A Scrambled Method of Moments and other qMCMD Estimators (In Progress)

Jean-Jacques Forneron

May 10, 2019

Boston University

Introduction: Simulated Method of Moments

- Want to estimate a static/dynamic model that is intractable:
  - GMM moments/Likelihood not analytical tractable
  - e.g. Dynamic Discrete Choice, DSGE, Asset Pricing,...
  - often because of latent variables and non-closed form policy functions
- Solution: use Simulation-Based Estimation
  - simulate $S \geq 1$ artificial samples $y_t^s(\theta)$, $t = 1, \dots, n$
  - match sample moments $\hat{\psi}_n$ with simulated $\hat{\psi}_n^S(\theta)$

$$\hat{\theta}_{SMD} = \text{argmin}_\theta \|\hat{\psi}_n - \hat{\psi}_n^S(\theta)\|_W$$

- Under regularity conditions:

$$\sqrt{n}(\hat{\theta}_{SMD} - \theta_0) \xrightarrow{d} \mathcal{N}\Big(0, (1 + \underbrace{1/S}_{\text{Simulation Noise}}) \times V\Big)$$

- Question: can we make the estimator efficient, i.e. no simulation noise, with the same moments, same S, etc.?

Introduction: Simulated Method of Moments

- In theory:
  - more simulations $\Rightarrow$ smaller variance
- In practice:
  - more informative moments are often more costly to evaluate: *tradeoff between larger S and better $\hat{\psi}_n$*
  - sample may be large: implies $n \times S$ very large...
  - optimization may be slow: $\nearrow S$ takes even more time
- Yet $S = 1 \Rightarrow$ yields a variance twice as large as $S = +\infty$
- Breaking the tradeoff between S and $\hat{\psi}_n$ could be very useful
  - smaller variance due to reduced simulation noise
  - better $\hat{\psi}_n$ possible without increasing simulation noise

Reducing Simulation Error

- Rely on **(randomized) quasi-Monte Carlo** methods instead of random draws to simulate y_t^s
- Call these Scrambled Method of Moments and qMCMD
- **Static models**: under standard assumptions, get for $S = 1$:

$$\sqrt{n}(\hat{\theta}_{SMM/qMCMD} - \theta_0) \xrightarrow{d} \mathcal{N}(0, V_\infty)$$

  no simulation noise (asymptotically), same class of moments
- **Dynamic Models**: for a certain class of moments, simulate in a way such that:

$$\sqrt{n}(\hat{\theta}_{SMM/qMCMD} - \theta_0) \xrightarrow{d} \mathcal{N}(0, V_\infty + V_{MC-qMC}/S)$$

  where $V_{MC} \geq V_{MC-qMC}$ under some conditions (**in progress**)

The Big Picture

- SMM and Indirect Inference are very popular methods
  - major applied journals: 225 have GMM in title/abstract/keywords in 2015 vs. 73 for SMM/II ($\sim$ 32%)
  - more and more popular with $\nearrow$ of heterogeneous agents, production networks, dynamic entry/exit models, etc.
- Powerful method with some caveats, mainly:
  i. assumes a fully parametric model
  ii. identification is assumed, hard to check
  iii. simulation noise: statistical/computational efficiency tradeoff

Related Literatures

- **(Randomized) quasi-Monte Carlo** methods used in
  - **state-space filtering**: sequential quasi-Monte Carlo (Gerber and Chopin, 2015, 2017)
  - **Bayesian sampling**, e.g.: ABC (Buchholz and Chopin, 2017)
  - **Option pricing** in finance (Paskov and Traub, 1995)
  - Review: Lemieux (2009)
  - Underlying theory: Dick and Pillichshammer (2010)
- **Simulation-Based Estimation**:
  - Classics: McFadden (1989); Pakes and Pollard (1989); Gouriéroux et al. (1993); Duffie and Singleton (1993); Gallant and Tauchen (1996)...
  - More recent: Evdokimov for static models $\nearrow$ number of moments instead of S

Background: Monte-Carlo and quasi-Monte Carlo Integration

Monte Carlo Integration

- We want to approximate the following integral:

$$\int_{[0,1]^d} f(u)du \quad \text{where } f : [0,1]^d \to \mathbb{R}$$

- Monte Carlo approach: draw $u_1, \ldots, u_n \overset{iid}{\sim} \mathcal{U}_{[0,1]^d}$ and

$$\frac{1}{n}\sum_{i=1}^{n} f(u_i) - \int_{[0,1]^d} f(u)du = O_p(n^{-1/2})$$

if $f \in \mathbb{L}^2([0,1]^d)$ i.e. $f(u_i)$ has finite variance

- Worse-case probabilistic error bound is $O(\sqrt{\log\log(n)/n})$ by the Law of the Iterated Logarithm
- Can we do better?

