u \in \mathbb{U}$, this class is known to be ϕ-Glivenko-Cantelli whenever there exists a measurable function $H\colon \mathbb{U} \to \mathbb{R}$ such that $\int H d\phi < \infty$ and $|h_y(u)| \leq H(u)$ for every $y \in \mathbb{G}$.[9] Since $0 \leq h_y(u) \leq \|V\|$ for every $u \in \mathbb{U}$, this envelope condition is satisfied, and $\{h_y\}_{y\in\mathbb{G}}$ is ϕ-Glivenko-Cantelli. This concludes the proof. *Q.E.D.*

5. RANDOM FITTED VFI

When implementing VFI in continuous state and action spaces, one issue that must be addressed is approximation of the iterates: Except in special cases, the iterates $v, Tv, \ldots, T^k v$ cannot be stored on a computer. The same is true for the iterates of the random Bellman operator R_n. Indeed, a function such as Rv cannot even be evaluated, since evaluation requires computation of a maximum at an (uncountable) infinity of points $x \in \mathbb{X}$. Thus, the iterates must

[7] If $\mathbb{G}$ is a singleton, then (10) converges to zero with probability one by the scalar strong law of large numbers. In fact, if $\mathbb{G}$ is finite, this same scalar law of large numbers can be used to show probability one convergence to zero for the term in (10). However, in our case, the set $\mathbb{G}$ is, in general, uncountably infinite, and hence the need for empirical process theory.

[8] Intuitively, the ϕ-Glivenko-Cantelli property holds if $\mathscr{H}$ is not too "diverse".

[9] See, for example, van der Vaart, 1998, p. 272.

be approximated. We call this process *random fitted VFI*. In this section, our aim is to develop convergence results for random fitted VFI analogous to those obtained for random VFI in section 4.

Our random fitted VFI algorithm is the same as algorithm 2, but replacing R_n with $\hat{R}_n := L \circ R_n =: LR_n$. Here $L\colon \mathscr{C}(\mathbb{X}) \to \mathscr{C}(\mathbb{X})$ is an approximation operator mapping a given function $w \in \mathscr{C}(\mathbb{X})$ into $Lw \approx w$, where Lw can be stored on a computer. For example, the mapping $w \mapsto Lw$ might proceed by evaluating w on a fixed and finite grid of points $\{x_j\}_{j=1}^J$, and then constructing a spline Lw based on these "interpolation" points. In this example, when applying $\hat{R}_n = LR_n$ to some $v \in \mathscr{C}(\mathbb{X})$, the function $R_n v$ need only be evaluated J times; the function $LR_n v$ is the computed from this information.

In order to derive theoretical results on random fitted VFI, we assume throughout that the operator $L\colon \mathscr{C}(\mathbb{X}) \to \mathscr{C}(\mathbb{X})$ is nonexpansive, in the sense that

$$\|Lv - Lw\| \leq \|v - w\| \text{ for all pairs } v, w \in \mathscr{C}(\mathbb{X})$$

EXAMPLE 5.1 Piecewise linear interpolation is well suited to VFI (see, e.g., Santos and Vigo-Aguiar, 1998). To describe it, let $\mathbb{X}$ be a subset of $\mathbb{R}^d$, let $\mathbb{V}$ be a finite subset of $\mathbb{X}$ such that the convex hull of $\mathbb{V}$ is $\mathbb{X}$, and let T be a $\mathbb{V}$-triangularization of $\mathbb{X}$.[10] Given a simplex $\Delta \in T$ with vertices $\zeta_1, \ldots, \zeta_{d+1}$, each $x \in \Delta$ can be represented uniquely as $\sum_{i=1}^{d+1} \lambda(x,i)\zeta_i$, where $\lambda(x,i)$ is its i-th barycentric coordinate relative to Δ.[11] For $w \in \mathscr{C}(\mathbb{X})$, we define $L\colon \mathscr{C}(\mathbb{X}) \to \mathscr{C}(\mathbb{X})$ by $Lw(x) = \sum_{i=1}^{d+1} \lambda(x,i) w(\zeta_i)$. The operator L is nonexpansive.

Several other well-known approximation schemes are nonexpansive. Stachurski (2008) contains an extensive discussion.

