The Annals of Statistics
2011, Vol. 39, No. 2, 673–701
DOI: 10.1214/10-AOS831

CONSISTENCY OF MARKOV CHAIN QUASI-MONTE CARLO ON CONTINUOUS STATE SPACES

BY S. CHEN[1], J. DICK[2] AND A. B. OWEN[1]

Stanford University, University of New South Wales and Stanford University

The random numbers driving Markov chain Monte Carlo (MCMC) simulation are usually modeled as independent $U(0, 1)$ random variables. Tribble [Markov chain Monte Carlo algorithms using completely uniformly distributed driving sequences (2007) Stanford Univ.] reports substantial improvements when those random numbers are replaced by carefully balanced inputs from completely uniformly distributed sequences. The previous theoretical justification for using anything other than i.i.d. $U(0, 1)$ points shows consistency for estimated means, but only applies for discrete stationary distributions. We extend those results to some MCMC algorithms for continuous stationary distributions. The main motivation is the search for quasi-Monte Carlo versions of MCMC. As a side benefit, the results also establish consistency for the usual method of using pseudo-random numbers in place of random ones.

1. Introduction. In Markov chain Monte Carlo (MCMC), one simulates a Markov chain and uses sample averages to estimate corresponding means of the stationary distribution of the chain. MCMC has become a staple tool in the physical sciences and in Bayesian statistics. When sampling the Markov chain, the transitions are driven by a stream of independent $U(0, 1)$ random numbers.

In this paper, we study what happens when the i.i.d. $U(0, 1)$ random numbers are replaced by deterministic sequences, or by some dependent $U(0, 1)$ values. The motivation for replacing i.i.d. $U(0, 1)$ points is that carefully stratified inputs may lead to more accurate sample averages. One must be cautious though, because as with adaptive MCMC [3, 21], the resulting simulated points do not have the Markov property.

The utmost in stratification is provided by quasi-Monte Carlo (QMC) points. There were a couple of attempts at merging QMC into MCMC around 1970, and then again starting in the late 1990s. It is only recently that significant improvements have been reported in numerical investigations. For example, Tribble [43] reports variance reductions of several thousand fold and an apparent improved

Received August 2009; revised May 2010.
[1]Supported by NSF Grants DMS-06-04939 and DMS-09-06056.
[2]Supported by a Queen Elizabeth II fellowship.
MSC2010 subject classifications. Primary 65C40, 62F15; secondary 26A42, 65C05.
Key words and phrases. Completely uniformly distributed, coupling, iterated function mappings, Markov chain Monte Carlo.

convergence rate for some Gibbs sampling problems. Those results motivate our theoretical work. They are described more fully in the literature survey below.

To describe our contribution, represent MCMC sampling via $\mathbf{x}_{i+1} = \phi(\mathbf{x}_i, \mathbf{u}_i)$ for $i = 1, \dots, n$, where $\mathbf{x}_0$ is a nonrandom starting point and $\mathbf{u}_i \in (0,1)^d$. The points $\mathbf{x}_i$ belong to a state space $\Omega \subset \mathbb{R}^s$. The function ϕ is chosen so that $\mathbf{x}_i$ form an ergodic Markov chain with the desired stationary distribution π when $\mathbf{u}_i \sim U(0,1)^d$ independently. For a bounded continuous function $f : \Omega \to \mathbb{R}$, let $\theta(f) = \int_\Omega f(\mathbf{x})\pi(\mathbf{x})\,d\mathbf{x}$ and $\hat{\theta}_n(f) = (1/n)\sum_{i=1}^n f(\mathbf{x}_i)$. Then $\hat{\theta}_n(f) \to_{\mathbb{P}} \theta(f)$ as $n \to \infty$. In this paper, we supply sufficient conditions on ϕ and on the deterministic sequences $\mathbf{u}_i$ so that $\hat{\theta}_n(f) \to \theta(f)$ holds when those deterministic sequences are used instead of random ones. The main condition is that the components of $\mathbf{u}_i$ be taken from a completely uniformly distributed (CUD) sequence, as described below.

Ours are the first results to prove that deterministic sampling applied to MCMC problems on continuous state spaces is consistent. In practice, of course, floating point computations take place on a large discrete state space. But invoking finite precision does not provide a satisfying description of continuous MCMC problems. In a finite state space argument, the resulting state spaces are so big that vanishingly few states will ever be visited in a given simulation. Then if one switches from 32 to 64 to 128 bit representations, the problem seemingly requires vastly larger sample sizes, but in reality is not materially more difficult.

To avoid using the finite state shortcut, we adopt a computational model with infinite precision. As a side benefit, this paper shows that the standard practice of replacing genuine i.i.d. values $\mathbf{u}_i$ by deterministic pseudo-random numbers is consistent for some problems with continuous state spaces. We do not think many people doubted this, but neither has it been established before, to our knowledge. It is already known from Roberts, Rosenthal and Schwartz [40] that, under certain conditions, a geometrically ergodic Markov chain remains so under small perturbations, such as rounding. That work does not address the replacement of random points by deterministic ones that we make here.

1.1. *Literature review.* There have been a small number of prior attempts to apply QMC sampling to MCMC problems. The first appears to have been Chentsov [7], whose work appeared in 1967, followed by Sobol' [42] in 1974. Both papers assume that the Markov chain has a discrete state space and that the transitions are sampled by inversion. Unfortunately, QMC does not usually bring large performance improvements on such unsmooth problems and inversion is not a very convenient method.

Chentsov replaces i.i.d. samples by one long CUD sequence, and this is the method we will explain and then adapt to continuous problems. Sobol' uses what is conceptually an $n \times \infty$ matrix of values from the unit interval. Each row is used to make transitions until the chain returns to its starting state. Then the sampling starts using the next row. It is like deterministic regenerative sampling. Sobol'

shows that the error converges as $O(1/n)$ in the very special case where the transition probabilities are all rational numbers with denominator a power of 2. These methods were not widely cited and, until recently, were almost forgotten, probably due to the difficulty of gaining large improvements in discrete problems, and the computational awkwardness of inversion as a transition mechanism for discrete state spaces.

The next attempt that we found is that of Liao [27] in 1998. Liao takes a set of QMC points in $[0, 1]^d$ shuffles them in random order, and uses them to drive an MCMC. He reports 4- to 25-fold efficiency improvements, but gives no theory. An analysis of Liao's method is given in [44]. Later, Chaudary [6] tried a different strategy using QMC to generate balanced proposals for Metropolis–Hastings sampling, but found only small improvements and did not publish the work. Craiu and Lemieux [8] also consider multiple-try Metropolis and find variance reductions of up to 30%, which is still modest. Earlier, Lemieux and Sidorsky [26] report variance reduction factors ranging from about 1.5 to about 18 in some work using QMC in conjunction with the perfect sampling method of Propp and Wilson [38].

Only recently have there been significantly large benefits from the combination of QMC and MCMC. Those benefits have mainly arisen for problems on continuous state spaces. Tribble's [43] best results come from Gibbs sampling problems computing posterior means. For problems with d parameters, he used every d-tuple from a small custom built linear feedback shift register (LFSR). One example is the well-known model used by Gelfand and Smith [16] for failure events of 10 pumps from the article by Gaver and O'Murcheartaigh [15]. There are 11 unknown parameters, one for each pump and one for the scale parameter in the distribution of pump failure rates. A second example is a 42 parameter probit model for vasorestriction based on a famous data set from [14] and analyzed using latent variables as in Albert and Chib [1]. Of those 42 parameters, the 3 regression coefficients are of greatest interest and 39 latent variables are nuisance variables. Table 1 sets out variance reduction factors found for randomized CUD versus i.i.d. sampling. The improvements appear to grow with n, and are evident at very small sample sizes.

TABLE 1

Variance reduction factors from Tribble [43] *for two Gibbs sampling problems. For the pumps data, the greatest and least variance reduction for a randomized CUD sequence versus i.i.d. sampling is shown. For the vasorestriction data, greatest and least variance reductions for the three regression parameters are shown. See* [43] *for simulation details*

	$n = 2^{10}$		$n = 2^{12}$		$n = 2^{14}$	
Data	**min**	**max**	**min**	**max**	**min**	**max**
Pumps	286	1543	304	5003	1186	16089
Vasorestriction	14	15	56	76	108	124

There is another line of research in which large improvements have been obtained by combining QMC with MCMC. This is the array-RQMC method described in L'Ecuyer, Lecot and Tuffin [24] and other articles. That method simulates numerous chains in parallel using quasi-Monte Carlo to update all the chains. It requires a complicated method to match the update variables for each step to the various evolving chains. This method has achieved variance reductions of many thousand fold on some problems from queuing and finance. Very few properties have been established for it, beyond the case of heat particles in one dimension that was considered by Morokoff and Caflisch [31].

Finally, Jim Propp's rotor-router method is a form of deterministic Markov chain sampling. It has brought large efficiency improvements for some problems on a discrete state space and has been shown to converge at better than the Monte Carlo rate on some problems. See, for example, Doerr and Friedrich [13].

The use of CUD sequences that we study has one practical advantage compared to the rotor-router, array-RQMC, regenerative sampling, and the other methods. It only requires replacing the i.i.d. sequence used in a typical MCMC run by some other list of numbers.

1.2. *Outline*. The paper is organized around our main results which appear in Section 3. Theorem 2 gives sufficient conditions for consistency of QMC-MCMC sampling by Metropolis–Hastings. Theorem 3 gives sufficient conditions for consistency of QMC-MCMC sampling for the systematic scan Gibbs sampler.

Section 2 contains necessary background and notation for the two main theorems of Section 3. It introduces quasi-Monte Carlo and Markov chain Monte Carlo giving key definitions we need in each case. That section presents the Rosenblatt–Chentsov transformation. We have combined a classic sequential inversion method based on the Rosenblatt transformation with an elegant coupling argument that Chentsov [7] used.

The consistency results for Metropolis–Hastings (Theorem 2) make moderately strong assumptions in order to ensure that a coupling occurs. Section 4 shows that those assumptions are satisfied by some Metropolized independence samplers and also by some slice samplers. We also assumed some Riemann integrability properties for our MCMC proposals. The Riemann integral is awkward compared to the Lebesgue integral, but considering it is necessary when we want to study specific algorithms on deterministic inputs. Section 5 gives sufficient conditions for an MCMC algorithm to satisfy the required Riemann integrability conditions.

Our consistency results for the Gibbs sampler (Theorem 3) require some contraction properties and some Jordan measurability. Section 6 shows that these properties hold under reasonable conditions. Section 7 has a brief discussion on open versus closed intervals for uniform random numbers. Our conclusions are in Section 8. The lengthier or more technical proofs are placed in the Appendix.

2. Background on QMC and MCMC.

2.1. *Notation.* Our random vectors are denoted by $\mathbf{x} = (x_1, \dots, x_s) \in \Omega \subseteq \mathbb{R}^s$ for $s \geq 1$. Points in the unit cube $[0,1]^d$ are denoted by $\mathbf{u} = (u_1, \dots, u_d)$. Two points $\mathbf{a}, \mathbf{b} \in \mathbb{R}^d$ with $a_j < b_j$ for $j = 1, \dots, d$ define a rectangle $\prod_{j=1}^d [a_j, b_j]$, denoted by $[\mathbf{a}, \mathbf{b}]$ for short. The indicator (or characteristic) function of a set $A \subset \mathbb{R}^d$ is written 1_A.

We assume the reader is familiar with the definition of the (proper) Riemann integral, for a bounded function on a finite rectangle $[\mathbf{a}, \mathbf{b}] \subset \mathbb{R}^d$. The bounded set $A \subset \mathbb{R}^d$ is Jordan measurable if 1_A is Riemann integrable on a bounded rectangle containing A. By Lebesgue's theorem (see Section 5) A is Jordan measurable if $\lambda_d(\partial A) = 0$. Here λ_d denotes Lebesgue measure on $\mathbb{R}^d$, and ∂A is the boundary of A, that is, the set on which 1_A is discontinuous.

2.2. *QMC background.* Here, we give a short summary of quasi-Monte Carlo. Further information may be found in the monograph by Niederreiter [34].

QMC is ordinarily used to approximate integrals over the unit cube $[0,1]^d$, for $d \in \mathbb{N}$. Let $\mathbf{x}_1, \dots, \mathbf{x}_n \in [0,1]^d$. The QMC estimate of $\theta(f) = \int_{[0,1]^d} f(\mathbf{x})\, d\mathbf{x}$ is $\hat{\theta}_n(f) = \frac{1}{n} \sum_{i=1}^n f(\mathbf{x}_i)$, just as we would use in plain Monte Carlo. The difference is that in QMC, distinct points $\mathbf{x}_i$ are chosen deterministically to make the discrete probability distribution with an atom of size $1/n$ at each $\mathbf{x}_i$ close to the continuous $U[0,1]^d$ distribution.

The distance between these distributions is quantified by discrepancy measures. The local discrepancy of $\mathbf{x}_1, \dots, \mathbf{x}_n$ at $\mathbf{a} \in [0,1]^d$ is

$$\delta(\mathbf{a}) = \delta(\mathbf{a}; \mathbf{x}_1, \dots, \mathbf{x}_n) = \frac{1}{n} \sum_{i=1}^n 1_{[\mathbf{0}, \mathbf{a})}(\mathbf{x}_i) - \prod_{j=1}^d a_j. \tag{1}$$

The star discrepancy of $\mathbf{x}_1, \dots, \mathbf{x}_n$ in dimension d is

$$D_n^{*d} = D_n^{*d}(\mathbf{x}_1, \dots, \mathbf{x}_n) = \sup_{\mathbf{a} \in [0,1]^d} |\delta(\mathbf{a}; \mathbf{x}_1, \dots, \mathbf{x}_n)|. \tag{2}$$

For $d = 1$, the star discrepancy reduces to the Kolmogorov–Smirnov distance between a discrete and a continuous uniform distribution.