Numerical Integration

- For $d = 1$, $u_1 = 0, u_2 = 1/n, \ldots, u_{n+1} = 1$ and f Lipschitz

$$\left|\frac{1}{n}\sum_{i=1}^{n} f(u_i) - \int_{[0,1]} f(u)du\right| \leq C_f \times n^{-1}$$

- For $d = 2$, use the regular lattice grid and

$$\left|\frac{1}{n}\sum_{i=1}^{n} f(u_i) - \int_{[0,1]^2} f(u)du\right| \leq C_f \times n^{-1/2}$$

- For $d = 3$, use the regular lattice grid and

$$\left|\frac{1}{n}\sum_{i=1}^{n} f(u_i) - \int_{[0,1]^3} f(u)du\right| \leq C_f \times n^{-1/3}$$

- Summary: $d = 1$ better, $d = 2$ same, $d \geq 3$ worse, ...
- ...that's the curse of dimensionality

Approximation Error for Arbitrary Sequences

- **Koksma-Hlawka inequality**: pick any $(u_i)_{i=1,\dots,n} \in [0,1]^d$. For $f : [0,1]^d \to \mathbb{R}$, the integration error is bounded by:

$$\Big|\frac{1}{n}\sum_{i=1}^{n} f(u_i) - \int_{[0,1]^d} f(u)du\Big| \leq \|f\|_{TV} \times D_n^\star(u_1,\dots,u_n)$$

- where $\|f\|_{TV}$ is the **Total Variation** in the sense of Hardy and Krause, i.e.:

$$\int_{[0,1]}\Big|\frac{\partial f(u)}{\partial u}\Big|du \text{ for } d=1, \quad \sum_{\mathfrak{u}\subseteq\mathcal{I}_d}\int_{[0,1]^{|\mathfrak{u}|}}\Big|\frac{\partial^{|\mathfrak{u}|} f(\mathfrak{u})}{\partial \mathfrak{u}}\Big|d\mathfrak{u} \text{ for } d>1$$

- $D_n^\star(u_1,\dots,u_n)$ is the **Star Discrepancy** of $u_1,\dots,u_n$

Dispersion and Discrepancy

- The dispersion of $(u_1, \ldots, u_n)$ is:

$$DP(u_1, \ldots, u_n) = \sup_{u \in [0,1]^d} \left(\inf_{i=1,\ldots,n} \|u - u_i\| \right)$$

  lower bound: $DP(u_1, \ldots, u_n) \geq n^{-1/d}$, up to constants

- The **Star Discrepancy** of $(u_1, \ldots, u_n)$ is:

$$D_n^\star(u_1, \ldots, u_n) = \sup_{\mathfrak{u} \in [0,1]^d} \left| \frac{\sum_{i=1}^n \mathbb{1}_{u_i \in [0,\mathfrak{u})}}{n} - \int_{[0,\mathfrak{u})} 1 du \right|$$

  it measures **how well** $(u_1, \ldots, u_n)$ **approximates the Lebesgue measure for rectangles** $[0, \mathfrak{u}) \subseteq [0, 1)$

- It's equal to the KS distance between the two measures

Number Theory to the Rescue

- Number Theory provides the following lower-bound (proved by Schmidt for $d = 2$ and Roth for $d > 2$)

$$D_n^\star(u_1, \ldots, u_n) \geq C_d \frac{\log(n)^{d-1}}{n}$$

where C_d depends only on $d \geq 1$; it suggests a $n^{-1+\varepsilon}$ rate is feasible for any $\varepsilon > 0$

- For random draws $u_i \overset{iid}{\sim} \mathcal{U}_{[0,1]^d}$

$$D_n^\star(u_1, \ldots, u_n) = O(\sqrt{\log\log(n)/n})$$

Monte-Carlo is inefficient for $d \geq 1$

- Lattice rule:

$$D_n^\star(u_1, \ldots, u_n) = O(n^{-1/d})$$

Lattice rule is inefficient for $d > 1$

- Deterministic sequences close (in rate) to the lower bound
- **Sobol Sequence**:

$$D_n^\star(u_1, \dots, u_n) \asymp \underbrace{2^d}_{\text{increases very quickly with d}} \times \frac{\log(n)^d}{n}$$

- Also: **van der Corput, Hammersley, Halton**,...

$$D_n^\star(u_1, \dots, u_n) \asymp \underbrace{C_d}_{\text{increases rapidly with d}} \times \frac{\log(n)^d}{n}$$

- for fixed d, these are faster than MC methods
- increasing/large d: $C_d \to +\infty$, curse of dimensionality

Randomized quasi-Monte Carlo

- **Digital Shift** take $u \sim \mathcal{U}_{[0,1]^d}$ and u_i low-discrepancy

$$\tilde{u}_i = u_i + u \text{ modulo } 1$$

  shifts all dimensions independently

- **The Scramble** (Owen, 1997), more complicated write:

$$u_i = \sum_{j=1}^{k} \frac{x_{i,j}}{b^j}$$

  e.g. for Sobol $b = 2$ (binary expansion)

  - do a uniform permutation of the $x_{i,j}$ across the u_i
  - **yields a $\mathcal{U}_{[0,1]^d}$ draw which is still low-discrepancy wp. 1**
  - can compute the variance the usual way

The Scramble: Abracadabra

Theorem (Owen (1997))

- *Let $u_1, \ldots, u_n$ be a scrambled low-discrepancy sequence and $f \in \mathbb{L}^2([0,1]^d)$ (possibly non-smooth/discontinuous), then:*