LEMMA 5.1 *The operator $\hat{R}_n(\omega) := LR_n(\omega)$ is a contraction of modulus ρ on $\mathscr{C}(\mathbb{X})$ for all $n \in \mathbb{N}$ and all $\omega \in \Omega$.*

This result follows immediately from lemmas 2.1 and 4.1. As a consequence, for each $\omega \in \Omega$, there exists a unique $\hat{V}_n(\omega) \in \mathscr{C}(\mathbb{X})$ such that $\hat{R}_n(\omega)\hat{V}_n(\omega) =$

[10] That is, T is a finite collection of non-degenerate simplexes such that the vertices of each simplex lie in $\mathbb{V}$ and any two simplexes intersect on a common face or not at all. (A simplex is called non-degenerate if it has positive measure in $\mathbb{R}^d$.)

[11] Note that, by definition, $\lambda(x,i) \geq 0$ and $\sum_{i=1}^{d+1} \lambda(x,i) = 1$.

$\hat{V}_n(\omega)$. As was the case for V_n, we refer to $\hat{V}_n$ as a random function, although $\omega \mapsto \hat{V}_n(\omega)$ may not be Borel measurable as mapping from Ω to $\mathscr{C}(\mathbb{X})$.

Our goal is to show that $\|\hat{V}_n - V\|$ becomes small as $n \to \infty$ with high probability. Denoting the fixed point of $\hat{T} := LT$ by $\hat{V}$, the triangle inequality yields

$$\|\hat{V}_n - V\| \leq \|\hat{V}_n - \hat{V}\| + \|\hat{V} - V\| \qquad \forall\, n \in \mathbb{N} \tag{13}$$

Regarding the second term in the sum, one can show (using lemma 2.1) that $\|\hat{V} - V\| \leq (1-\rho)^{-1}\|LV - V\|$, and hence this term can be made arbitrarily small if L approximates V arbitrarily well. Further discussion of L is somewhat orthogonal to the concerns of this paper, and hence we just assume that $\|LV - V\|$ is small, and focus on convergence of the first term in the sum.

THEOREM 5.1 $\|\hat{V}_n - \hat{V}\| \overset{a.s.^*}{\to} 0$ *as* $n \to \infty$.

PROOF OF THEOREM 5.1: By lemma 2.1 and the nonexpansiveness of L,

$$\|\hat{V}_n - \hat{V}\| \leq \frac{1}{1-\rho}\|\hat{R}_n\hat{V} - \hat{T}\hat{V}\| \leq \frac{1}{1-\rho}\|R_n\hat{V} - T\hat{V}\| \tag{14}$$

We can repeat the arguments of the proof of theorem 4.1, substituting $\hat{V}$ in the place of V. Using (7), we obtain

$$|R_n\hat{V}(x) - T\hat{V}(x)| \leq \rho \max_{a\in\Gamma(x)} \left| \frac{1}{n}\sum_{i=1}^{n} \hat{V}[F(x,a,U_i)] - \int \hat{V}[F(x,a,u)]\phi(du) \right|$$

where $x \in \mathbb{X}$ is arbitrary. Taking the supremum over all $x \in \mathbb{X}$, we now have

$$\|R_nV - TV\| \leq \rho \max_{(x,a)\in\mathbb{G}} \left| \frac{1}{n}\sum_{i=1}^{n} \hat{V}[F(x,a,U_i)] - \int \hat{V}[F(x,a,u)]\phi(du) \right|$$

As before, let $y = (x,a)$ denote a typical element of $\mathbb{G}$, and let

$$\hat{h}_y(u) :=: \hat{h}_{(x,a)}(u) := \hat{V}[F(x,a,u)] :=: \hat{V}[F(y,u)] \tag{15}$$

Using the notation from the proof of theorem 4.1 and (14), we can now write

$$\|\hat{V}_n - \hat{V}\| \leq \frac{\rho}{1-\rho} \sup_{y\in\mathbb{G}} |\phi_n(\hat{h}_y) - \phi(\hat{h}_y)| \qquad \forall\, n \in \mathbb{N} \tag{16}$$

The class $\{\hat{h}_y\}_{y\in\mathbb{G}}$ is ϕ-Glivenko-Cantelli, as can be seen from identical arguments to the case of theorem 4.1. Therefore, the right-hand side converges to zero $\mathbf{P}^*$-almost surely. *Q.E.D.*

6. RATES OF CONVERGENCE

The result in theorem 5.1 gives no indication as to the *rate* of convergence. To obtain a rate, we need to give a rate for the right-hand side of (16). Since $\mathbb{G}$ is infinite, we require the convergence results of empirical process theory. These results hinge on the diversity of the function class $\{\hat{h}_y\}_{y\in\mathbb{G}}$. The ϕ-Glivenko-Cantelli property used in that proof of theorem 5.1 is not sufficient for rates, so further restrictions on $\{\hat{h}_y\}_{y\in\mathbb{G}}$ are required.