A uniformly distributed sequence is one for which $D_n^{*d} \to 0$ as $n \to \infty$. If $\mathbf{x}_i$ are uniformly distributed then $\hat{\theta}_n(f) \to \theta(f)$ provided that f is Riemann integrable.

Under stronger conditions than Riemann integrability, we can get rates of convergence for QMC. The Koksma–Hlawka inequality is

$$|\hat{\theta}_n(f) - \theta(f)| \leq D_n^{*d} V_{\mathrm{HK}}(f), \tag{3}$$

where V_{HK} is the total variation of f in the sense of Hardy and Krause. For properties of V_{HK} and other multidimensional variation measures, see [36].

Equation (3) gives a deterministic upper bound on the integration error, and it factors into a measure of the points' quality and a measure of the integrand's roughness. There exist constructions $\mathbf{x}_1, \dots, \mathbf{x}_n$ where $D_n^{*d} = O(n^{-1+\epsilon})$ holds for any $\epsilon > 0$. Therefore, functions of finite variation can be integrated at a much better rate by QMC than by MC. Rates of convergence of $O(n^{-\alpha}(\log n)^{d\alpha})$, where $\alpha \geq 1$ denotes the smoothness of the integrand which can therefore be arbitrarily large, can also be achieved [12].

Equation (3) is not usable for error estimation. Computing the star discrepancy is very difficult [19], and computing $V_{\mathrm{HK}}(f)$ is harder than integrating f. Practical error estimates for QMC may be obtained using randomized quasi-Monte Carlo (RQMC). In RQMC each $\mathbf{x}_i \sim U[0,1]^d$ individually while the ensemble $\mathbf{x}_1, \dots, \mathbf{x}_n$ has $\Pr(D_n^{*d}(\mathbf{x}_1, \dots, \mathbf{x}_n) < C(\log n)^d/n) = 1$ for some $C < \infty$. For an example, see [35]. A small number of independent replicates of the RQMC estimate can be used to get an error estimate. RQMC has the further benefit of making QMC unbiased. For a survey of RQMC, see [25].

A key distinction between QMC and MC is that the former is effective for Riemann integrable functions, while the latter, in principle, works for Lebesgue integrable functions. In practice, MC is usually implemented with deterministic pseudo-random numbers. The best generators are proved to simulate independent $U[0,1]$ random variables based on either discrepancy measures over rectangles or on spectral measures. Those conditions are enough to prove convergence for averages of Riemann integrable functions, but not for Lebesgue integrable functions. As a result, ordinary Monte Carlo with pseudo-random numbers is also problematic for Lebesgue integrable functions that are not Riemann integrable.

2.3. *Completely uniformly distributed.* In the Markov chain context, we need a lesser known QMC concept as follows. A sequence $u_1, u_2, \dots \in [0,1]$ is completely uniformly distributed (CUD) if for any $d \geq 1$ the points $\mathbf{x}_i^{(d)} = (u_i, \dots, u_{i+d-1})$ satisfy $D_n^{*d}(\mathbf{x}_1^{(d)}, \dots, \mathbf{x}_n^{(d)}) \to 0$ as $n \to \infty$. This is one of the definitions of a random sequence from Knuth [22], and it is an important property for modern random number generators.

Using a CUD sequence in an MCMC is akin to using up the entire period of a random number generator, as remarked by Niederreiter [33] in 1986. It is then necessary to use a small random number generator. The CUD sequences used by Tribble [43] are miniature versions of linear congruential generators and feedback shift register generators. As such, they are no slower than ordinary pseudo-random numbers.

In the QMC context, we need to consider nonoverlapping d-tuples $\widetilde{\mathbf{x}}_i^{(d)} = (u_{di-d+1}, \dots, u_{di})$ for $i \geq 1$. It is known [7] that

$$\begin{aligned} &D_n^{*d}\bigl(\mathbf{x}_1^{(d)}, \dots, \mathbf{x}_n^{(d)}\bigr) \to 0 \qquad \forall d \geq 1, \\ &\qquad\qquad \Longleftrightarrow \\ &D_n^{*d}\bigl(\widetilde{\mathbf{x}}_1^{(d)}, \dots, \widetilde{\mathbf{x}}_n^{(d)}\bigr) \to 0 \qquad \forall d \geq 1. \end{aligned} \tag{4}$$

2.4. *MCMC iterations*. In the QMC context, the function f subsumes all the necessary transformations to turn a finite list of i.i.d. $U[0,1]$ random variables into the desired nonuniformly distributed quantities, as well as the function of those quantities whose expectation we seek. In some problems, we are unable to find such transformations, and so we turn to MCMC methods.

Suppose that we want to sample $\mathbf{x} \sim \pi$ for a density function π defined with respect to Lebesgue measure on $\Omega \subseteq \mathbb{R}^s$. For definiteness, we will seek to approximate $\theta(f) = \int_\Omega f(\mathbf{x})\pi(\mathbf{x})\,d\mathbf{x}$. In this section, we briefly present MCMC. For a full description of MCMC, see the monographs by Liu [28] or Robert and Casella [39].

In an MCMC simulation, we choose an arbitrary $\mathbf{x}_0 \in \Omega$ with $\pi(\mathbf{x}_0) > 0$ and then for $i \geq 1$ update via

$$\mathbf{x}_i = \phi(\mathbf{x}_{i-1}, \mathbf{u}_i), \tag{5}$$

where $\mathbf{u}_i \in [0,1]^d$ and ϕ is an update function described below. The distribution of $\mathbf{x}_i$ depends on $\mathbf{x}_0, \ldots, \mathbf{x}_{i-1}$ only through $\mathbf{x}_{i-1}$ and so these random variables have the Markov property. The function ϕ is chosen so that the stationary distribution of $\mathbf{x}_i$ is π. Then we estimate $\theta(f)$ by $\hat{\theta}_n(f) = \frac{1}{n}\sum_{i=1}^n f(\mathbf{x}_i)$ as before. If a burn-in period was used, we assume that $\mathbf{x}_0$ is the last point of it.

First, we describe the Metropolis–Hastings algorithm for computing $\phi(\mathbf{x}, \mathbf{u})$ from the current point $\mathbf{x} \in \Omega$ and $\mathbf{u} \in [0,1]^d$. It begins with a proposal $\mathbf{y}$ taken from a transition kernel $P(\mathbf{x}, d\mathbf{y})$. With genuinely random proposals, the transition kernel gives a complete description. But for either quasi-Monte Carlo or pseudo-random sampling, it matters how we actually generate the proposal. We will assume that $d-1$ $U[0,1]$ random variables are used to generate $\mathbf{y}$ via $\mathbf{y} = \psi_{\mathbf{x}}(u_{1:(d-1)})$. Then the proposal $\mathbf{y}$ is either accepted or rejected with probability $A(\mathbf{x}, \mathbf{y})$. The decision is typically based on whether the dth random variable u_d is below A.

DEFINITION 1 (Generator). The function $\psi : [0,1]^d \to \mathbb{R}^s$ is a generator for the distribution F on $\mathbb{R}^s$ if $\psi(\mathbf{u}) \sim F$ when $\mathbf{u} \sim U[0,1]^d$.

DEFINITION 2 (Metropolis–Hastings update). For $\mathbf{x} \in \Omega$, let $\psi_{\mathbf{x}} : [0,1]^{d-1} \to \Omega$ be a generator for the transition kernel $P(\mathbf{x}, d\mathbf{y})$ with conditional density $p(\cdot \mid \mathbf{x})$. The Metropolis–Hastings sampler has

$$\phi(\mathbf{x}, \mathbf{u}) = \begin{cases} \mathbf{y}(\mathbf{x}, \mathbf{u}), & u_d \leq A(\mathbf{x}, \mathbf{u}), \\ \mathbf{x}, & u_d > A(\mathbf{x}, \mathbf{u}), \end{cases}$$

where $\mathbf{y}(\mathbf{x}, \mathbf{u}) = \psi_{\mathbf{x}}(\mathbf{u}_{1:(d-1)})$ and

$$A(\mathbf{x}, \mathbf{u}) = \min\left(1, \frac{\pi(\mathbf{y}(\mathbf{x}, \mathbf{u}))\, p(\mathbf{x} \mid \mathbf{y}(\mathbf{x}, \mathbf{u}))}{\pi(\mathbf{x})\, p(\mathbf{y}(\mathbf{x}, \mathbf{u}) \mid \mathbf{x})}\right).$$

EXAMPLE 1 [Metropolized independence sampler (MIS)]. The MIS update is a special case of the Metropolis–Hastings update in which $\mathbf{y}(\mathbf{x}, \mathbf{u}) = \psi(\mathbf{u}_{1:(d-1)})$ does not depend on $\mathbf{x}$.

EXAMPLE 2 [Random walk Metropolis (RWM)]. The RWM update is a special case of the Metropolis–Hastings update in which $\mathbf{y}(\mathbf{x}, \mathbf{u}) = \mathbf{x} + \psi(\mathbf{u}_{1:(d-1)})$ for some generator ψ not depending on $\mathbf{x}$.

DEFINITION 3 (Systematic scan Gibbs sampler). Let $\mathbf{x} = (x_1, \dots, x_s) \in \mathbb{R}^d$ with $x_j \in \mathbb{R}^{k_j}$ and $d = \sum_{j=1}^s k_j$. To construct the systematic scan Gibbs sampler, let $\psi_{j,\mathbf{x}_{-j}}(\mathbf{u}_j)$ be a k_j-dimensional generator of the full conditional distribution of x_j given x_ℓ for all $\ell \neq j$. This Gibbs sampler generates the new point using $\mathbf{u} \in [0,1]^d$. Write $\mathbf{u} = (\mathbf{u}_1, \dots, \mathbf{u}_s)$ with $\mathbf{u}_j \in [0,1]^{k_j}$. The systematic scan Gibbs sampler has

$$\phi(\mathbf{x}, \mathbf{u}) = (\phi_1(\mathbf{x}, \mathbf{u}), \phi_2(\mathbf{x}, \mathbf{u}), \dots, \phi_s(\mathbf{x}, \mathbf{u})),$$

where, for $1 \le j \le s$,

$$\phi_j(\mathbf{x}, \mathbf{u}) = \psi_{j,\mathbf{x}_{[j]}}(\mathbf{u}_j)$$

and $\mathbf{x}_{[j]} = (\phi_1(\mathbf{x}, \mathbf{u}), \dots, \phi_{j-1}(\mathbf{x}, \mathbf{u}), x_{j+1}, \dots, x_d)$.

EXAMPLE 3 (Inversive slice sampler). Let π be a probability density function on $\Omega \subset \mathbb{R}^s$. Let $\Omega' = \{(y, \mathbf{x}) \mid \mathbf{x} \in \Omega, 0 \le y \le \pi(\mathbf{x})\} \subset \mathbb{R}^{s+1}$ and let π' be the uniform distribution on Ω'. The inversive slice sampler is the systematic scan Gibbs sampler for π' with each $k_j = 1$ using inversion for every $\psi_{j,\mathbf{x}_{[j]}}$.

There are many other slice samplers. See [32]. It is elementary that $(y, \mathbf{x}) \sim \pi'$ implies $\mathbf{x} \sim \pi$. It is more usual to use $(\mathbf{x}, y)$, but our setting simplifies when we assume y is updated first.

2.5. *Some specific generators*. We generate our random variables as functions of independent uniform random variables. The generators we consider require a finite number of inputs, so acceptance-rejection is not directly covered, but see the note in Section 8.

For an encyclopedic presentation of methods to generate nonuniform random vectors, see Devroye [9]. Here, we limit ourselves to inversion and some generalizations culminating in the Rosenblatt–Chentsov transformation introduced below. We will not need to assume that π can be sampled by inversion. We only need inversion for an oracle used later in a coupling argument.

Let F be the CDF of $x \in \mathbb{R}$, and for $0 < u < 1$ define

$$F^{-1}(u) = \inf\{x \mid F(x) \ge u\}.$$

Take $F^{-1}(0) = \lim_{u\to 0^+} F^{-1}(u)$ and $F^{-1}(1) = \lim_{u\to 1^-} F^{-1}(u)$, using extended reals if necessary. Then $x = F^{-1}(u)$ has distribution F on $\mathbb{R}$ when $u \sim U[0,1]$.

Multidimensional inversion is based on inverting the Rosenblatt transformation [41]. Let F be the joint distribution of $\mathbf{x} \in \mathbb{R}^s$. Let F_1 be the marginal CDF of x_1 and for $j = 2, \dots, s$, let $F_j(\cdot; \mathbf{x}_{1:(j-1)})$ be the conditional CDF of x_j given $x_1, \dots, x_{j-1}$. The inverse Rosenblatt transformation ψ_R of $\mathbf{u} \in [0,1]^s$ is $\psi_R(\mathbf{u}) = \mathbf{x} \in \mathbb{R}^s$ where

$$x_1 = F_1^{-1}(u_1)$$

and

$$x_j = F_j^{-1}\big(u_j; \mathbf{x}_{1:(j-1)}\big), \qquad j \ge 1.$$

If $\mathbf{u} \sim U[0,1]^s$, then $\psi_R(\mathbf{u}) \sim F$.