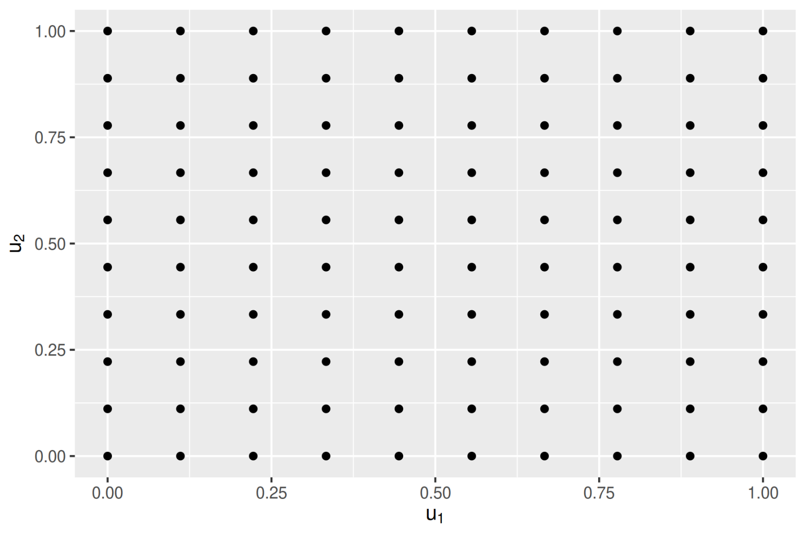
$$\frac{1}{n}\sum_{i=1}^{n} f(u_i) - \int_{[0,1]^d} f(u)du = o_p(n^{-1/2})$$

- *If additionally f is sufficiently smooth then the last term is $O_p(n^{-3/2+\varepsilon})$ for any $\varepsilon > 0$ – d is fixed here*

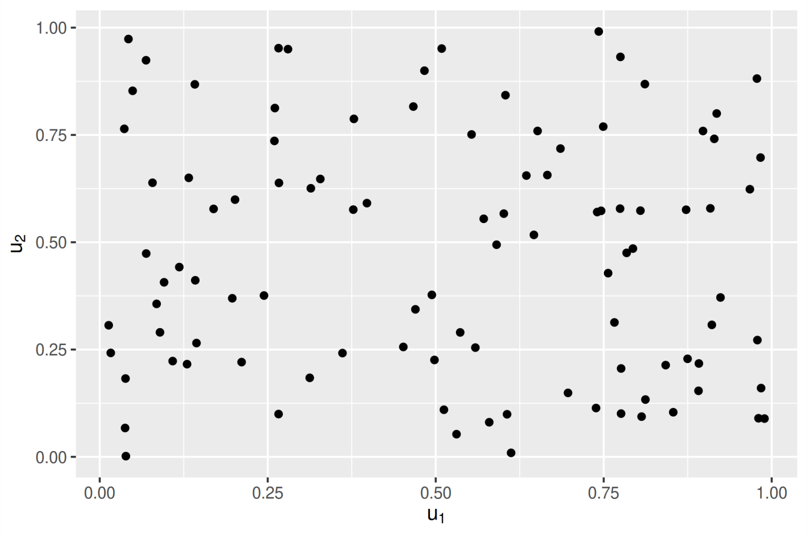
Remarks:

- d is fixed, curse of dimensionality still applies
- conditions are identical to the CLT for iid draws, rate is faster