6.1. Donsker Classes

Let $\mathscr{H}$ be a class of uniformly bounded, measurable functions from $\mathbb{U}$ into $\mathbb{R}$, and let $(b\mathscr{H}, \|\cdot\|)$ be the Banach space of bounded, real valued functions on $\mathscr{H}$ with the supremum norm. The class $\mathscr{H}$ is called *ϕ-Donsker* if the empirical process

$$\nu_n(h) := \sqrt{n}(\phi_n(h) - \phi(h)) \qquad (n \in \mathbb{N},\, h \in \mathscr{H})$$

converges in distribution to a tight Gaussian process ν on $b\mathscr{H}$. Here $\omega \mapsto \nu_n(\cdot)(\omega)$ and $\omega \mapsto \nu(\cdot)(\omega)$ are maps from Ω into $b\mathscr{H}$. The maps $\omega \mapsto \nu_n(\cdot)(\omega)$ are not necessarily measurable, and convergence in distribution is to be understood in the sense of $\nu_n \xrightarrow{d^*} \nu$.

Letting $\{\hat{h}_y\}_{y\in\mathbb{G}}$ be the class of functions defined in (15), we can now state the following result.

PROPOSITION 6.1 *If $\{\hat{h}_y\}_{y\in\mathbb{G}}$ is ϕ-Donsker, then $\|\hat{V}_n - \hat{V}\| = O_{\mathbf{P}^*}(n^{-1/2})$.*

PROOF: We will need some preliminary results and additional notation. Let

$$G_n(y) := \nu_n(h_y) := \sqrt{n}(\phi_n(\hat{h}_y) - \phi(\hat{h}_y)) \qquad (n \in \mathbb{N},\, y \in \mathbb{G})$$

G_n can be understood as a real-valued stochastic process indexed by $y \in \mathbb{G}$:

$$G_n(y)(\omega) = \sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n} \hat{h}_y(U_i(\omega)) - \int \hat{h}_y(u)\phi(du)\right) \in \mathbb{R}$$

Regarding measurability, we have the following result, proved in the appendix:

LEMMA 6.1 *For each $n \in \mathbb{N}$, the following measurability results hold:*

1. $\omega \mapsto G_n(\cdot)(\omega)$ *is a* $\mathscr{C}(\mathbb{G})$*-valued random variable, and*
2. $\omega \mapsto \|G_n(\cdot)(\omega)\| = \sup_{y\in\mathbb{G}} |G_n(y)(\omega)|$ *is a real-valued random variable.*

In view of (16), we have

$$\|\hat{V}_n - \hat{V}\| \leq \frac{\rho}{1-\rho} n^{-1/2} \sup_{y\in\mathbb{G}} |n^{1/2}(\phi_n(h_y) - \phi(h_y))| \leq \frac{\rho}{1-\rho} n^{-1/2} \sup_{y\in\mathbb{G}} |G_n(y)|$$

Since $\mathscr{H} := \{\hat{h}_y\}_{y\in\mathbb{G}}$ is ϕ-Donsker, we have $\nu_n \xrightarrow{d^*} \nu$, where ν is a Gaussian process on $\mathscr{H}$. In view of lemma 2.2 and continuity of the norm $\|\cdot\|$ on $b\mathscr{H}$, we then have $\|\nu_n\| \xrightarrow{d^*} \|\nu\|$ in $\mathbb{R}$. Observe that

$$\|\nu_n\| = \sup_{h\in\mathscr{H}} |\nu_n(h)| = \sup_{y\in\mathbb{G}} |\nu_n(h_y)| = \sup_{y\in\mathbb{G}} |G_n(y)|$$

$$\therefore \quad \sup_{y\in\mathbb{G}} |G_n(y)| \xrightarrow{d^*} \|\nu\|$$

By part 2 of lemma 6.1, this is convergence in distribution in the regular sense, and, as a consequence, $\sup_{y\in\mathbb{G}} |G_n(y)| = O_{\mathbf{P}}(1)$. We then have

$$\|\hat{V}_n - \hat{V}\| \leq \frac{\rho}{1-\rho} n^{-1/2} O_{\mathbf{P}}(1) = O_{\mathbf{P}}(n^{-1/2})$$