We will use the inverse Rosenblatt transformation as a first step in a coupling argument which extends the one in Chentsov [7].

DEFINITION 4 (Rosenblatt–Chentsov transformation). Let ψ_R be the inverse Rosenblatt transformation for the stationary distribution π and let ϕ be the update function for MCMC. The Rosenblatt–Chentsov transformation of the finite sequence $\mathbf{u}_0, \mathbf{u}_1, \dots, \mathbf{u}_m \in [0,1]^d$ is the finite sequence $\mathbf{x}_0, \dots, \mathbf{x}_m \in \Omega \subset \mathbb{R}^s$, with $s \le d$, where $\mathbf{x}_0 = \psi_R(\mathbf{u}_{0,1:s})$ and $\mathbf{x}_i = \phi(\mathbf{x}_0, \mathbf{u}_i)$ for $i = 1, \dots, m$.

The Rosenblatt–Chentsov transformation starts off using $\mathbf{u}_0$ and inversion to generate $\mathbf{x}_0$ and then it applies whatever generators are embedded in ϕ with the innovations $\mathbf{u}_i$, to sample the transition kernel. The transition function ϕ need not be based on inversion.

3. Consistency for MCQMC sampling. In this section, we prove sufficient conditions for some deterministic MCQMC samplers to sample consistently. The same proof applies to deterministic pseudo-random sampling. First, we define consistency, then some regularity conditions, and then we give the main results.

3.1. *Definition of consistency.* Our definition of consistency is that the empirical distribution of the MCMC samples converges weakly to π.

DEFINITION 5. The triangular array $\mathbf{x}_{n,1}, \dots, \mathbf{x}_{n,n} \in \mathbb{R}^s$ for n in an infinite set $\mathbb{N}^* \subset \mathbb{N}$ consistently samples the probability density function π if

$$\lim_{\substack{n\to\infty \\ n\in\mathbb{N}^*}} \frac{1}{n} \sum_{i=1}^{n} f(\mathbf{x}_{n,i}) = \int f(\mathbf{x})\pi(\mathbf{x})\, d\mathbf{x} \tag{6}$$

holds for all bounded continuous functions $f : \Omega \to \mathbb{R}$. The infinite sequence $\mathbf{x}_1, \mathbf{x}_2, \ldots \in \mathbb{R}^s$ consistently samples π if the triangular array of initial subsequences with $\mathbf{x}_{n,i} = \mathbf{x}_i$ for $i = 1, \dots, n$ does.

In practice, we use a finite list of vectors and so the triangular array formulation is a closer description of what we do. However, to simplify the presentation and avoid giving two versions of everything, we will work only with the infinite sequence version of consistency. Triangular array versions of CUD sampling for discrete state spaces are given in [44].

It suffices to use functions f in a convergence-determining class. For example, we may suppose that f is uniformly continuous [4], or that $f = 1_{(\mathbf{a},\mathbf{b}]}$ [5]. When π is a continuous distribution, we may use $f = 1_{[\mathbf{a},\mathbf{b}]}$.

3.2. *Regularity conditions*. Here, we define some assumptions that we need to make on the MCMC update functions.

DEFINITION 6. Let $\mathcal{C} \subset [0,1]^d$ have positive Jordan measure. If $u \in \mathcal{C}$ implies that $\phi(\mathbf{x}, \mathbf{u}) = \phi(\mathbf{x}', \mathbf{u})$ for all $\mathbf{x}, \mathbf{x}' \in \Omega$, then $\mathcal{C}$ is a coupling region.

Consider two iterations $\mathbf{x}_i = \phi(\mathbf{x}_{i-1}, \mathbf{u}_i)$ and $\mathbf{x}'_i = \phi(\mathbf{x}'_{i-1}, \mathbf{u}_i)$ with the same innovations $\mathbf{u}_i$ but possibly different starting points $\mathbf{x}_0$ and $\mathbf{x}'_0$. If $\mathbf{u}_i \in \mathcal{C}$, then $\mathbf{x}_j = \mathbf{x}'_j$ holds for all $j \geq i$. In Section 4, we give some nontrivial examples of MCMC updates with coupling regions.

DEFINITION 7 (Regular MCMC). Let $\mathbf{x}_m = \mathbf{x}_m(\mathbf{u}_0, \ldots, \mathbf{u}_m)$ be the last point generated in the Rosenblatt–Chentsov transformation, viewed as a function on $[0,1]^{d(m+1)}$. The MCMC is *regular* (*for bounded continuous functions*) if the function $f(\mathbf{x}_m(\mathbf{u}_0, \ldots, \mathbf{u}_m))$ is Riemann integrable on $[0,1]^{d(m+1)}$ whenever f is bounded and continuous.

Note that if an MCMC is regular, then the definition of the Rosenblatt–Chentsov transformation implies that

$$\int_{[0,1]^{d(m+1)}} f(\mathbf{x}_m(\mathbf{u}_0, \ldots, \mathbf{u}_m))\, d\mathbf{u}_0 \cdots d\mathbf{u}_m = \int_\Omega f(\mathbf{x})\pi(\mathbf{x})\, d\mathbf{x}$$

for any $m \geq 0$ and all bounded continuous functions f.

We can, of course, define regularity for MCMC also with respect to other classes of functions. Indeed, there are numerous equivalent conditions for regularity. For example, the Portmanteau theorem ([5], Chapter 1.2) implies that it is enough to assume that the functions f are bounded and uniformly continuous. Of interest are also indicator functions of rectangles since they appear in the definition of the local discrepancy at (1). The following theorem states some equivalent conditions. To simplify the statements, we write that MCMC is *regular for indicator functions* whenever $1_A(\mathbf{x}_m(\mathbf{u}_0, \ldots, \mathbf{u}_m))$ is Riemann integrable on $[0,1]^{d(m+1)}$, where A is either $A = [\mathbf{a}, \mathbf{b}]$ with $\mathbf{a}, \mathbf{b}$ finite or $A = \Omega$.

THEOREM 1. *The following statements are equivalent*:

(i) *MCMC is regular for bounded continuous functions.*
(ii) *MCMC is regular for bounded uniformly continuous functions.*
(iii) *MCMC is regular for indicator functions* $1_{[\mathbf{a},\mathbf{b}]}$ *of rectangles* $[\mathbf{a}, \mathbf{b}]$.

PROOF. This result follows by applying the Portmanteau theorem ([5], Chapter 1.2) and some methods from real analysis. □

A regular MCMC is one that satisfies any (and hence all) of the above.

3.3. *Main results for Metropolis–Hastings.* Theorem 2 below is the main result that we will use for Metropolis–Hastings sampling. One does not expect CUD sampling to correct for an MCMC algorithm that would not be ergodic when sampled with i.i.d. inputs. Ergodicity is assured through our assumption that there is a coupling region. Section 4 below shows that some nontrivial MCMC methods have such regions. Theorem 2 does not require the detailed balance condition that Metropolis–Hastings satisfies, and so it may apply to some nonreversible chains too.

THEOREM 2. *Let* $\Omega \subseteq \mathbb{R}^s$ *and let* $\mathbf{x}_0 \in \Omega$, *and for* $i \geq 1$ *let* $\mathbf{x}_i = \phi(\mathbf{x}_{i-1}, \mathbf{u}_i)$ *where* ϕ *is the update function of a regular MCMC with a coupling region* $\mathcal{C}$. *If* $\mathbf{u}_i = (v_{d(i-1)+1}, \ldots, v_{di})$ *for a CUD sequence* $(v_i)_{i\geq 1}$, *then* $\mathbf{x}_1, \ldots, \mathbf{x}_n$ *consistently samples* π.

The proof of Theorem 2 is in the Appendix. It shows that the fraction of points $\mathbf{x}_i$ in a bounded rectangle $[\mathbf{a}, \mathbf{b}]$ converges to $\int_{[\mathbf{a},\mathbf{b}]} \pi(\mathbf{x})\, d\mathbf{x}$. Almost the same proof technique applies to expectations of bounded continuous functions.

3.4. *Main results for Gibbs sampling.* The Gibbs sampler can be viewed as a special case of Metropolis–Hastings with acceptance probability one. However, it is more straightforward to study it by applying results on iterated function mappings to (5) using methods from Diaconis and Freedman [10] and Alsmeyer and Fuh [2].

In this subsection, we assume that (Ω, d) is a complete separable metric space. We assume that the update function $\phi(\mathbf{x}, \mathbf{u})$ is jointly measurable in $\mathbf{x}$ and $\mathbf{u}$ and that it is Lipschitz continuous in $\mathbf{x}$ for any $\mathbf{u}$. Lipschitz continuity is defined through the metric $d(\cdot, \cdot)$ on Ω. The Lipschitz constant, which depends on $\mathbf{u}$, is

$$\ell(\mathbf{u}) = \sup_{\mathbf{x}\neq\mathbf{x}'} \frac{d(\phi(\mathbf{x}, \mathbf{u}), \phi(\mathbf{x}', \mathbf{u}))}{d(\mathbf{x}, \mathbf{x}')}. \tag{7}$$

For each $\mathbf{u}_n \in [0, 1]^d$, define $L_n = \ell(\mathbf{u}_n)$.

Next, we present a theorem from Alsmeyer and Fuh [2] on iterated random mappings. The n step iteration, denoted ϕ_n, is defined by $\phi_1(\mathbf{x}; \mathbf{u}_1) = \phi(\mathbf{x}, \mathbf{u}_1)$ and for $n \geq 2$: $\phi_n(\mathbf{x}; \mathbf{u}_1, \ldots, \mathbf{u}_n) = \phi(\phi_{n-1}(\mathbf{x}; \mathbf{u}_1, \ldots, \mathbf{u}_{n-1}), \mathbf{u}_n)$.

THEOREM 3. *Let the update function $\phi(\mathbf{x}, \mathbf{u})$ be jointly measurable in $\mathbf{x}$ and $\mathbf{u}$ with $\int_{[0,1]^d} \log(\ell(\mathbf{u}))\, d\mathbf{u} < 0$ and, for some $p > 0$, $\int_{[0,1]^d} \ell(\mathbf{u})^p\, d\mathbf{u} < \infty$. Assume that there is a point $\mathbf{x}' \in \Omega$ with $\int_{[0,1]^d} \log^+(d(\phi(\mathbf{x}', \mathbf{u}), \mathbf{x}'))\, d\mathbf{u} < \infty$ and $E(d(\phi(\mathbf{x}', \mathbf{u}), \mathbf{x}')^p) < \infty$. Then there is a $\gamma^* \in (0, 1)$ such that for all $\gamma \in (\gamma^*, 1)$ there is a $\alpha_\gamma \in (0, 1)$ such that for every $\mathbf{x}, \widehat{\mathbf{x}} \in \Omega$*

$$\lim_{m\to\infty} \alpha_\gamma^{-m} \Pr\bigl(d(\phi_m(\mathbf{x}; \cdot), \phi_m(\widehat{\mathbf{x}}; \cdot)) > \gamma^m\bigr) = 0. \tag{8}$$

PROOF. This follows by specializing Corollary 2.5(a) of [2] to the present setting. □

THEOREM 4. *Let (Ω, d) be a complete separable metric space and let $(v_i)_{i\ge1}$ be a CUD sequence such that for every sequence $(d_n)_{n\ge1}$ of natural numbers with $d_n = O(\log n)$, we have $\lim_{n\to\infty} D_n^{*d_n} = 0$. Let $\mathbf{x}_0 \in \Omega$, and for $i \ge 1$ let $\mathbf{x}_i = \phi(\mathbf{x}_{i-1}, \mathbf{u}_i)$ be the Gibbs sampler update for stationary distribution π. Assume that ϕ satisfies the conditions of Theorem 3 and that there is a $\gamma \in (\gamma^*, 1)$ such that*

$$\mathcal{B}_m(\mathbf{x}, \widehat{\mathbf{x}}) = \{\mathbf{v} \in [0, 1]^{dm} : d(\phi_m(\mathbf{x}, \mathbf{v}), \phi_m(\widehat{\mathbf{x}}, \mathbf{v})) > \gamma^m\}$$

is Jordan measurable for all $m \ge 1$ and $\mathbf{x}, \widehat{\mathbf{x}} \in \Omega$. Under these conditions, if the Gibbs sampler is regular, then $\mathbf{x}_1, \dots, \mathbf{x}_n$ consistently samples π.

The proof of Theorem 4 is in the Appendix. Like Theorem 2, it shows that bounded rectangles $[\mathbf{a}, \mathbf{b}]$ have asymptotically the correct proportion of points. Once again, similar arguments apply for bounded continuous functions of $\mathbf{x}$.

Although not explicitly stated there, the proof of [11], Theorem 1, shows the existence of sequences $(v_i)_{i\ge1}$ for which

$$D_n^{*d}\bigl(\{(v_{d(i-1)+1}, \dots, v_{di}), i = 1, \dots, n\}\bigr) \le C\sqrt{\frac{d\log(n+1)}{n}},$$

for all $n, d \in \mathbb{N}$, where $C > 0$ is a constant independent of n and d. Unfortunately, no explicit construction of such a sequence is given in [11]. Then for any sequence $(d_n)_{n\ge1}$ of natural numbers with $d_n = O(\log n)$ we obtain that

$$D_n^{*d_n}\bigl(\{(v_{d_n(i-1)+1}, \dots, v_{d_n i}), i = 1, \dots, n\}\bigr) \le C'\frac{\log(n+1)}{\sqrt{n}} \to 0 \qquad \text{as } n \to \infty.$$

In Theorem 2, we assumed that the coupling region $\mathcal{C}$ is Jordan measurable In Theorem 4, we do not have a coupling region, but still have an analogous assumption, namely that the sets $\mathcal{B}_m(\mathbf{x}, \widehat{\mathbf{x}})$ are Jordan measurable. A condition on ϕ which guarantees that $\mathcal{B}_m(\mathbf{x}, \widehat{\mathbf{x}})$ is Jordan measurable is given in Section 6.