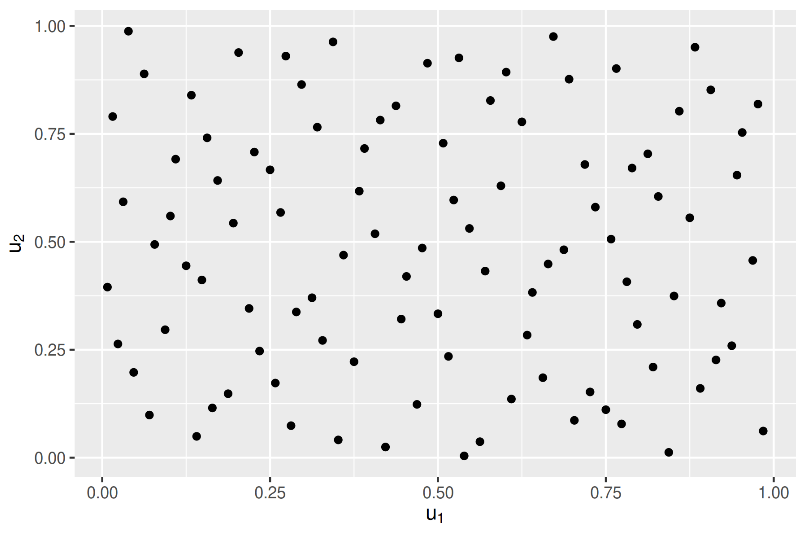
Visualisation: Lattice Rule



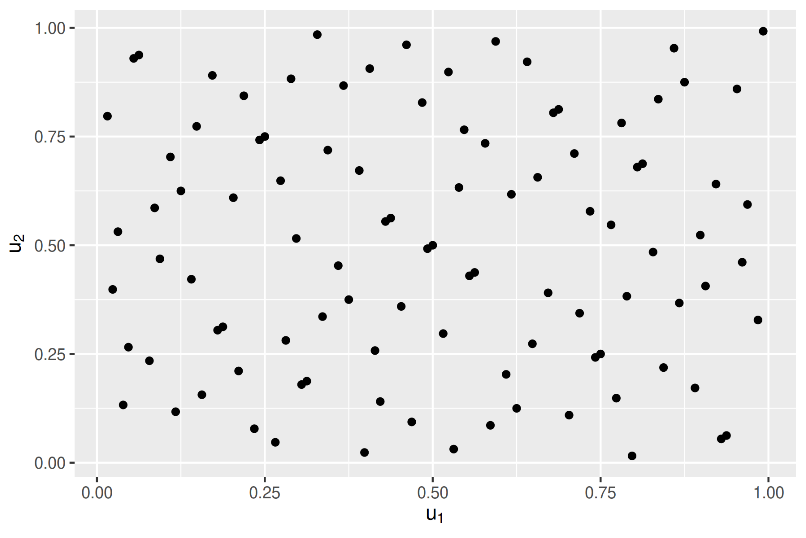
Visualisation: Random Uniform Draws



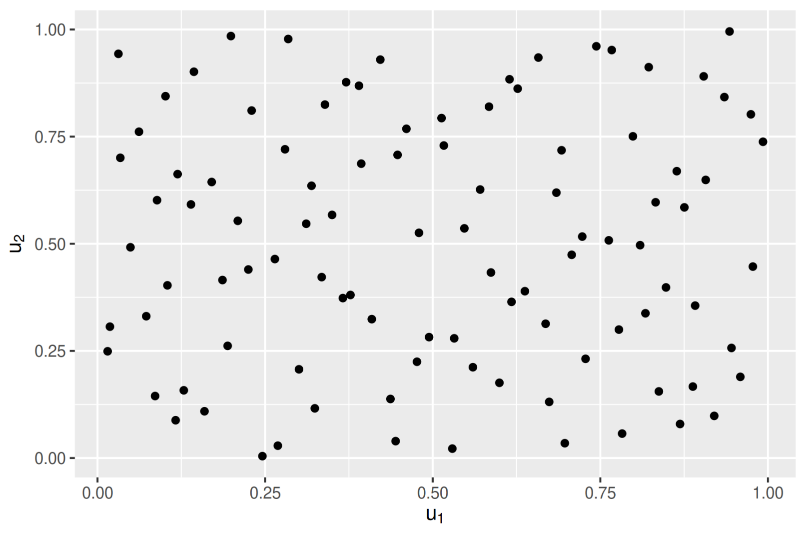
Visualisation: Halton Sequence



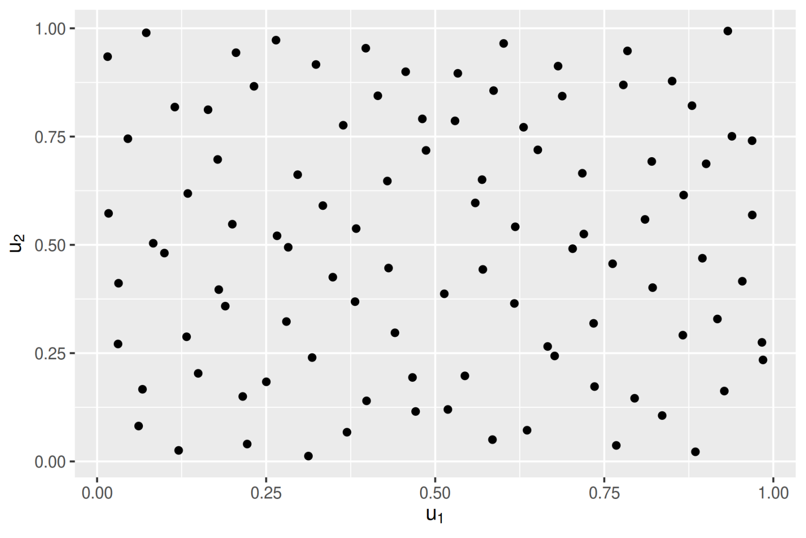
Visualisation: Sobol Sequence



Visualisation: Scrambled Sobol Sequence 1/2



Visualisation: Scrambled Sobol Sequence 2/2



Quasi-Monte Carlo Estimation: Static Models

Static Models: the setting

- Static Models - cross-sectional:

$$y_i = g(\theta, u_i), u_i \overset{iid}{\sim} \mathcal{U}_{[0,1]^d}$$

- Sample Moments:

$$\hat{\psi}_n = \frac{1}{n}\sum_{i=1}^{n} \psi(y_i)$$

- Simulated Moments:

$$\hat{\psi}_n^s(\theta) = \frac{1}{n}\sum_{i=1}^{n} \psi(y_i^s), \quad y_i^s = g(\theta, u_i^s), u_i^s \overset{iid}{\sim} \mathcal{U}_{[0,1]^d}$$

- Simulation-based estimator:

$$\hat{\theta}_n^S = \text{argmin}_{\Theta} \|\hat{\psi}_n - \hat{\psi}_n^S(\theta)\|_W$$

Static Models: Simulated Method of Moments

Theorem (Pakes and Pollard (1989))

Suppose $\theta \to \lim_{n\to\infty} \mathbb{E}(\hat{\psi}_n^s(\theta))$ *is injective, that* $\hat{\psi}_n$ *satisfies a LLN and a CLT. Suppose* $\hat{\psi}_n^s(\cdot)$ *satisfies a uniform CLT and stochastic equicontinuity conditions then:*

$$\sqrt{n}\left(\hat{\theta}_n^S - \theta_0\right) \xrightarrow{d} \mathcal{N}\left(0, (1 + 1/S) \times V\right)$$

Static Models: Scrambled Method of Moments

Theorem (Scrambled Method of Moments)

Suppose Pakes and Pollard (1989)'s conditions holds and u_i^s is a scrambled net sequence then:

$$\sqrt{n}\left(\hat{\theta}_n^S - \theta_0\right) \xrightarrow{d} \mathcal{N}\left(0, V\right)$$

for any $S \geq 1$ fixed.