This concludes the proof of proposition 6.1. *Q.E.D.*

6.2. The Lipschitz Case

In this section and the next, we use proposition 6.1 to obtain sufficient conditions for rates of convergence in different (and somewhat specialized) settings. Our first result is based on a Lipschitz condition. To apply the method, we now specialize to the case $\mathbb{G} \subset \mathbb{R}^d$ for some $d \in \mathbb{N}$, and add the following assumptions to those imposed in sections 2–5:

(i) Lw is Lipschitz continuous for every $w \in \mathscr{C}(\mathbb{X})$.[12]
(ii) There exists a measurable $m_0 \colon \mathbb{U} \to \mathbb{R}$ with $\int m_0^2 \, d\phi < \infty$ and

$$\|F(y,u) - F(y',u)\|_2 \leq m_0(u)\|y - y'\|_2 \qquad \forall\, y, y' \in \mathbb{G},\ u \in \mathbb{U} \tag{17}$$

[12]This condition depends on the approximation architecture used in the fitted VFI routine, and is satisfied by, for example, the piecewise linear interpolation operator in Example 5.1.

Here $\|\cdot\|_2$ represents the euclidean norm on $\mathbb{R}^d$.

PROPOSITION 6.2 *If* (i)–(ii) *hold, then* $\|\hat{V}_n - \hat{V}\| = O_{\mathbf{P}^*}(n^{-1/2})$.

The proof is given in the appendix.

EXAMPLE 6.1 If $\mathbb{U} \subset \mathbb{R}^k$, and F is linear, in the sense that

$$F(x,a,u) = Ax + Ba + Cu \qquad (x \in \mathbb{X},\ a \in \Gamma(x),\ u \in \mathbb{U})$$

for matrices A, B and C, then assumption (ii) is satisfied. To see this, observe that for any $y = (x,a) \in \mathbb{G}$, $y' = (x',a') \in \mathbb{G}$, and $u \in \mathbb{U}$,

$$\begin{aligned}\|Ax + Ba + Cu - Ax' - Ba' - Cu\|_2 \\ = \|A(x-x') + B(a-a')\|_2 \le \gamma(\|x-x'\|_2 + \|a-a'\|_2)\end{aligned}$$

where γ is the maximum of the operator norms of A and B. Since $y = (x,a) \mapsto \|x\|_2 + \|a\|_2 \in \mathbb{R}$ defines a norm on $\mathbb{R}^d$, and since all norms on $\mathbb{R}^d$ are equivalent, we obtain

$$\|F(y,u) - F(y',u)\|_2 \le M\gamma\|y - y'\|_2 \qquad \forall\, y, y' \in \mathbb{G},\ u \in \mathbb{U}$$

for some $M < \infty$. This verifies (ii).

6.3. The Monotone Case

Another way to establish the ϕ-Donsker property is via monotonicity. To this end, we make the following assumptions (in addition to the basic assumptions imposed in sections 2–5). Let $\mathbb{X} \subset \mathbb{R}^d$ and let $\mathbb{U} \subset \mathbb{R}$. We assume now that

(i) L maps $i\mathscr{C}(\mathbb{X})$ to itself, where $i\mathscr{C}(\mathbb{X})$ is the increasing functions in $\mathscr{C}(\mathbb{X})$.
(ii) For all $x, x' \in \mathbb{X}$ with $x \le x'$,
  (a) $\Gamma(x) \subset \Gamma(x')$,
  (b) $r(x,a) \le r(x',a)$ for all $a \in \Gamma(x)$, and
  (c) $F(x,a,u) \le F(x',a,u)$ for all $a \in \Gamma(x)$ and $u \in \mathbb{U}$.
(iii) For all $y \in \mathbb{G}$, we have $F(y,u) \le F(y,u')$ whenever $u \le u'$.

Assumption (i) depends on the approximation architecture, and is satisfied by, for example, the piecewise linear interpolation operator in Example 5.1. The other assumptions are discussed below.

PROPOSITION 6.3 *If* (i)–(iii) *hold, then* $\|\hat{V}_n - \hat{V}\| = O_{\mathbf{P}^*}(n^{-1/2})$.

The proof is given in the appendix.

7. CONCLUSION

We studied VFI when the integral in the Bellman operator is evaluated using Monte Carlo. Employing the techniques of empirical process theory, we showed that iterating with the resulting random Bellman operator R_n generates a sequence of functions that converges, and that the limit point is close to V when the sample size is large. We also proved a version of the same result when the iterates of the Bellman operator must be approximated, as is necessary in continuous state spaces. In addition, we provided results on asmptotic rates of convergence.