4. Examples of coupling regions. Theorem 2 used coupling regions. These are somewhat special. But they do exist for some realistic MCMC algorithms.

LEMMA 1. *Let ϕ be the update for the Metropolized independence sampler on $\Omega \subseteq \mathbb{R}^s$ obtaining the proposal $\mathbf{y} = \psi(\mathbf{u}_{1:(d-1)})$, where ψ generates samples from the density p, which are accepted when*

$$u_d \le \frac{\pi(\mathbf{y})\,p(\mathbf{x})}{\pi(\mathbf{x})\,p(\mathbf{y})}.$$

Assume that the importance ratio is bounded above, that is,

$$\kappa \equiv \sup_{\mathbf{x}\in\Omega} \frac{\pi(\mathbf{x})}{p(\mathbf{x})} < \infty.$$

Suppose also that there is a rectangle $[\mathbf{a}, \mathbf{b}] \subset [0,1]^{d-1}$ of positive volume with

$$\eta \equiv \inf_{\mathbf{u}\in[\mathbf{a},\mathbf{b}]} \frac{\pi(\psi(\mathbf{u}))}{p(\psi(\mathbf{u}))} > 0.$$

Then $\mathcal{C} = [\mathbf{a}, \mathbf{b}] \times [0, \eta/\kappa]$ is a coupling region.

PROOF. The set $\mathcal{C}$ has positive Jordan measure. Suppose that $\mathbf{u} \in \mathcal{C}$. Then

$$\pi(\mathbf{y})\,p(\mathbf{x}) \ge \eta p(\mathbf{y}) \frac{1}{\kappa}\pi(\mathbf{x}) \ge u_d\,p(\mathbf{y})\pi(\mathbf{x}),$$

and so $\phi(\mathbf{x}, \mathbf{u}) = \mathbf{y}$, regardless of $\mathbf{x}$. □

LEMMA 2. *Let π be a density on a bounded rectangular region $\Omega = [\mathbf{a}, \mathbf{b}] \subset \mathbb{R}^s$. Assume that $0 < \eta \le \pi(\mathbf{x}) \le \kappa < \infty$ holds for all $\mathbf{x} \in \Omega$. Let $\Omega' = \{(y, \mathbf{x}) \mid 0 \le y \le \pi(\mathbf{x})\} \subset [\mathbf{a}, \mathbf{b}] \times [0, \kappa]$ be the domain of the inversive slice sampler. Let $(y_i, \mathbf{x}_i) = \phi((y_{i-1}, \mathbf{x}_{i-1}), \mathbf{u}_i)$ for $\mathbf{u}_i \in [0,1]^{s+1}$ be the update for the inversive slice sampler and put $(y'_i, \mathbf{x}'_i) = \phi((y'_{i-1}, \mathbf{x}'_{i-1}), \mathbf{u}_i)$. If $\mathbf{u}_i \in \mathcal{C} = [0, \eta/\kappa] \times [0,1]^s$, then $\mathbf{x}_i = \mathbf{x}'_i$.*

PROOF. If $u_{i,1} \le \eta/\kappa$, then $y_i = u_{i1}\pi(\mathbf{x}_{i-1})$ and $y'_i = u_{i1}\pi(\mathbf{x}'_{i-1})$ are in the set $[0, \eta/\kappa]$. The distribution of $\mathbf{x}$ given y for any $y \in [0, \eta/\kappa]$ is $U[\mathbf{a}, \mathbf{b}]$. Therefore, $\mathbf{x}_i = \mathbf{x}'_i = \mathbf{a} + u_{2:(s+1)}(\mathbf{b} - \mathbf{a})$ (componentwise). □

Lemma 2 does not couple the chains because y_i and y'_i are different in general. But because $\mathbf{x}_i = \mathbf{x}'_i$, a coupling will happen at the next step, that is, $(y_{i+1}, \mathbf{x}_{i+1}) = (y'_{i+1}, \mathbf{x}'_{i+1})$ when $\mathbf{u}_i \in [0, \eta/\kappa] \times [0,1]^s$. One could revise Theorem 2 to include couplings that happen within some number t of steps after $\mathbf{u} \in \mathcal{C}$ happens. In this case, it is simpler to say that the chain whose update comprises two iterations of the inversive slice sampler satisfies Theorem 2. For a chain whose update is just one iteration, the averages over odd and even numbered iterations both converge

properly and so that chain is also consistent. Alternatively, we could modify the space of y values so that all $y \in [0, \eta/\kappa]$ are identified as one point. Then $\mathcal{C}$ is a coupling region.

The result of Lemma 2 also applies to slice samplers that sample $y \mid \mathbf{x}$ and then $\mathbf{x} \mid y \sim U\{\mathbf{x} \mid \pi(\mathbf{x}) \le y\}$ using an s-dimensional generator that is not necessarily inversion.

5. Riemann integrability. Theorem 2 proves that MCMC consistently samples π when implemented using CUD sequences. We required certain Riemann integrability conditions in defining regular Rosenblatt–Chentsov transformations. Here, we verify that nontrivial MCMC algorithms can have regular Rosenblatt–Chentsov transformations.

It seems odd to use the Riemann integral over 100 years after Lebesgue [23]. But pseudo-random number generators are now typically designed to meet an equidistribution criterion over rectangular regions [29]. Other times they are designed with a spectral condition in mind. This again is closely related to Riemann integrability via the Weyl [45] condition where $\hat{\theta}_n(f) \to \theta(f)$ for all trigonometric polynomials $f(\mathbf{x}) = e^{2\pi\sqrt{-1}k'\mathbf{x}}$ if and only if $\mathbf{x}_1, \dots, \mathbf{x}_n$ are uniformly distributed. Unless one is using physical random numbers, the Riemann integral, or perhaps the improper Riemann integral is almost implicit.

5.1. *Definitions and basic theorems.* A function from $A \subset \mathbb{R}^d$ to $\mathbb{R}^s$ for $s \ge 1$ is Riemann integrable if all of its s components are. To study how Riemann integrability propagates, we will use the following two definitions.

DEFINITION 8. For a function $f : \mathbb{R}^k \to \mathbb{R}$, the discontinuity set of f is

$$D(f) = \{\mathbf{x} \in \mathbb{R}^k \mid f \text{ discontinuous at } \mathbf{x}\}.$$

If f is only defined on $A \subset \mathbb{R}^k$, then $D(f) = D(f_0)$ where $f_0(\mathbf{x}) = f(\mathbf{x})$ for $\mathbf{x} \in A$ and $f_0(\mathbf{x}) = 0$ for $\mathbf{x} \notin A$.

DEFINITION 9. For a function $f : \mathbb{R}^k \to \mathbb{R}$, the graph of f is

$$G(f) = \{(\mathbf{x}, y) \in \mathbb{R}^{k+1} \mid y = f(\mathbf{x})\}.$$

Lebesgue's theorem, next, provides a checkable characterization of Riemann integrability.

THEOREM 5 (Lebesgue's theorem). *Let $A \subset \mathbb{R}^d$ be bounded and let $f : A \to \mathbb{R}$ be a bounded function. Then f is Riemann integrable iff $\lambda_d(D(f)) = 0$.*

PROOF. See Marsden and Hoffman [30], page 455. □

5.2. *Need for Riemann integrable proposals*. Here, we show that Riemann integrability adds a special requirement to the way an algorithm is implemented. Then we give an example to show that propagation rules for Riemann integrability are more complicated than are those for continuity and differentiability.

Suppose that F is the $\mathcal{N}(\binom{0}{0}, \left(\begin{smallmatrix} 1 & \rho \\ \rho & 1 \end{smallmatrix}\right))$ distribution for some $\rho \in (-1, 1)$. If we take

$$x_1(\mathbf{u}) = \Phi^{-1}(u_1)$$

and

$$x_2(\mathbf{u}) = \rho x_1(\mathbf{u}) + \sqrt{1-\rho^2}\Phi^{-1}(u_2),$$

then we find that $f(\mathbf{u}) = 1_{a_1 \le x_1(\mathbf{u}) \le b_1} \times 1_{a_2 \le x_2(\mathbf{u}) \le b_2}$ is discontinuous only on a set of measure zero. It is trivially bounded, and these two facts imply it is Riemann integrable on $[0, 1]^2$.

Another transformation for the same distribution F is

$$x_1 = \Phi^{-1}(u_1)$$

and

$$x_2 = \begin{cases} \rho x_1(\mathbf{u}) + \sqrt{1-\rho^2}\Phi^{-1}(u_2), & u_1 \notin \mathbb{Q}, \\ -\rho x_1(\mathbf{u}) - \sqrt{1-\rho^2}\Phi^{-1}(u_2), & u_1 \in \mathbb{Q}. \end{cases}$$

Changing the conditional distribution of x_2 given x_1 on a set of measure 0 leaves the distribution F of $\mathbf{x}$ unchanged. But for this version, we find f can be discontinuous on more than a set of measure 0 and so this inverse Rosenblatt transformation of F is not regular.

In practice, of course, one would use the regular version of the transformation. But propagating Riemann integrability to a function built up from several other functions is not always straightforward. The core of the problem is that the composition of two Riemann integrable functions need not be Riemann integrable.

As an example [18], consider Thomae's function on $(0, 1)$,

$$f(x) = \begin{cases} 1/q, & x = p/q \in \mathbb{Q}, \\ 0, & \text{else}, \end{cases}$$

where it is assumed that p and q in the representation p/q have no common factors. This f is continuous except on $\mathbb{Q} \cap (0, 1)$ and so it is Riemann integrable. The function $g(x) = 1_{0<x\le 1}$ is also Riemann integrable. But $g(f(x)) = 1_{x\in\mathbb{Q}}$ for $x \in (0, 1)$, which is famously not Riemann integrable. The class of Riemann integrable functions, while more restrictive than we might like for conclusions, is also too broad to use in propagation rules.

5.3. *Specializing to MCMC.* First, we show that the acceptance-rejection step in Metropolis–Hastings does not cause problems with Riemann integrability.

LEMMA 3. *Let $k \in \mathbb{N}$ and suppose that g, h and A are real-valued Riemann integrable functions on $[0,1]^k$. For $\mathbf{u} \in [0,1]^{k+1}$ define*

$$f(\mathbf{u}) = \begin{cases} g(\mathbf{u}_{1:k}), & u_{k+1} \le A(\mathbf{u}_{1:k}), \\ h(\mathbf{u}_{1:k}), & \text{else.} \end{cases}$$

Then f is Riemann integrable on $[0,1]^{k+1}$.

PROOF. First, $D(f) \subset ((D(g) \cup D(h)) \times [0,1]) \cup G(A)$. Riemann integrability of g gives $\lambda_k(D(g)) = 0$. Similarly, $\lambda_k(D(h)) = 0$. Therefore, $\lambda_{k+1}(D(g) \cup D(h)) \times [0,1]) = 0$.

Turning to $G(A)$, we split the domain $[0,1]^k$ of A into n^k congruent subcubes $C_{n,1}, \dots, C_{n,n^k}$ (whose boundaries overlap). Then $G(A) \subseteq \bigcup_{i=1}^{n^k} C_{n,i} \times [m_{i,n}, M_{i,n}]$, where $m_{i,n} = \inf_{\mathbf{u}_{1:k} \in C_{n,i}} A(\mathbf{u}_{1:k})$ and $M_{i,n} = \sup_{\mathbf{u}_{1:k} \in C_{n,i}} A(\mathbf{u}_{1:k})$. As a result $\lambda_{k+1}(G(h)) \le n^{-k} \sum_i (M_{i,n} - m_{i,n})$. Riemann integrability of A implies this upper bound vanishes as $n \to \infty$. Therefore, $\lambda_{k+1}(G(A)) = 0$ and so $\lambda_{k+1}(D(f)) = 0$ and the result follows by Lebesgue's theorem. □

In the MCMC context, g and h are the jth component of the proposal and the previous state, respectively, A is the acceptance probability, and $\mathbf{u}$ is the ensemble of uniform random variables used in m stage Rosenblatt–Chentsov coupling and $k = (m+1)d - 1$.

For consistency results, we study the proportion of times $f(\mathbf{u}) \in [\mathbf{a}, \mathbf{b}]$. It is enough to consider the components one at a time and in turn to show $1_{f_j(\mathbf{u}) \le b_j}$ and $1_{f_j(\mathbf{u}) < a_j}$ are Riemann integrable. However, as the example with Thomae's function shows, even the indicator function of an interval applied to a Riemann integrable function can give a non-Riemann integrable composite function.

We may avoid truncation by employing bounded continuous test functions. We will use the following simple corollary of Lebesgue's theorem.

LEMMA 4. *For $k \ge 1$ and $r \ge 1$, let $g_1, \dots, g_r$ be Riemann integrable functions from $[0,1]^k$ to a bounded interval $[a,b] \subset \mathbb{R}$. Let h be a continuous function from $[a,b]^k$ to $\mathbb{R}$. Then*

$$f(\mathbf{u}) = h(g_1(\mathbf{u}), \dots, g_k(\mathbf{u}))$$

is Riemann integrable on $[0,1]^k$.