Remarks:

- computation is identical
- simply change the random number generator
- asymptotic variance is equivalent to SMM with $S(n) \to +\infty$

Proof of the Theorem (Sketch)

- Note that:

$$\mathbb{E}(\hat{\psi}_n^s(\theta)) = \int_{[0,1]^d} \psi \circ g(u;\theta) du$$

  is within the framework of quasi-Monte Carlo integration
- Scrambled $u_i^s \overset{iid}{\sim} \mathcal{U}_{[0,1]^d}$; invoke Pakes and Pollard (1989):

$$\sqrt{n}\left(\hat{\theta}_{SMM} - \theta_0\right) = -(G'WG)^{-1}G'W\left(\hat{\psi}_n - \hat{\psi}_n^S(\theta_0)\right) + o_p(1)$$

  where $G = \partial_\theta \mathbb{E}(\hat{\psi}_n^S(\theta_0))$
- Finally $\hat{\psi}_n^s(\theta_0)$ is scrambled so:

$$\sqrt{n}\left(\hat{\theta}_{SMM} - \theta_0\right) \xrightarrow{d} \mathcal{N}(0, V)$$

  where $V = avar(\hat{\psi}_n)$ does not depend on $S \geq 1$
- Discussion about covariates x_i at the end

Illustration: a Pen and Pencil Example

- Very basic example:

$$y_i = \mu + \sigma e_i, e_i \overset{iid}{\sim} \mathcal{N}(0,1)$$

- Method of moments: $\hat{\mu} = \bar{y}_n$,

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(y_i - \bar{y}_n)^2 \text{ (biased)}$$

- Gouriéroux et al. (1993), SMM:

$$\hat{\sigma}^2_{SMM} = \frac{\hat{\sigma}^2}{\frac{1}{nS}\sum_{s=1}^{S}\sum_{i=1}^{n}(e_i^s - \bar{e}_n^s)^2} \overset{S\to\infty}{\to} \frac{n}{n-1}\hat{\sigma}^2 \text{ (unbiased)}$$

- Minimum Distance ($S = +\infty$):

$$\hat{\sigma}^2_{MD} = \frac{n}{n-1}\hat{\sigma}^2 \text{ (unbiased)}$$

Scrambled vs. Simulated Method of Moments vs. Analytical

$S = 1$	$\hat{\sigma}^2_{MM}$	$\hat{\sigma}^2_{MD}$	$\hat{\sigma}^2_{SMM}$	$\hat{\sigma}^2_{qMCMD}$
$n \times variance$	2.00	2.02	4.10	2.08
$n \times bias$	-1.11	-0.11	1.38	-2.11
$n \times MSE$	2.01	2.02	4.11	2.10
$S = 2$	$\hat{\sigma}^2_{MM}$	$\hat{\sigma}^2_{MD}$	$\hat{\sigma}^2_{SMM}$	$\hat{\sigma}^2_{qMCMD}$
$n \times variance$	2.00	2.02	3.14	2.02
$n \times bias$	-1.11	-0.11	1.29	-1.71
$n \times MSE$	2.01	2.02	3.14	2.04
$S = 20$	$\hat{\sigma}^2_{MM}$	$\hat{\sigma}^2_{MD}$	$\hat{\sigma}^2_{SMM}$	$\hat{\sigma}^2_{qMCMD}$
$n \times variance$	2.00	2.02	2.10	1.95
$n \times bias$	-1.11	-0.11	-0.12	-2.22
$n \times MSE$	2.01	2.02	2.10	1.97

Quasi-Monte Carlo Estimation: Dynamic Models

Common Application: Option Pricing

- Value of an option at time T: $P_T = e^{-rT}\mathbb{E}(H(S_T))$
  - $S_t =$ stock price at t
  - $H =$ some function
- (quasi)-Monte Carlo approach:
  - simulate n (large) paths of $(S_t)_{t\in[0,T]}$
  - approximate the expectation by computing a sample mean
  - use $d = T \Rightarrow$ curse of dimensionality
  - not an issue there because they need n very large anyways...
- For our setup: would like S fixed; not large
- The dynamics are going to be an issue if you are not careful

Dynamic Models: Setting

- Dynamic Models

$$y_t = g_{obs}(y_{t-1}, \theta, z_t)$$
$$z_t = g_{latent}(z_{t-1}, \theta, u_t), \quad u_t \overset{iid}{\sim} \mathcal{U}_{[0,1]^d}$$

- Consider the sample moments:

$$\hat{\psi}_n = \frac{1}{n}\sum_{t=1}^{n} \psi(y_t, \dots, y_{t-L})$$
$$= \frac{1}{n}\sum_{t=1}^{n} \psi_t(e_t, \dots, e_1, z_0, y_0)$$

- Several issues:
    - $\mathbb{E}(\hat{\psi}_n)$ involves a $n-$dimensional integral over $(u_n, \dots, u_1)$
    - (z_0, y_0) fixed $\Rightarrow$ non-stationarity (integral changes with $t\dots$)

Dynamic Models: things that do not work

- Even m-dependent models (simplest):

$$y_t = g_{obs}(u_t, \ldots, u_{t-m}, \theta), \quad u_t \overset{iid}{\sim} \mathcal{U}_{[0,1]^d}$$

  are not straightforward because of the overlap (lags)