We treated only VFI for stationary, infinite horizon SDPs. We did not treat nonstationary SDPs, finite horizon models, or optimal stopping. Providing convergence results for such models will require somewhat different methods, and we leave these problems for future research.

8. REMAINING PROOFS

PROOF OF LEMMA 4.1: Fix $n \in \mathbb{N}$ and $\omega \in \Omega$. Let $R := R_n(\omega)$. Fix $w, w' \in \mathscr{C}(\mathbb{X})$ and $x \in \mathbb{X}$. In view of (7), we have

$$|Rw(x) - Rw'(x)| \leq \rho \max_{a \in \Gamma(x)} \left| \frac{1}{n} \sum_{i=1}^{n} w[F(x, a, U_i(\omega))] - \frac{1}{n} \sum_{i=1}^{n} w'[F(x, a, U_i(\omega))] \right|$$

Using the triangle inequality and the definition of $\|\cdot\|$, we obtain

$$|Rw(x) - Rw'(x)| \leq \rho \|w - w'\|$$

Taking the supremum over $x \in \mathbb{X}$ yields the desired result. *Q.E.D.*

PROOF OF LEMMA 6.1: We begin by proving measurability of $\omega \mapsto H(\cdot)(\omega)$, where

$$H(y)(\omega) = \hat{h}_y(U(\omega)) = \hat{V}[F(y, U(\omega))]$$

Note that since $\mathbb{X}$ and $\mathbb{A}$ are compact, and since Γ is continuous and compact-valued, $\mathbb{G}$ is compact in the product topology. From the Stone–Weierstrass theorem it follows that $\mathscr{C}(\mathbb{G})$ is separable. Hence, by the Pettis measurability theorem, we need only show that $\omega \mapsto \ell(H(\cdot)(\omega))$ is measurable for each ℓ in the dual space $\mathscr{C}(\mathbb{G})^*$ of $\mathscr{C}(\mathbb{G})$. By the Riesz representation theorem, $\mathscr{C}(\mathbb{G})^*$ can be identified with $\mathscr{M}(\mathbb{G})$, the space of finite signed Borel measures on $\mathbb{G}$. Thus, it remains to show that

$$\Omega \ni \omega \mapsto \int H(y)(\omega)\,\gamma(dy) \in \mathbb{R} \text{ is measurable} \quad \forall\, \gamma \in \mathscr{M}(\mathbb{G})$$

To this end it is sufficient to show that $H(y)(\omega) = \hat{V}[F(y, U(\omega))]$ is measurable with respect to the product σ-algebra $\mathscr{B}_{\mathbb{G}} \otimes \mathscr{F}$, where $\mathscr{B}_{\mathbb{G}}$ is the Borel σ-algebra on $\mathbb{G}$. Since H is continuous with respect to y and measurable with respect to ω, H is a Carathéodory function (Aliprantis and Border, 1999, def. 4.49). As $\mathbb{G}$ is separable, measurability with respect to $\mathscr{B}_{\mathbb{G}} \otimes \mathscr{F}$ is established (Aliprantis and Border, 1999, lem. 4.50).

Given measurability of $\omega \mapsto H(\cdot)(\omega)$, measurability of $\omega \mapsto G_n(\cdot)(\omega)$ follows from the fact that linear combinations of measurable random elements of a separable Banach space are themselves measurable.

Regarding the second claim in the lemma, measurability of $\omega \mapsto \|G(\cdot)(\omega)\|$ follows from measurability of $\omega \mapsto G(\cdot)(\omega)$, continuity of the norm as a map from $\mathscr{C}(\mathbb{G})$ to $\mathbb{R}$, and the fact that continuous transformations of Borel measurable mappings are Borel measurable. *Q.E.D.*

PROOF OF PROPOSITION 6.2: By proposition 6.1, it suffices to show that the class $\{\hat{h}_y\}_{y\in\mathbb{G}}$ is ϕ-Donsker when (i)–(iii) hold. A sufficient condition for $\{\hat{h}_y\}_{y\in\mathbb{G}}$ to be ϕ-Donsker is the existence of a measurable function $m\colon \mathbb{U} \to \mathbb{R}$ such that $\int m^2 d\phi < \infty$ and

$$|h_y(u) - h_{y'}(u)| \leq m(u)\|y - y'\|_2 \qquad \forall\, y, y' \in \mathbb{G},\ u \in \mathbb{U} \tag{18}$$