PROOF. Because h is continuous, $D(f) \subset \bigcup_{j=1}^{r} D(g_k)$. But $\lambda_k(D(g_k)) = 0$. Therefore, $\lambda_k(D(f)) = 0$ and so f is Riemann integrable by Lebesgue's theorem. □

We can also propagate Riemann integrability through monotonicity. If g is a monotone function from $\mathbb{R}$ to $\mathbb{R}$ and f is the indicator of an interval, then $f \circ g$ is the indicator of an interval too, and hence is Riemann integrable, when that interval is of finite length.

LEMMA 5. *Let $F_1(x_1)$ be the CDF of x_1 and for $j = 2, \dots, s$, let $F_j(x_j \mid \mathbf{x}_{1:(j-1)})$ be the conditional CDF of x_j given $\mathbf{x}_{1:(j-1)}$. Suppose that the CDFs $F_j(x_j \mid \mathbf{x}_{1:(j-1)})$ are continuous functions of $\mathbf{x}_{1:j}$ and that the quantile functions $F_j^{-1}(u_j \mid \mathbf{x}_{1:(j-1)})$ are continuous in $(u_j, \mathbf{x}_{1:(j-1)}) \in [0,1] \times \mathbb{R}^{j-1}$, for $j = 2, \dots, s$. Define functions $z_1(\mathbf{u}) = F_1^{-1}(u_1)$ and $z_j(\mathbf{u}) = F_j^{-1}(u_j \mid \mathbf{z}_{1:(j-1)}(\mathbf{u}))$ for $j = 2, \dots, s$, where $\mathbf{z}_{1:(j-1)} = (z_1, \dots, z_{j-1})$. Then for $\mathbf{b} \in \mathbb{R}^s$, the set*

$$S(\mathbf{b}) = \{\mathbf{u} \mid z_j(\mathbf{u}) \le b_j, 1 \le j \le s\}$$

is Jordan measurable.

PROOF. By hypothesis, z_k is a continuous function of $\mathbf{u} \in [0,1]^s$, for $k = 1, \dots, s$, and so is $F_k(b_k \mid \mathbf{z}_{1:(k-1)}(\mathbf{u}))$. This latter only depends on $\mathbf{u}_{1:(k-1)}$, for $k = 2, \dots, s$, and so we write it as $g_k(\mathbf{u}_{1:(k-1)})$.

For $k = 1, \dots, s$, let $S_k = \{\mathbf{u}_{1:k} \mid u_j \le g_j(\mathbf{u}_{1:(j-1)}) \text{ for } j = 1, \dots, k\}$. The set S_1 is the interval $[0, F_1^{-1}(b_1)]$, and hence is Jordan measurable. Suppose S_k is Jordan measurable for $k < s$. Then

$$S_{k+1} = (S_k \times [0,1]) \cap G_{k+1} \qquad \text{where } G_{k+1} = \{\mathbf{u}_{1:(k+1)} \mid u_{k+1} \le g_{k+1}(\mathbf{u}_{1:k})\}.$$

The set $S_k \times [0,1]$ is Jordan measurable because S_k is. The boundary of G_{k+1} is contained within the intersection of the graph of g_{k+1} and the boundary of $[0,1]^{k+1}$ and so G_{k+1} is Jordan measurable. The result follows by induction because $S(\mathbf{b}) = S_s$. □

5.4. *Regularity of Rosenblatt–Chentsov.* Here, we give sufficient conditions for the Rosenblatt–Chentsov transformation to be regular.

THEOREM 6. *For integer $m \ge 0$, let $\mathbf{x}_m$ be the endpoint of the Rosenblatt–Chentsov transformation of $[0,1]^{(d+1)m}$, started with a Riemann integrable function ψ_R and continued via the Metropolis–Hastings update ϕ. Let ϕ be defined in terms of the proposal function $\mathbf{y} : \mathbb{R}^s \times [0,1]^{d-1} \to \mathbb{R}^s$ with proposal density $p(\cdot,\cdot) : \mathbb{R}^s \times \mathbb{R}^s \to [0,\infty)$ and target density $\pi : \mathbb{R}^s \to [0,\infty)$. Let f be a bounded continuous function on $\mathbb{R}^s$.*

If ψ is bounded and $\mathbf{y}$, P and π are bounded continuous functions, then $f(\mathbf{x}_m(\mathbf{u}_0, \dots, \mathbf{u}_m))$ is a Riemann integrable function of the variables $[0,1]^{(d+1)m}$ used in the Rosenblatt–Chentsov transformation.

PROOF. We only need to show that $\mathbf{x}_m$ is a Riemann integrable function of $(\mathbf{u}_0, \dots, \mathbf{u}_m) \in [0,1]^{d(m+1)}$ and then the result follows by Lemma 4.

We proceed by induction. For $m = 0$, $\mathbf{x}_0 = \psi(\mathbf{u}_0)$ is bounded and continuous on $[0,1]^d$, hence it is Riemann integrable.

Now suppose that $\mathbf{x}_{m-1}$ is a Riemann integrable function on $[0,1]^{dm}$. Let $h(\mathbf{u}_0, \dots, \mathbf{u}_{m-1}, \mathbf{u}_{m1:(d-1)})$ be the value $\mathbf{x}_{m-1}$, written as a Riemann integrable function on $[0,1]^{dm+d-1}$, so it ignores its last $d-1$ arguments. Let $g(\mathbf{u}_0, \dots, \mathbf{u}_{m-1}, \mathbf{u}_{m1:(d-1)})$ be the proposal $\mathbf{y}_m = \mathbf{y}(\mathbf{x}_{m-1}, \mathbf{u}_{m1:(d-1)}) = \mathbf{y}(g(\cdot), \mathbf{x}_{m-1}, \mathbf{u}_{m1:(d-1)})$. This is a continuous function $\mathbf{y}(\cdot,\cdot)$ of two Riemann integrable functions on $[0,1]^{d(m+1)-1}$ and so it is Riemann integrable. Next, $A(\cdot,\cdot)$ is a continuous function of both $\mathbf{x}_{m-1}$ and $\mathbf{y}_m$ which are in turn Riemann integrable functions on $[0,1]^{dm+d-1}$, and so $A(\cdot,\cdot)$ is Riemann integrable. Then $\mathbf{x}_m$ is a Riemann integrable function on $[0,1]^{dm+d}$, by Lemma 3, completing the induction. □

6. Conditions for the Gibbs sampler. In studying the Gibbs sampler, we made several assumptions. First, we required Jordan measurability for the sets $\mathcal{B}_m(\mathbf{x}, \widehat{\mathbf{x}})$. Second, we required a contraction property. In this section, we show that those assumptions are reasonable.

6.1. *Jordan measurability of* $\mathcal{B}_m(\mathbf{x}, \widehat{\mathbf{x}})$. We give an example where the conditions of Theorem 4 are satisfied, that is, the sets $\mathcal{B}_m(\mathbf{x}, \widehat{\mathbf{x}})$ are Jordan measurable for all $m \geq 1$ and $\mathbf{x}, \widehat{\mathbf{x}} \in \Omega$ (for some suitable domain $\Omega \subset \mathbb{R}^s$). Assume (additionally to the assumptions made in Theorem 4) that $\phi(\mathbf{x}, \mathbf{u})$ is totally differentiable with continuous derivative with respect to $\mathbf{u}$ for each $\mathbf{x} \in \Omega$ and that d is based on the L_p norm for some $1 \leq p < \infty$. Further, assume that the gradient of $d(\phi(\mathbf{x}, \mathbf{u}), \phi(\widehat{\mathbf{x}}, \mathbf{u}))$ with respect to $\mathbf{u}$ vanishes only on a null set for all $\mathbf{x}, \widehat{\mathbf{x}} \in \Omega$, $\mathbf{x} \neq \widehat{\mathbf{x}}$, that is,

$$\lambda\big(\{\mathbf{u} \in [0,1]^d : \nabla_{\mathbf{u}}\, d(\phi(\mathbf{x}, \mathbf{u}), \phi(\widehat{\mathbf{x}}, \mathbf{u})) = \mathbf{0}\}\big) = 0,$$

for all $\mathbf{x}, \widehat{\mathbf{x}} \in \Omega$, $\mathbf{x} \neq \widehat{\mathbf{x}}$, where λ denotes the Lebesgue measure and where $\nabla_{\mathbf{u}}\, d(\phi(\mathbf{x}, \mathbf{u}), \phi(\widehat{\mathbf{x}}, \mathbf{u})) = (\frac{\partial}{\partial u_j}\, d(\phi(\mathbf{x}, \mathbf{u}), \phi(\widehat{\mathbf{x}}, \mathbf{u})))_{j=1,\dots,d}$ denotes the gradient.

Then, for all $m \geq 1$, we also have

$$\lambda\big(\{\mathbf{u} \in [0,1]^{dm} : \nabla_{\mathbf{u}}\, d(\phi_m(\mathbf{x}, \mathbf{u}), \phi_m(\widehat{\mathbf{x}}, \mathbf{u})) = \mathbf{0}\}\big) = 0$$

for all $\mathbf{x}, \widehat{\mathbf{x}} \in \Omega$, $\mathbf{x} \neq \widehat{\mathbf{x}}$. Let $\mathbf{x}, \widehat{\mathbf{x}} \in \Omega$ with $\mathbf{x} \neq \widehat{\mathbf{x}}$ be fixed. Then for almost all $\mathbf{u}^* \in [0,1]^{dm}$ we have $\nabla_{\mathbf{u}}\, d(\phi_m(\mathbf{x}, \mathbf{u}^*), \phi_m(\widehat{\mathbf{x}}, \mathbf{u}^*)) \neq \mathbf{0}$. Therefore, there is a $\delta > 0$ such that $\nabla_{\mathbf{u}} d(\phi_m(\mathbf{x}, \mathbf{u}), \phi_m(\widehat{\mathbf{x}}, \mathbf{u})) \neq \mathbf{0}$ for all $\mathbf{u} \in N_\delta(\mathbf{u}^*)$, where $N_\delta(\mathbf{u}^*) = \{\mathbf{v} \in [0,1]^{dm} : \|\mathbf{u}^* - \mathbf{v}\|_{L_2} < \delta\}$ is a neighborhood of $\mathbf{u}^*$. Therefore, the directional derivative at a point $\mathbf{u} \in N_\delta(\mathbf{u}^*)$ is different from 0, except on a hyperplane, that is, almost everywhere. Hence, by the mean value theorem, the function $d(\phi_m(\mathbf{x}, \mathbf{u}), \phi_m(\widehat{\mathbf{x}}, \mathbf{u}))$ for $\mathbf{u} \in N_\delta(\mathbf{u}^*)$ can at most be constant on a hyperplane, which has Lebesgue measure 0. Note that $N_\delta(\mathbf{u}^*) \cap \mathbb{Q}^{dm} \neq \varnothing$, therefore there is a

countable number of elements $\mathbf{u}_1^*, \mathbf{u}_2^*, \ldots$ and numbers $\delta_1, \delta_2, \ldots$ with the properties of $\mathbf{u}^*$ and δ described above and for which we have $\bigcup_{n=1}^{\infty} N_{\delta_n}(\mathbf{u}_n^*) = [0,1]^{dm}$. Therefore, we have

$$\lambda\big(\{\mathbf{u} \in [0,1]^{dm} : d(\phi_m(\mathbf{x}, \mathbf{u}), \phi_m(\widehat{\mathbf{x}}, \mathbf{u})) = c\}\big) = 0,$$

for any $c > 0$.

The set of points where $1_{\mathcal{B}_m(\mathbf{x}, \widehat{\mathbf{x}})}$ is discontinuous is given by

$$D = \{\mathbf{u} \in [0,1]^{dm} : \forall \delta > 0\ \exists \mathbf{v}, \mathbf{v}' \in N_\delta(\mathbf{u}) \text{ such that } d(\phi_m(\mathbf{x}, \mathbf{v}), \phi_m(\widehat{\mathbf{x}}, \mathbf{v})) > \gamma^m \text{ and } d(\phi_m(\mathbf{x}, \mathbf{v}'), \phi_m(\widehat{\mathbf{x}}, \mathbf{v}')) \le \gamma^m\}.$$

As $\mathcal{B}_m(\mathbf{x}, \widehat{\mathbf{x}})$ and $\{\mathbf{u} \in [0,1]^{dm} : d(\phi_m(\mathbf{x}, \mathbf{u}), \phi_m(\widehat{\mathbf{x}}, \mathbf{u})) < \gamma^m\}$ are open, it follows that

$$D \subseteq \{\mathbf{u} \in [0,1]^{dm} : d(\phi_m(\mathbf{x}, \mathbf{u}), \phi_m(\widehat{\mathbf{x}}, \mathbf{u})) = \gamma^m\}.$$

Therefore, $\lambda_{dm}(D) = 0$ and Lebesgue's theorem (see Theorem 5) implies that $\mathcal{B}_m(\mathbf{x}, \widehat{\mathbf{x}})$ is Jordan measurable.

6.2. *Contraction*. Here, we illustrate how the Gibbs sampler yields a contraction for the probit model. In this model,

$$Z_i = \mathbf{x}_i^{\mathsf{T}} \beta + \epsilon_i$$

and

$$Y_i = \mathbf{1}_{Z_i > 0},$$

for $i = 1, \ldots, n$ for independent $\epsilon_i \sim \mathcal{N}(0,1)$. The coefficient $\beta \in \mathbb{R}^p$ has a noninformative prior distribution. The predictors are $\mathbf{x}_i \in \mathbb{R}^p$. We define the matrix X with ij element x_{ij}. We assume that X has rank p.