- **Brute-force approach** (don't do this); suppose lags do not matter much:

$$\|\psi_t(u_t, \ldots, u_1, z_0, y_0, \theta) - \psi_t(u_t, \ldots, u_{t-m}, 0, \ldots, 0, \theta)\| \leq C_1^m$$

  up to a constant, for some $C_1 \in [0, 1)$

- And suppose that:

$$\left\| \frac{1}{n} \sum_{t=1}^{n} \psi_t(e_t, \ldots, e_{t-m}, 0, \ldots, 0, \theta) - \mathbb{E}(\psi_t(e_t, \ldots, e_{t-m}, 0, \ldots, 0, \theta)) \right\|$$

$$\leq C_2(\psi) \times C_3(m) \log(n)^m n^{-1}$$

- Use $m-$dimensional Sobol sequence: $C_3(m) = 2^m$

Dynamic Models: things that do not work

- Can't balance out the rates: $m = \alpha \times \frac{\log(n)}{\log\log(n)+\log(2)}$ for some well chosen $\alpha > 0$
- Yields rates that are always slower than $\sqrt{n}$
  - either the bias dominates (m small)
  - or the variance is too large (m large)
- That's because of $\log(n)^m$: curse of dimensionality
- Gets worse for $C_1 \simeq 1$ - very persistent DGPs
- So... back to Monte Carlo?... Not necessarily

- Simple trick: suppose we know (i.e. can draw from) the stationary distribution $f(y_t, z_t; \theta)$
- Consider moments of the form:

$$\hat{\psi}_n = \frac{1}{n} \sum_{t=1}^{n} \psi(y_t, \ldots, y_{t-L}), \quad \text{L fixed}$$

- Draw $(y_t^L, z_t^L) \overset{iid}{\sim} f(y_t, z_t; \theta)$ and simulate independent short time-series, then we have:

$$(y_t^1, \ldots, y_t^L) \overset{iid}{\sim} f(y_t, \ldots, y_{t-L}; \theta)$$

- Only depends on $L-1$ iid $\mathcal{U}_{[0,1]^d}$ shocks $(u_t^1, \ldots, u_t^{L-1})$

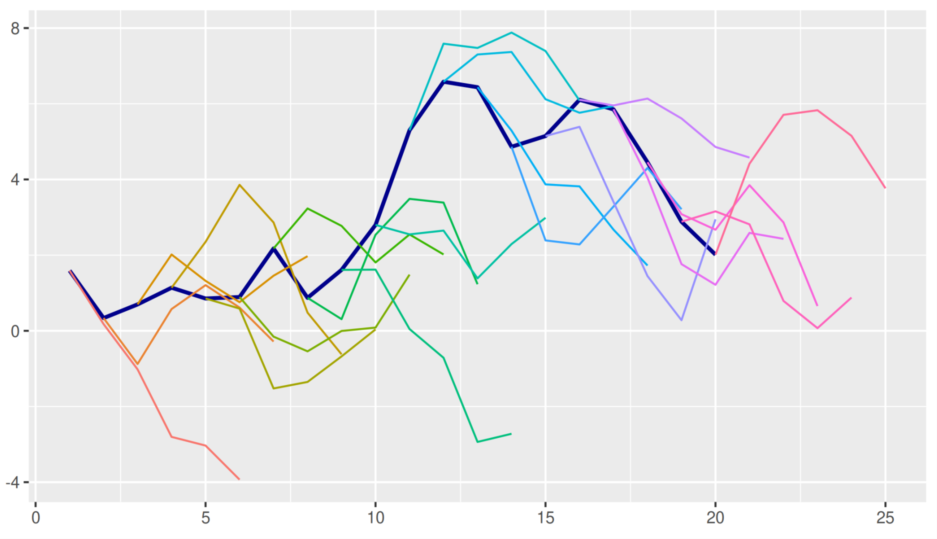
Getting around the curse of dimensionality

- We can use $d = L$ and simulate $n \times S$ short time-series
- We are within the quasi-Monte Carlo framework again:

$$\mathbb{E}(\hat{\psi}_n^s(\theta)) = \int_{\mathbf{u}\in[0,1]^{L-1},y,z} \psi \circ g_L(\mathbf{u}, y, z; \theta) f(y, z) dydzd\mathbf{u}$$

- The difficulty is when $f(y_t, z_t; \theta)$ is not known
  - We can simulate $(y_t, z_t)_{t=1,\ldots,n}$ once with Monte-Carlo and then simulate mini-samples with qMC
  - Computationally: simulate more than MC alone, but qMC simulations can be done in parallel...
- Typically (iid case) yields variance: $V_{qMC} \leq V_{MC/qMC} \leq V_{MC}$ (see e.g. Ökten et al., 2006; Buchholz and Chopin, 2017)

Illustration



Algorithm

- Construct $L \times d$-dimensional sequence $(\tilde{u}_t^1, \ldots, \tilde{u}_t^L) \in [0,1]^L$
- Draw $(\tilde{y}_t^0, \tilde{z}_t^0) \sim f(y_t, z_t; \theta)$ for $t = 1, \ldots, n \times S$
- Simulate $n \times S$ short samples:

$$\tilde{y}_t^\ell = g_{obs}(\tilde{y}_{t-1}^\ell, \theta, \tilde{z}_t^\ell)$$
$$\tilde{z}_t^\ell = g_{latent}(\tilde{z}_{t-1}^\ell, \theta, \tilde{u}_t^\ell), \ell = 1, \ldots, L$$

- And compute:

$$\begin{aligned}\hat{\psi}_n^S(\theta) &= \frac{1}{nS} \sum_{t=1}^{nS} \psi(\tilde{y}_t^L, \ldots, \tilde{y}_t^1) \\ &= \frac{1}{nS} \sum_{t=1}^{nS} \psi(\tilde{u}_t^1, \ldots, \tilde{u}_t^L, \tilde{y}_t^0, \tilde{z}_t^0; \theta)\end{aligned}$$