(see, e.g., van der Vaart, 1998, p. 271). To find such an m, observe that $\hat{V}$ is Lipschitz, as follows from (ii) and the relation $\hat{V} = LT\hat{V}$. As a consequence, there exists a $K < \infty$ such that, for any $y, y' \in \mathbb{G}$ and $u \in \mathbb{U}$, we have

$$\begin{aligned} |h_y(u) - h_{y'}(u)| &:= |\hat{V}[F(y,u)] - \hat{V}[F(y',u)]| \\ &\leq K\|F(y,u) - F(y',u)\|_2 \leq Km_0(u)\|y - y'\|_2 \end{aligned}$$

where m_0 is the function in (iii). Letting $m := Km_0$, we see that $\int m^2 d\phi = K^2 \int m_0^2 d\phi < \infty$. All the conditions are now verified, and hence $\{\hat{h}_y\}_{y\in\mathbb{G}}$ is ϕ-Donsker. *Q.E.D.*

PROOF OF PROPOSITION 6.3: From van der Vaart (1998, p. 273), it suffices to show that the class $\{\hat{h}_y\}_{y\in\mathbb{G}}$ is uniformly bounded on $\mathbb{U}$, and that each element $\hat{h}_y$ is monotone increasing on $\mathbb{U}$. Since $\hat{h}_y(u) = \hat{V}[F(y,u)]$, uniform boundedness will hold if $\hat{V}$ is bounded on $\mathbb{X}$. That this is the case follows from the fact that $\mathbb{X}$ is compact and $\hat{V} \in \mathscr{C}(\mathbb{X})$.

Regarding monotonicity, we begin by showing that $\hat{V}$ is monotone increasing. To see that this is the case, observe that $\hat{V}$ is the fixed point of LT in $\mathscr{C}(\mathbb{X})$. Since $i\mathscr{C}(\mathbb{X})$ is a closed subset of $\mathscr{C}(\mathbb{X})$, we need only show that LT maps $i\mathscr{C}(\mathbb{X})$ into itself. Since $L\colon i\mathscr{C}(\mathbb{X}) \to i\mathscr{C}(\mathbb{X})$ by assumption, it remains to verify that T also has this property. For a proof of this fact, see Stachurski (2009, theorem 12.1.2). As a result, $\hat{V}$ is increasing, and the claim in the proposition now follows from assumption 3 above. *Q.E.D.*

REFERENCES

[1] Aliprantis, C. D. and K. C. Border (1999): *Infinite Dimensional Analysis,* Springer-Verlag, New York.

[2] Bellman, R.E. (1957): *Dynamic Programming,* Princeton University Press, Princeton, NJ.

[3] Den Haan, W. and A. Marcet (1990): "Solving the Stochastic Growth Model by Parametrizing Expectations," *Journal of Business and Economic Statistics,* 8, 31–34.

[4] Keane, M. P. and K. I. Wolpin (1994): "The Solution and Estimation of Discrete Choice Dynamic Programming Models by Simulation," The Review of Economics and Statistics, 76 (4), 648–672.

[5] Longstaff, F. A. and E. S. Schwartz (2001): "Valuing American Options by Simulation: A Simple Least-squares Approach," *Review of Financial Studies,* 14, 113–147.

[6] Pakes, A. and P. McGuire (2001): "Stochastic Algorithms, Symmetric Markov Perfect Equilibrium, and the 'Curse' of Dimensionality," *Econometrica,* 69 (5), 1261–1281.

[7] Puterman, M. (1994): *Markov Decision Processes: Discrete Stochastic Dynamic Programming,* John Wiley & Sons, New York.

[8] Rust, J. (1997): "Using Randomization to Break the Curse of Dimensionality," *Econometrica,* 65 (3), 487–516.

[9] Santos, M.S. and J. Vigo-Aguiar (1998): "Analysis of a Numerical Dynamic Programming Algorithm Applied to Economic Models," *Econometrica,* 66(2), 409–426.

[10] Stachurski, J. (2008): "Continuous State Dynamic Programming via Nonexpansive Approximation," *Computational Economics,* 31 (2), 141–160.

[11] Stachurski, J. (2009): *Economic Dynamics: Theory and Computation,* The MIT Press, Cambridge MA.

[12] van der Vaart, A. W. (1998): *Asymptotic Statistics,* Cambridge series in statistical and probabilistic mathematics, Cambridge University Press, UK.