The state of the Markov chain is $(\beta, \mathbf{Z}) \in \Omega \subset \mathbb{R}^{p+n}$, where $\mathbf{Z} = (Z_1, \ldots, Z_n)^{\mathsf{T}}$. Given the observed data $(y_1, \ldots, y_n, \mathbf{x}_1, \ldots, \mathbf{x}_n)$, we can use the Gibbs sampler to simulate the posterior distribution of β and $\mathbf{Z} = (Z_1, \ldots, Z_n)^{\mathsf{T}}$. A single step of the Gibbs sampler makes the transition

$$\begin{pmatrix} \beta^{(k-1)} \\ \mathbf{Z}^{(k-1)} \end{pmatrix} \overset{u_1, \ldots, u_n}{\longrightarrow} \begin{pmatrix} \beta^{(k-1)} \\ \mathbf{Z}^{(k)} \end{pmatrix} \overset{u_{n+1}, \ldots, u_{n+p}}{\longrightarrow} \begin{pmatrix} \beta^{(k)} \\ \mathbf{Z}^{(k)} \end{pmatrix}$$

for $k \ge 1$ using generators given explicitly below. The values $u_1, \ldots, u_{n+p}$ are the components of $\mathbf{u}_k \in (0,1)^{n+p}$. We also write the transitions as

$$(\beta, \mathbf{Z}) \to \phi((\beta, \mathbf{Z}), \mathbf{u}) = \big(\phi^{(1)}((\beta, \mathbf{Z}), \mathbf{u}), \phi^{(2)}((\beta, \mathbf{Z}), \mathbf{u})\big),$$

where ϕ and its components $\phi^{(1)}$ and $\phi^{(2)}$ (for β and $\mathbf{Z}$, resp.) are given explicitly below.

Given β, the components of $\mathbf{Z}$ are independent, with

$$Z_i \sim \begin{cases} \mathcal{N}(\mathbf{x}_i^\mathsf{T}\beta, 1)|Z_i > 0, & \text{if } Y_i = 1, \\ \mathcal{N}(\mathbf{x}_i^\mathsf{T}\beta, 1)|Z_i \le 0, & \text{if } Y_i = 0. \end{cases}$$

We may generate them from $u_1, \dots, u_n \in (0,1)$ by

$$Z_i = \begin{cases} \mathbf{x}_i^\mathsf{T}\beta + \Phi^{-1}\big(\Phi(-\mathbf{x}_i^\mathsf{T}\beta) + u_i\Phi(\mathbf{x}_i^\mathsf{T}\beta)\big), & \text{if } Y_i = 1, \\ \mathbf{x}_i^\mathsf{T}\beta + \Phi^{-1}(u_i\Phi(-\mathbf{x}_i^\mathsf{T}\beta)), & \text{if } Y_i = 0. \end{cases} \tag{9}$$

Given $\mathbf{Z}$, the distribution of β is $\beta \sim \mathcal{N}((X^\mathsf{T}X)^{-1}X^\mathsf{T}\mathbf{Z}, (X^\mathsf{T}X)^{-1})$. We may generate it using $u_{n+1}, \dots, u_{n+p} \in (0,1)$ via

$$\beta = (X^\mathsf{T}X)^{-1}X^\mathsf{T}\mathbf{Z} + (X^\mathsf{T}X)^{-1/2}\begin{pmatrix} \Phi^{-1}(u_{n+1}) \\ \vdots \\ \Phi^{-1}(u_{n+p}) \end{pmatrix}. \tag{10}$$

Thus equation (10) defines $\phi^{(1)}$ while (9) defines $\phi^{(2)}$.

The framework in [2] allows one to pick a metric that conforms to the problem. We use the metric $d((\beta, \mathbf{Z}), (\beta', \mathbf{Z}')) = \max(d_1(\beta, \beta'), d_2(\mathbf{Z}, \mathbf{Z}'))$, where

$$d_1(\beta, \beta') = d_1(\beta - \beta') = \sqrt{(\beta - \beta')^\mathsf{T}(X^\mathsf{T}X)(\beta - \beta')} \tag{11}$$

and

$$d_2(\mathbf{Z}, \mathbf{Z}') = d_2(\mathbf{Z} - \mathbf{Z}') = \sqrt{(\mathbf{Z} - \mathbf{Z}')^\mathsf{T}(\mathbf{Z} - \mathbf{Z}')}. \tag{12}$$

We show below that

$$d\big((\beta^{(k)}, \mathbf{Z}^{(k)}), (\beta'^{(k)}, \mathbf{Z}'^{(k)})\big) \le d\big((\beta^{(k-1)}, \mathbf{Z}^{(k-1)}), (\beta'^{(k-1)}, \mathbf{Z}'^{(k-1)})\big) \tag{13}$$

for pairs $(\beta^{(k-1)}, \mathbf{Z}^{(k-1)}), (\beta'^{(k-1)}, \mathbf{Z}'^{(k-1)})$ of distinct points in Ω. Both metrics d_1 and d_2 are also norms, which simplifies our task.

Suppose first that $\beta^{(k-1)} = \beta'^{(k-1)}$. Then it follows easily that $\mathbf{Z}^{(k)} = \mathbf{Z}'^{(k)}$ and $\beta^{(k)} = \beta'^{(k)}$, so then the left-hand side of (13) is 0. As a result, we may assume without loss of generality that $d_1(\beta^{(k-1)} - \beta'^{(k-1)}) > 0$. With this assumption, we will use the bound

$$\begin{aligned} &\frac{d((\beta^{(k)}, \mathbf{Z}^{(k)}), (\beta'^{(k)}, \mathbf{Z}'^{(k)}))}{d((\beta^{(k-1)}, \mathbf{Z}^{(k-1)}), (\beta'^{(k-1)}, \mathbf{Z}'^{(k-1)}))} \\ &\qquad \le \max\left(\frac{d_1(\beta^{(k)} - \beta'^{(k)})}{d_1(\beta^{(k-1)} - \beta'^{(k-1)})}, \frac{d_2(\mathbf{Z}^{(k)} - \mathbf{Z}'^{(k)})}{d_1(\beta^{(k-1)} - \beta'^{(k-1)})}\right). \end{aligned} \tag{14}$$

We begin by studying the update to $\mathbf{Z}$. Subtracting $\mathbf{x}_i^\mathsf{T}\beta$ from both sides of (9), applying $\Phi(\cdot)$, differentiating with respect to β and gathering up terms, we find

that $\frac{\partial}{\partial\beta}Z_i = \lambda_i \mathbf{x}_i$ where

$$\lambda_i = \begin{cases} 1 - \dfrac{(1-u_i)\varphi(\mathbf{x}_i^\mathsf{T}\beta)}{\varphi(Z_i - \mathbf{x}_i^\mathsf{T}\beta)}, & \text{if } Y_i = 1, \\ 1 - \dfrac{u_i\varphi(-\mathbf{x}_i^\mathsf{T}\beta)}{\varphi(Z_i - \mathbf{x}_i^\mathsf{T}\beta)}, & \text{if } Y_i = 0, \end{cases} \tag{15}$$

and φ is the $\mathcal{N}(0,1)$ probability density function.

It is clear that $\lambda_i < 1$. Next, we show that $\lambda_i \geq 0$. We begin by inverting (9) to get

$$u_i = \begin{cases} \dfrac{\Phi(Z_i - \mathbf{x}_i^\mathsf{T}\beta) - \Phi(-\mathbf{x}_i^\mathsf{T}\beta)}{\Phi(\mathbf{x}_i^\mathsf{T}\beta)}, & \text{if } Y_i = 1, \\ \dfrac{\Phi(Z_i - \mathbf{x}_i^\mathsf{T}\beta)}{\Phi(-\mathbf{x}_i^\mathsf{T}\beta)}, & \text{if } Y_i = 0. \end{cases} \tag{16}$$

Substituting (16) into (15) and simplifying yields

$$1 - \lambda_i = \begin{cases} \dfrac{\varphi(\mathbf{x}_i^\mathsf{T}\beta)\Phi(-Z_i + \mathbf{x}_i^\mathsf{T}\beta)}{\Phi(\mathbf{x}_i^\mathsf{T}\beta)\varphi(Z_i - \mathbf{x}_i^\mathsf{T}\beta)}, & \text{if } Y_i = 1, \\ \dfrac{\varphi(-\mathbf{x}_i^\mathsf{T}\beta)\Phi(Z_i - \mathbf{x}_i^\mathsf{T}\beta)}{\Phi(-\mathbf{x}_i^\mathsf{T}\beta)\varphi(Z_i - \mathbf{x}_i^\mathsf{T}\beta)}, & \text{if } Y_i = 0. \end{cases} \tag{17}$$

Now consider the function $\tau(x) = \varphi(x)/\Phi(x)$. This function is nonnegative and decreasing, using a Mill's ratio bound from [20]. When $Y_i = 1$, then $1 - \lambda_i = \tau(\mathbf{x}_i^\mathsf{T}\beta)/\tau(\mathbf{x}_i^\mathsf{T}\beta - Z_i) \leq 1$ because then $Z_i \geq 0$. We also used symmetry of $\varphi(\cdot)$. If instead $Y_i = 0$, then $1 - \lambda_i = \tau(-\mathbf{x}_i^\mathsf{T}\beta)/\tau(-\mathbf{x}_i^\mathsf{T}\beta + Z_i) \leq 1$ because then $Z_i \leq 0$. Either way, $1 - \lambda_i \leq 1$ and therefore $\lambda_i \in [0,1)$ for all i.

Writing the previous results in a compact matrix form, we have

$$\frac{\partial \mathbf{Z}}{\partial\beta} = \left(\frac{\partial z_i}{\partial\beta_j}\right)_{ij} = \Lambda X,$$

where $\Lambda = \Lambda(\beta, \mathbf{Z}) = \operatorname{diag}(\lambda_1, \ldots, \lambda_n)$. Similarly, equation (10) yields

$$\frac{\partial\beta}{\partial \mathbf{Z}} = (X^\mathsf{T}X)^{-1}X^\mathsf{T}.$$

Thus, for the $\mathbf{Z}$ update with any $\mathbf{u}_k \in (0,1)^{n+p}$,

$$\begin{aligned} \frac{d_2(\mathbf{Z}^{(k)} - \mathbf{Z}'^{(k)})}{d_1(\beta^{(k-1)} - \beta'^{(k-1)})} &\leq \sup_{\substack{\widetilde{\beta}^{(k-1)}, \widetilde{\mathbf{Z}}^{(k)} \\ d_1(\xi)=1}} d_2\left(\frac{\partial \widetilde{\mathbf{Z}}^{(k)}}{\partial \widetilde{\beta}^{(k-1)}}\xi\right) \\ &\leq \sup_{\substack{\beta, \mathbf{Z} \\ (X\xi)^\mathsf{T}X\xi = 1}} \|\Lambda(\beta, \mathbf{Z})X\xi\| < 1. \end{aligned} \tag{18}$$

For the β update, applying the chain rule gives

$$\frac{\partial \beta^{(k)}}{\partial \beta^{(k-1)}} = \frac{\partial \beta^{(k)}}{\partial \mathbf{Z}^{(k-1)}} \frac{\partial \mathbf{Z}^{(k-1)}}{\partial \beta^{(k-1)}} = (X^{\mathsf{T}}X)^{-1}X^{\mathsf{T}}\Lambda X$$

and then

$$\begin{aligned}
\frac{d_1(\beta^{(k)} - \beta'^{(k)})}{d_1(\beta^{(k-1)} - \beta'^{(k-1)})} &\le \sup_{\widetilde{\beta}, d_1(\xi)=1} d_1\biggl(\frac{\partial \widetilde{\beta}^{(k)}}{\partial \widetilde{\beta}^{(k-1)}}\xi\biggr) \\
&= \sup_{\substack{\beta,\mathbf{Z} \\ (X\xi)^{\mathsf{T}}X\xi=1}} d_1((X^{\mathsf{T}}X)^{-1}X^{\mathsf{T}}\Lambda X\xi) \\
&= \sup_{\beta,\mathbf{Z},\|\eta\|=1} d_1((X^{\mathsf{T}}X)^{-1}X^{\mathsf{T}}\Lambda\eta) \\
&= \sup_{\beta,\mathbf{Z},\|\eta\|=1} \|X(X^{\mathsf{T}}X)^{-1}X^{\mathsf{T}}\Lambda\eta\| \\
&\le \max_{1\le i\le n} \lambda_i \\
&< 1,
\end{aligned} \tag{19}$$

using the nonexpansive property of the projection matrix $X(X^{\mathsf{T}}X)^{-1}X^{\mathsf{T}}$.

By combining (18) with (19), we establish the contraction (13).

7. Open versus closed intervals. In the Lebesgue formulation, $U(0,1)^d$ and $U[0,1]^d$ are the same distribution, in that they cannot be distinguished with positive probability from any countable sample of independent values. Riemann integrals are usually defined for $[0,1]^d$ and discrepancy measures are usually defined for either $[0,1]^d$ or $[0,1)^d$. These latter theories are designed for bounded functions.

In Monte Carlo simulations, sometimes values $u_{ij} \in \{0,1\}$ are produced. These end points can be problematic with inversion, where they may yield extended real values, and hence good practice is to select random number generators supported in the open interval $(0,1)$.

For our Gibbs sampler example with the probit model, we required $\mathbf{u}_k \in (0,1)^{n+p}$. This was necessary because otherwise the values $\phi(\mathbf{x},\mathbf{u})$ might fail to belong to Ω.

Our slice sampler example had Ω equal to the bounded rectangle $[\mathbf{a},\mathbf{b}]$. Then values $u_{ij} \in \{0,1\}$ do not generate sample points outside Ω.