- if $\tilde{y}_t^0$ not iid $\Rightarrow$ complications in the asymptotics (when dependence is combined with qMC)

Scrambled Method of Moments

- If we can simulate directly from the stationary distribution: can apply the Static model theorem
- Otherwise: Theorem in progress
- Complications:
  - need a CLT for MC-qMC with dependence
  - initial value (y_0, z_0) bias as in Duffie and Singleton (1993)
- Can extend the results to some indirect inference estimators
  - if auxiliary parameters computed from moments with L-lags
  - e.g. $AR(p)$ OLS estimates as moments for $ARMA(1,1)$ model

Simple Example: ARMA(1,1)

- Simple time-series model:

$$y_t = \rho y_{t-1} + \sigma[e_t + \theta e_{t-1}], e_t \overset{iid}{\sim} \mathcal{N}(0,1)$$

here: $e_t = \Phi^{-1}(u_t)$, $u_t \overset{iid}{\sim} \mathcal{U}_{[0,1]}$, $z_t = (e_t, e_{t-1})$

- SMM/qMCMD estimator, auxiliary model:

$$y_t = \beta_1 y_{t-1} + \cdots + \beta_p y_{t-1} + \tilde{e}_t$$

auxiliary moments: $\hat{\psi}_n = (\hat{\beta}_1, \ldots, \hat{\beta}_p, \hat{\sigma}_{\tilde{e}})$ + efficient W

- Two experiments for (y_{t-L-1}, e_{t-L-1}):
    1. draw from the stationary distribution directly (qMC)
    2. simulate a time-series and use simulated pairs (MC-qMC)

SMM vs. MLE vs. qMCMD – $p = 4$ Lags

Table 1: $\sqrt{n}\times$ standard deviation

$S=1$	MLE	SMM	qMCMD-1	qMCMD-2
$\hat{\rho}$	1.09	1.62	1.20	1.43
$\hat{\theta}$	1.13	1.89	1.35	1.55
$\hat{\sigma}$	0.72	1.05	0.76	0.97

Table 2: $\sqrt{n}\times$ standard deviation

$S=2$	MLE	SMM	qMCMD-1	qMCMD-2
$\hat{\rho}$	1.09	1.41	1.16	1.28
$\hat{\theta}$	1.13	1.57	1.27	1.45
$\hat{\sigma}$	0.72	0.93	0.74	0.82

Table 3: $\sqrt{n}\times$ standard deviation

SMM vs. MLE vs. qMCMD – $p = 12$ Lags

Table 4: $\sqrt{n}\times$ standard deviation

$S = 1$	MLE	SMM	qMCMD-1	qMCMD-2
$\hat{\rho}$	1.09	1.86	1.30	1.49
$\hat{\theta}$	1.13	1.85	1.37	1.46
$\hat{\sigma}$	0.72	1.10	0.81	0.93

Table 5: $\sqrt{n}\times$ standard deviation

$S = 2$	MLE	SMM	qMCMD-1	qMCMD-2
$\hat{\rho}$	1.09	1.57	1.29	1.35
$\hat{\theta}$	1.13	1.60	1.34	1.32
$\hat{\sigma}$	0.72	0.94	0.74	0.83

Table 6: $\sqrt{n}\times$ standard deviation

SMM vs. MLE vs. qMCMD – $p = 4$ Lags

Table 7: $\sqrt{n}\times$ standard deviation

$S = 1$	MLE	SMM	qMCMD-1	qMCMD-2
$\hat{\rho}$	0.54	0.77	0.56	0.85
$\hat{\theta}$	0.94	1.53	1.09	1.13
$\hat{\sigma}$	0.68	1.03	0.75	0.82

Table 8: $\sqrt{n}\times$ standard deviation

$S = 2$	MLE	SMM	qMCMD-1	qMCMD-2
$\hat{\rho}$	0.54	0.63	0.50	0.71
$\hat{\theta}$	0.94	1.34	1.06	1.06
$\hat{\sigma}$	0.68	0.93	0.73	0.78

Table 9: $\sqrt{n}\times$ standard deviation

Dynamic Models: Some Conclusions

- As usual (near) unit-roots do not mix well with SMM (see e.g. Phillips, 2012)
- Here it seems to affect the MC-qMC properties: be careful

Covariates, MC-qMC and CLTs

Introducing Covariates - Static Models

- Suppose we are interested in moments of the form:

$$\hat{\psi}_n^s(\theta) = \frac{1}{n}\sum_{t=1}^{n}\psi(u_t, x_t;\theta)$$

- If $x_t \overset{iid}{\sim} f_x$, this is very similar to MC-qMC
- Idea similar to MC-qMC (Ökten et al., 2006):

$$\hat{\psi}_n^s(\theta) - \mathbb{E}(\hat{\psi}_n^s(\theta)) =$$

$$\underbrace{\frac{1}{n}\sum_{t=1}^{n}[\psi(u_t, x_t;\theta) - \mathbb{E}(\psi(u_t, x_t;\theta)|u_t)]}_{\text{non iid CLT}}$$

$$+\underbrace{\frac{1}{n}\sum_{t=1}^{n}[\mathbb{E}(\psi(u_t, x_t;\theta)|u_t) - \mathbb{E}(\psi(u_t, x_t;\theta))]}_{\text{Standard qMC: } o_p(n^{-1/2})}$$