Our Metropolized independence sampler did not require bounded support. It could produce extended real values. Those however are not problematic for weak convergence, which is based on averages of $1_{[\mathbf{a},\mathbf{b}]}(\mathbf{x}_i)$ or other bounded test functions. Also, the chain will not get stuck at an unbounded point.

8. Discussion. We have demonstrated that MCQMC algorithms formed by Metropolis–Hastings updates driven by completely uniformly distributed points can consistently sample a continuous stationary distribution. Some regularity conditions are required, but we have also shown that those conditions hold for many, though by no means all, MCMC updates. The result is a kind of ergodic theorem for QMC like the ones in [7] and [37] for finite state spaces.

When RQMC is used in place of QMC to drive an MCMC simulation, then instead of CUD points, we need to use weakly CUD points. These satisfy $\Pr(D_n^{*d} > \epsilon) \to 0$ for all $\epsilon > 0$ and all $d \in \mathbb{N}$.

Our version of MCMC above leaves out some methods in which one or more components of $\mathbf{u}_i$ are generated by acceptance-rejection sampling because then we cannot assume $d < \infty$. A modification based on splicing i.i.d. $U[0,1]$ random variables into a CUD sequence was proposed by Liao [27] and then shown to result in a weakly CUD sequence in [44].

We do not expect that a global substitution of QMC points will always bring a large improvement to MCMC algorithms. What we do expect is that means of smooth functions of the state vector in Gibbs samplers will often benefit greatly from more even sampling.

It is also a fair question to ask when one needs an MCMC result computed to the high precision that QMC sometimes makes possible. Gelman and Shirley [17] address this issue, distinguishing Task 1 (inference about a parameter θ) from Task 2 [precise determination of $\mathbb{E}(\theta)$ or more generally $\mathbb{E}(f(\theta))$ conditional on the data, or a posterior quantile of θ]. The accuracy of Task 1 problems may be limited more by sample size than by Monte Carlo effort. Task 2 problems include computation of normalizing constants and problems where one wants to report numerically stable, and hence more reproducible, simulation output.

APPENDIX: PROOFS

This Appendix contains the lengthier proofs.

We need one technical lemma about CUD points. Consider overlapping blocks of dk-tuples from u_i, with starting indices d units apart. If u_i are CUD then these overlapping blocks are uniformly distributed. The proof works by embedding the dk-tuples into nonoverlapping rdk-tuples. For large r, the boundary effect between adjacent blocks becomes negligible. This result is also needed for the argument in [37].

LEMMA 6. *For $j \ge 1$, let $u_j \in [0,1]$. For integers $d, i, k \ge 1$, let $\mathbf{x}_i = (u_{d(i-1)+1}, \dots, u_{d(i-1)+dk})$. If u_j are completely uniformly distributed, then $\mathbf{x}_i \in [0,1]^{dk}$ are uniformly distributed.*

PROOF. Choose any $\mathbf{c} \in [0,1]^{dk}$. Let $v = \prod_{j=1}^{dk} c_j$ be the volume of $[\mathbf{0}, \mathbf{c})$. For integers $r \ge 1$, define f_r on $[0,1]^{rdk}$ by $f_r(\mathbf{u}) = \sum_{j=0}^{(r-1)k} 1_{[\mathbf{0},\mathbf{c})}(u_{jd+1}, \dots,$

u_{jd+dk}). Each f_r has Riemann integral $((r-1)k+1)v$. We use f_r on nonoverlapping blocks of length rdk from u_j:

$$\frac{1}{n}\sum_{i=1}^{n} 1_{[\mathbf{0},\mathbf{c})}(\mathbf{x}_i) \geq \frac{1}{n}\sum_{i=1}^{\lfloor n/(rk)\rfloor} f_r\bigl(u_{(i-1)rdk+1},\ldots,u_{irdk}\bigr)$$
$$\to \quad \frac{(r-1)k+1}{rk}v > \frac{r-1}{r}v,$$

after using (4). Taking r as large as we like, we get $\liminf_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n} 1_{[\mathbf{0},\mathbf{c})}(\mathbf{x}_i) \geq v$. It follows that $\liminf_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n} 1_{[\mathbf{a},\mathbf{b})}(\mathbf{x}_i) \geq \mathrm{Vol}[\mathbf{a},\mathbf{b})$ for any rectangular subset $[\mathbf{a},\mathbf{b}) \subset [0,1]^{dk}$. Therefore, $\limsup_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n} 1_{[\mathbf{0},\mathbf{c})}(\mathbf{x}_i) \leq v$ too, for otherwise some rectangle $[\mathbf{a},\mathbf{b})$ would get too few points. □

Now, we prove the main theorems from Section 3.

PROOF OF THEOREM 2. Pick $\varepsilon > 0$. Now let $m \in \mathbb{N}$ and for $i = 1,\ldots,n$ define the sequence $\mathbf{x}'_{i,m,0},\ldots,\mathbf{x}'_{i,m,m} \in \Omega$ as the Rosenblatt–Chentsov transformation of $\mathbf{u}_i,\ldots,\mathbf{u}_{i+m}$.

Suppose that ϕ is regular and for a bounded rectangle $[\mathbf{a},\mathbf{b}] \subset \mathbb{R}^s$, let $f(\mathbf{x}) = 1_{[\mathbf{a},\mathbf{b}]}(\mathbf{x})$. Then

$$\int f(\mathbf{x})\pi(\mathbf{x})\,d\mathbf{x} - \frac{1}{n}\sum_{i=1}^{n} f(\mathbf{x}_i) = \Sigma_1 + \Sigma_2 + \Sigma_3, \tag{20}$$

where

$$\Sigma_1 = \int f(\mathbf{x})\pi(\mathbf{x})\,d\mathbf{x} - \frac{1}{n}\sum_{i=1}^{n} f(\mathbf{x}'_{i,m,m}),$$

$$\Sigma_2 = \frac{1}{n}\sum_{i=1}^{n} f(\mathbf{x}'_{i,m,m}) - f(\mathbf{x}_{i+m})$$

and

$$\Sigma_3 = \frac{1}{n}\sum_{i=1}^{n} f(\mathbf{x}_{i+m}) - f(\mathbf{x}_i).$$

For Σ_1, notice that $\mathbf{x}'_{i,m,m} \in [\mathbf{a},\mathbf{b}]$ if and only if $(v_{d(i-1)+1},\ldots,v_{d(i+m)})$ lies in a $d(m+1)$-dimensional region $\mathcal{B}_1$. The region $\mathcal{B}_1$ has volume $\int_{[\mathbf{a},\mathbf{b}]}\pi(\mathbf{x})\,d\mathbf{x}$ because $\Pr(\mathbf{x}'_{i,m,m} \in [\mathbf{a},\mathbf{b}])$ is $\int_{[\mathbf{a},\mathbf{b}]}\pi(\mathbf{x})\,d\mathbf{x}$ when $(v_{d(i-1)+1},\ldots,v_{d(i+m)}) \sim U[0,1]^{d(m+1)}$. It has a Riemann integrable indicator function by hypothesis. Then because $(v_i)_{i\geq 1}$ are CUD, and using Lemma 6 with $k = m+1$, we get

$$|\Sigma_1| = \left|\int f(\mathbf{x})\pi(\mathbf{x})\,d\mathbf{x} - \frac{1}{n}\sum_{i=1}^{n} f(\mathbf{x}'_{i,m,m})\right| \longrightarrow 0 \qquad \text{as } n \to \infty.$$

Now, consider Σ_2. The only nonzero terms arise when $\mathbf{x}_{i+m} \neq \mathbf{x}'_{i,m,m}$. This in turn requires that the coupling region $\mathcal{C}$ is avoided m consecutive times, by $\mathbf{u}_{i+1}, \ldots, \mathbf{u}_{i+m}$. Then $(v_{di+1}, \ldots, v_{d(i+m)})$ belongs to a region of volume at most $(1 - \mathrm{Vol}(\mathcal{C}))^m$. Choose m large enough that $(1 - \mathrm{Vol}(\mathcal{C}))^m < \varepsilon$. Then

$$\lim_{n\to\infty} \sup \left| \frac{1}{n} \sum_{i=1}^{n} f(\mathbf{x}'_{i,m,m}) - f(\mathbf{x}_{i+m}) \right| < \varepsilon.$$

For the third term, $|\Sigma_3|$ is at most m/n, which goes to 0 as $n \to \infty$. Thus, we have

$$\left| \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} 1_{\mathbf{x}_i \in [\mathbf{a},\mathbf{b}]} - \int_{[\mathbf{a},\mathbf{b}]} \pi(\mathbf{x})\, d\mathbf{x} \right| < \varepsilon.$$

As $\varepsilon > 0$ was chosen arbitrarily, the result follows for this case.

The result holds trivially for the function 1_Ω, hence we are done. □

PROOF OF THEOREM 3. We use the notation from the proof of Theorem 2. As in the proof of Theorem 2, we write $\int f(\mathbf{x})\pi(\mathbf{x})\, d\mathbf{x} - \frac{1}{n}\sum_{i=1}^{n} f(\mathbf{x}_i)$ as the sum of three terms. The first and third terms vanish by the same arguments we used in Theorem 2.

For the second term, we have

$$|\Sigma_2(n)| \le \frac{1}{n} \sum_{i=1}^{n} |f(\mathbf{x}'_{i,m,m}) - f(\mathbf{x}_{i+m})|.$$

Let $\varepsilon > 0$ be arbitrary. We show that $\limsup_{n\to\infty} |\Sigma_2(n)| \le \varepsilon$. As $\varepsilon > 0$ is arbitrary, this then implies that $\limsup_{n\to\infty} |\Sigma_2(n)| = 0$.

Assume that the Gibbs sampler is regular for rectangles and for a bounded positive volume rectangle $[\mathbf{a}, \mathbf{b}] \subset \mathbb{R}^s$ let $f(\mathbf{x}) = 1_{[\mathbf{a},\mathbf{b}]}(\mathbf{x})$. For $0 \le \delta < \min_{1\le j\le d}(b_j - a_j)$, let $\boldsymbol{\delta} = (\delta, \ldots, \delta) \in \mathbb{R}^s$ and put $f_\delta(\mathbf{x}) = 1_{[\mathbf{a}-\boldsymbol{\delta},\mathbf{b}+\boldsymbol{\delta}]}$ and $f_{-\delta}(\mathbf{x}) = 1_{[\mathbf{a}+\boldsymbol{\delta},\mathbf{b}-\boldsymbol{\delta}]}$.

Because $f_\delta(\mathbf{x}) \ge f(\mathbf{x}) \ge f_{-\delta}(\mathbf{x})$, the triple $(f_{-\delta}(\mathbf{x}'_{i,m,m}), f(\mathbf{x}'_{i,m,m}), f_\delta(\mathbf{x}'_{i,m,m}))$ must be in the set $S = \{(0,0,0), (0,0,1), (0,1,1), (1,1,1)\}$. Likewise $f(\mathbf{x}_{i+m}) \in \{0,1\}$. By inspecting all 8 cases in $S \times \{0,1\}$, we find that $|\Sigma_2| \le \sigma_1 + \sigma_2 + \sigma_3$, for

$$\sigma_1 = \frac{1}{n} \sum_{i=1}^{n} f_\delta(\mathbf{x}'_{i,m,m}) - f_{-\delta}(\mathbf{x}'_{i,m,m}),$$

$$\sigma_2 = \frac{1}{n} \sum_{i=1}^{n} \big(f_{-\delta}(\mathbf{x}'_{i,m,m}) - f(\mathbf{x}_{i+m})\big)_+$$

and

$$\sigma_3 = \frac{1}{n} \sum_{i=1}^{n} \big(f(\mathbf{x}_{i+m}) - f_\delta(\mathbf{x}'_{i,m,m})\big)_+,$$

where $z_+ = \max(z, 0)$.

Choose $\delta > 0$ such that

$$\int_{\Omega \cap ([\mathbf{a}-\boldsymbol{\delta}, \mathbf{b}+\boldsymbol{\delta}] \setminus [\mathbf{a}+\boldsymbol{\delta}, \mathbf{b}-\boldsymbol{\delta}])} \pi(\mathbf{x})\, d\mathbf{x} < \frac{\varepsilon}{3}.$$

As the Gibbs sampler is regular for rectangles, $(v_i)_{i \geq 1}$ is a CUD sequence, and $\mathbf{x}'_{i,m,m}$ is constructed using the Rosenblatt–Chentsov transformation we have

$$\begin{aligned}&\lambda(\{\mathbf{u} \in [0,1]^{dm+d} : \mathbf{x}'_{i,m,m} \in [\mathbf{a} - \boldsymbol{\delta}, \mathbf{b} + \boldsymbol{\delta}] \setminus [\mathbf{a}, \mathbf{b}]\}) \\ &\quad = \int_{\Omega \cap ([\mathbf{a}-\boldsymbol{\delta}, \mathbf{b}+\boldsymbol{\delta}] \setminus [\mathbf{a}+\boldsymbol{\delta}, \mathbf{b}+\boldsymbol{\delta}])} \pi(\mathbf{x})\, d\mathbf{x} < \frac{\varepsilon}{3},\end{aligned}$$

and so $\limsup_{n\to\infty} |\sigma_1(n)| \leq \varepsilon/3$.

The points $\mathbf{x}'_{i,m,m}$ and $\mathbf{x}_{i+m}$ have different starting points $\mathbf{x}'_{i,m,0}$ and $\mathbf{x}_i$, but are updated m times using the same $\mathbf{u}_{i+1}, \ldots, \mathbf{u}_{i+m}$, that is, $\mathbf{x}'_{i,m,m} = \phi_m(\mathbf{x}'_{i,m,0}, \mathbf{u}_{i+1}, \ldots, \mathbf{u}_m)$ and $\mathbf{x}_{i+m} = \phi_m(\mathbf{x}_i, \mathbf{u}_{i+1}, \ldots, \mathbf{u}_{i+m})$. Therefore, Theorem 3 implies that there is a constant $C > 0$ such that for all sufficiently large $m \geq m_i^*$ the region

$$\begin{aligned}\mathcal{B}_{m,i} = \{(\mathbf{v}_1, \ldots, \mathbf{v}_m) \in [0,1]^{dm} : d(&\phi_m(\mathbf{x}'_{i,m,0}, (\mathbf{v}_1, \ldots, \mathbf{v}_m)), \\ &\phi_m(\mathbf{x}_i, (\mathbf{v}_1, \ldots, \mathbf{v}_m))) > \gamma^m\},\end{aligned}$$

has volume at most $C\alpha_\gamma^m$. Let $\mathcal{B}_m = \bigcup_{i=1}^n \mathcal{B}_{m,i}$. Let $\beta = \infty$ if $[\mathbf{a}, \mathbf{b}] \cap \Omega = \varnothing$ or $\Omega \setminus [\mathbf{a} - \boldsymbol{\delta}, \mathbf{b} + \boldsymbol{\delta}] = \varnothing$ and $\beta = \inf\{d(\mathbf{y}, \mathbf{y}') : \mathbf{y} \in [\mathbf{a}, \mathbf{b}] \cap \Omega, \mathbf{y}' \in \Omega \setminus [\mathbf{a} - \boldsymbol{\delta}, \mathbf{b} + \boldsymbol{\delta}]\}$ otherwise.

Let $m_1 = m_1(n)$ be such that $Cn\alpha_\gamma^{m_1} < \varepsilon/3$ and $\gamma^{m_1} < \beta$. Now take $m_0 \geq \max\{m_1, m_1^*, \ldots, m_n^*\}$. For large enough n, we can take $m_0 = m_0(n) = \lceil \frac{\log n + \log(2C/\varepsilon)}{\log 1/\alpha_\gamma} \rceil + 1$. Then $\mathcal{B}_{m_0}$ has volume at most $\varepsilon/3$.

Thus, $f_{-\boldsymbol{\delta}}(\mathbf{x}'_{i,m_0,m_0}) > f(\mathbf{x}_{i+m_0})$ implies that $d(\mathbf{x}'_{i,m_0,m_0}, \mathbf{x}_{i+m_0}) \geq \beta$, which in turn implies that $(\mathbf{u}_{i+1}, \ldots, \mathbf{u}_{i+m_0}) \in \mathcal{B}_{m_0,i}$, and so $(\mathbf{u}_{i+1}, \ldots, \mathbf{u}_{i+m_0}) \in \mathcal{B}_{m_0}$. Therefore, we have

$$\limsup_{n\to\infty} |\sigma_2(n)| \leq \limsup_{n\to\infty} \frac{1}{n} \sum_{i=1}^n 1_{(\mathbf{u}_{i+1}, \ldots, \mathbf{u}_{i+m_0}) \in \mathcal{B}_{m_0}} = \limsup_{m_0\to\infty} \lambda(\mathcal{B}_{m_0}) \leq \frac{\varepsilon}{3}.$$

A similar argument shows that $\limsup_{n\to\infty} |\sigma_3(n)| \leq \varepsilon/3$.

Combining the three bounds yields

$$\begin{aligned}\limsup_{n\to\infty} |\Sigma_2(n)| &\leq \limsup_{n\to\infty} \sigma_1(n) + \limsup_{n\to\infty} \sigma_2(n) + \limsup_{n\to\infty} \sigma_3(n) \\ &\leq \frac{\varepsilon}{3} + \frac{\varepsilon}{3} + \frac{\varepsilon}{3} = \varepsilon,\end{aligned}$$

establishing consistency when the Gibbs sampler is regular.

Since the result holds trivially for the function 1_Ω, the result follows. □

The coupling region in Theorem 2 was replaced by a mean contraction assumption $\int_{[0,1]^d} \log(\ell(\mathbf{u}))\, d\mathbf{u} < 0$ in Theorem 4. This way we obtain (possibly different) coupling type regions $\mathcal{B}_{m,i}$ for each $i = 1, \ldots, n$. We remedy this situation by letting m depend on n, which in turn requires us to use a stronger assumption on the CUD sequence $(v_i)_{i\geq 1}$, namely, that $\lim_{n\to\infty} D_n^{*d_n} = 0$.

Acknowledgments. We thank Seth Tribble, Erich Novak, Ilya M. Sobol' and two anonymous reviewers for helpful comments.

REFERENCES

[1] ALBERT, J. and CHIB, S. (1993). Bayesian analysis of binary and polychotomous response data. *J. Amer. Statist. Assoc.* **88** 669–679. MR1224394

[2] ALSMEYER, G. and FUH, C.-D. (2001). Limit theorems for iterated function mappings. *Stochastic Process. Appl.* **96** 123–142. MR1856683

[3] ANDRIEU, C. and MOULINES, E. (2006). On the ergodicity properties of some MCMC algorithms. *Ann. Appl. Probab.* **16** 1462–1505. MR2260070

[4] ASH, R. B. (1972). *Real Analysis and Probability*. Academic Press, New York. MR0435320

[5] BILLINGSLEY, P. (1999). *Convergence of Probability Measures*. Wiley, New York. MR1700749

[6] CHAUDARY, S. (2004). Acceleration of Monte Carlo methods using low discrepancy sequences. Ph.D. thesis, UCLA.

[7] CHENTSOV, N. N. (1967). Pseudorandom numbers for modelling Markov chains. *Comput. Math. Math. Phys.* **7** 218–2332.

[8] CRAIU, R. V. and LEMIEUX, C. (2007). Acceleration of the multiple-try Metropolis algorithm using antithetic and stratified sampling. *Stat. Comput.* **17** 109–120. MR2380640

[9] DEVROYE, L. (1986). *Non-Uniform Random Variate Generation*. Springer, New York. MR0836973

[10] DIACONIS, P. and FREEDMAN, D. (1999). Iterated random functions. *SIAM Rev.* **41** 45–76. MR1669737

[11] DICK, J. (2007). A note on the existence of sequences with small star discrepancy. *J. Complexity* **23** 649–652. MR2372019

[12] DICK, J. (2009). On quasi-Monte Carlo rules achieving higher order convergence. In *Monte Carlo and Quasi-Monte Carlo Methods 2008* (P. L'Ecuyer and A. B. Owen, eds.) 73–96. Springer, Heidelberg.

[13] DOERR, B. and FRIEDRICH, T. (2009). Deterministic random walks on the two-dimensional grid. *Combin. Probab. Comput.* **18** 123–144. MR2497377

[14] FINNEY, D. J. (1947). The estimation from individual records of the relationship between dose and quantal response. *Biometrika* **34** 320–334.

[15] GAVER, D. and O'MURCHEARTAIGH, I. (1987). Robust empirical Bayes analysis of event rates. *Technometrics* **29** 1–15. MR0876882

[16] GELFAND, A. E. and SMITH, A. F. M. (1990). Sampling-based approaches to calculating marginal densities. *J. Amer. Statist. Assoc.* **85** 398–409. MR1141740

[17] GELMAN, A. and SHIRLEY, K. (2010). Inference from simulations and monitoring convergence. In *Handbook of Markov Chain Monte Carlo: Methods and Applications*. (S. Brooks, A. Gelman, G. Jones and X.-L. Meng, eds.) 131–143. Chapman and Hall/CRC Press, Boca Raton, FL.

[18] GHORPADE, S. R. and LIMAYE, B. V. (2006). *A Course in Calculus and Real Analysis*. Springer, New York. MR2229667

[19] GNEWUCH, M., SRIVASTAV, A. and WINZEN, C. (2008). Finding optimal volume subintervals with k points and computing the star discrepancy are NP-hard. *J. Complexity* **24** 154–172.
[20] GORDON, R. D. (1941). Value of Mill's ratio of area to bounding ordinate and of the normal probability integral for large values of the argument. *Ann. Math. Statist.* **18** 364–366. MR0005558
[21] HAARIO, H., SAKSMAN, E. and TAMMINEN, J. (2001). An adaptive Metropolis algorithm. *Bernoulli* **7** 223–242. MR1828504
[22] KNUTH, D. E. (1998). *The Art of Computer Programming*, 3rd ed. *Seminumerical Algorithms* **2**. Addison-Wesley, Reading, MA. MR0378456
[23] LEBESGUE, H. L. (1902). Intégrale, longueur, aire. Ph.D. thesis, Univ. de Paris.
[24] L'ECUYER, P., LECOT, C. and TUFFIN, B. (2008). A randomized quasi-Monte Carlo simulation method for Markov chains. *Oper. Res.* **56** 958–975.
[25] L'ECUYER, P. and LEMIEUX, C. (1999). Quasi-Monte Carlo via linear shift-register sequences. In *Proceedings of the 1999 Winter Simulation Conference* (P. A. Farrington, H. B. Nembhard, D. T. Sturrock and G. W. Evans, eds.) 632–639. IEEE Press, Piscataway, NJ.
[26] LEMIEUX, C. and SIDORSKY, P. (2006). Exact sampling with highly uniform point sets. *Math. Comput. Modelling* **43** 339–349. MR2214643
[27] LIAO, L. G. (1998). Variance reduction in Gibbs sampler using quasi random numbers. *J. Comput. Graph. Statist.* **7** 253–266.
[28] LIU, J. S. (2001). *Monte Carlo Strategies in Scientific Computing*. Springer, New York. MR1842342
[29] MATSUMOTO, M. and NISHIMURA, T. (1998). Mersenne twister: A 623-dimensionally equidistributed uniform pseudorandom number generator. *ACM Transactions on Modeling and Computer Simulation* **8** 3–30.
[30] MARSDEN, J. E. and HOFFMAN, M. J. (1993). *Elementary Classical Analysis*, 2nd ed. Macmillan, New York.
[31] MOROKOFF, W. and CAFLISCH, R. E. (1993). A quasi-Monte Carlo approach to particle simulation of the heat equ ation. *SIAM J. Numer. Anal.* **30** 1558–1573. MR1249033
[32] NEAL, R. M. (2003). Slice sampling. *Ann. Statist.* **31** 705–767. MR1994729
[33] NIEDERREITER, H. (1986). Multidimensional integration using pseudo-random numbers. *Math. Programming Stud.* **27** 17–38. MR0836749
[34] NIEDERREITER, H. (1992). *Random Number Generation and Quasi-Monte Carlo Methods*. SIAM, Philadelphia, PA. MR1172997
[35] OWEN, A. B. (1995). Randomly permuted (t, m, s)-nets and (t, s)-sequences. In *Monte Carlo and Quasi-Monte Carlo Methods in Scientific Computing* (H. Niederreiter and P. Jau-Shyong Shiue, eds.) 299–317. Springer, New York. MR1445791
[36] OWEN, A. B. (2005). Multidimensional variation for quasi-Monte Carlo. In *Contemporary Multivariate Analysis and Design of Experiments: In Celebration of Prof. Kai-Tai Fang's 65th Birthday* (J. Fan and G. Li, eds.). World Sci. Publ., Hackensack, NJ. MR2271076
[37] OWEN, A. B. and TRIBBLE, S. D. (2005). A quasi-Monte Carlo Metropolis algorithm. *Proc. Natl. Acad. Sci. USA* **102** 8844–8849. MR2168266
[38] PROPP, J. G. and WILSON, D. B. (1996). Exact sampling with coupled Markov chains. *Random Structures and Algorithms* **9** 223–252. MR1611693
[39] ROBERT, C. P. and CASELLA, G. (2004). *Monte Carlo Statistical Methods*, 2nd ed. Springer, New York. MR2080278
[40] ROBERTS, G. O., ROSENTHAL, J. S. and SCHWARTZ, P. O. (1998). Convergence properties of perturbed Markov chains. *J. Appl. Probab.* **35** 1–11. MR1622440
[41] ROSENBLATT, M. (1952). Remarks on a multivariate transformation. *Ann. Math. Statist.* **23** 470–472. MR0049525
[42] SOBOL', I. M. (1974). Pseudo-random numbers for constructing discrete Markov chains by the Monte Carlo method. *USSR Comput. Math. Math. Phys.* **14** 36–45. MR0339444

[43] TRIBBLE, S. D. (2007). Markov chain Monte Carlo algorithms using completely uniformly distributed driving sequences. Ph.D. thesis, Stanford Univ. MR2710331

[44] TRIBBLE, S. D. and OWEN, A. B. (2008). Construction of weakly CUD sequences for MCMC sampling. *Electron. J. Stat.* **2** 634–660. MR2426105

[45] WEYL, H. (1916). Über die gleichverteilung von zahlen mod. eins. *Math. Ann.* **77** 313–352. MR1511862

S. CHEN
STANFORD UNIVERSITY
SEQUOIA HALL
STANFORD, CALIFORNIA 94305
USA

J. DICK
SCHOOL OF MATHEMATICS AND STATISTICS
UNIVERSITY OF NEW SOUTH WALES
SYDNEY
AUSTRALIA

A. B. OWEN
STANFORD UNIVERSITY
SEQUOIA HALL
STANFORD, CALIFORNIA 94305
USA
E-MAIL: owen@stat.stanford.edu