Introducing Covariates - Static Models

- The main difficulty is in deriving a CLT for:

$$\frac{1}{n}\sum_{t=1}^{n}[\psi(u_t, x_t;\theta) - \mathbb{E}(\psi(u_t, x_t;\theta)|u_t)]$$

- Need a Lindeberg/Lyapunov condition using

$$s_n^2 = \sum_{t=1}^{n}\mathbb{E}\left(\|\psi(u_t, x_t;\theta) - \mathbb{E}(\psi(u_t, x_t;\theta)|u_t)\|^2 \Big| u_t\right)$$

- Since u_t is a low-discrepancy sequence, we have:

$$s_n^2/n \to \mathbb{E}\left(\mathbb{E}\left(\|\psi(u_t, x_t;\theta) - \mathbb{E}(\psi(u_t, x_t;\theta)|u_t)\|^2 \Big| u_t\right)\right) = \sigma > 0$$

Introducing Covariates - Static Models

- Suppose the sum:

$$0 < \lim_{n\to\infty} \frac{1}{n} \sum_{t=1}^{n} \mathbb{E}\left(\|\psi(u_t, x_t; \theta) - \mathbb{E}(\psi(u_t, x_t; \theta)|u_t)\|^{2+\delta} \Big| u_t \right) < +\infty$$

- for some $\delta > 0$, then we have:

$$\frac{\sum_{t=1}^{n} \mathbb{E}\left(\|\psi(u_t, x_t; \theta) - \mathbb{E}(\psi(u_t, x_t; \theta)|u_t)\|^{2+\delta} \Big| u_t \right)}{s_n^{2+\delta}} = O(n^{-\delta/2}) \to 0$$

- This implies a CLT, i.e. conditional on $u_{1:n}$:

$$\sqrt{n}\left(\hat{\psi}_n^s(\theta) - \mathbb{E}[\hat{\psi}_n^s(\theta)|u_{1:n}] \right) \Big| u_{1:n} \xrightarrow{d} \mathcal{N}(0, V_{MC-qMC})$$

- V_{MC-qMC} is smaller than the usual V_{MC} (Ökten et al., 2006; Buchholz and Chopin, 2017)

Central Limit Theorem

- Recall that we split the sum into:
  - a non iid sample mean in $x_t \Rightarrow$ CLT
  - an iid sample mean only in $u_t \Rightarrow o_p(n^{-1/2})$
- Hence an overall CLT applies with 2nd term negligible
- Focused on $S = 1$, $S > 1$ need to ensure $u_t \perp\!\!\!\perp x_t$ holds
  - scramble: ok
  - deterministic sequence: shuffle the x_t
- For dynamic models: similar idea with non-iid, non-stationary dependent sequences (in progress)
- Covariates are problematic in dynamic models without further assumptions
  - need to draw triplets: (y_t, x_t, z_t)
  - and simulate paths: $(y_t^\ell, x_t^\ell, z_t^\ell)$

Conclusion

Conclusion

- Static models:
  - should always use Scrambled Method of Moments or qMCMD
- Dynamic models:
  - most empirical applications rely on sample variances, covariances - could use SMM/qMCMD instead
- Dynamic Panel Data models:
  - fixed T (small): see static models
  - increasing T: see dynamic models

Thank You!

References

Buchholz, A. and Chopin, N. (2017). Improving approximate Bayesian computation via quasi-Monte Carlo.

Dick, J. and Pillichshammer, F. (2010). *Digital Nets and Sequences*. Cambridge University Press, Cambridge.

Duffie, D. and Singleton, K. J. (1993). Simulated Moments Estimation of Markov Models of Asset Prices. *Econometrica*, 61(4):929.

Gallant, a. R. and Tauchen, G. (1996). Which Moments to Match? *Econometric Theory*, 12(04):657.

Gerber, M. and Chopin, N. (2015). Sequential quasi Monte Carlo. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 77(3):509–579.

Gerber, M. and Chopin, N. (2017). Convergence of sequential quasi-Monte Carlo smoothing algorithms. *Bernoulli*, 23(4B):2951–2987.

Gouriéroux, C., Monfort, A., and Renault, E. (1993). Indirect inference. *Journal of Applied Econometrics*, 8(S1):S85—-S118.

References ii

Lemieux, C. (2009). *Monte Carlo and Quasi-Monte Carlo Sampling*. Springer Series in Statistics. Springer New York, New York, NY.

McFadden, D. (1989). A Method of Simulated Moments for Estimation of Discrete Response Models Without Numerical Integration. *Econometrica*, 57(5):995.

Ökten, G., Tuffin, B., and Burago, V. (2006). A central limit theorem and improved error bounds for a hybrid-Monte Carlo sequence with applications in computational finance. *Journal of Complexity*, 22(4):435–458.

Owen, A. B. (1997). Scrambled net variance for integrals of smooth functions. *The Annals of Statistics*, 25(4):1541–1562.

Pakes, A. and Pollard, D. (1989). Simulation and the Asymptotics of Optimization Estimators. *Econometrica*, 57(5):1027.

Paskov, S. H. and Traub, J. F. (1995). Faster Valuation of Financial Derivatives. *The Journal of Portfolio Management*, 22(1):113–123.

Phillips, P. C. (2012). Folklore Theorems, Implicit Maps, and Indirect Inference. *Econometrica*, 80(1):425–454.