Applications of numerical integration methods to high-dimensional econometric parameter estimation

Alexandros Gilch

December 5, 2019

Institut für Numerische Simulation
Rheinische Friedrich-Wilhelms-Universität Bonn

Motivation

- Many econometric models require the computation of high-dimensional integrals which have no analytic solution
- Commonly used techniques include Monte Carlo and Quasi Monte Carlo simulations of those integrals
- Mathematics provide quadrature rules with better convergence rates but which require more knowledge about the integrands
- Econometric models are rarely classified in a general form possible

$\Rightarrow$ Cannot give general statement about the applicability of advanced quadrature rules.
$\Rightarrow$ Need to define econometric model classes and corresponding mathematical function spaces to provide guidelines when to use which quadrature rule.

Motivation for this talk

- Present modern techniques for the approximation of high-dimensional integrals
- Get feedback about the econometric foundations

Econometric models

Two examples from Discrete choice theory

The Probit model

Want to find choice probabilities for a decision maker a facing m alternatives

$$\Pr(y(a)=i \mid X) = \Phi(\eta_i) \qquad \text{where } \eta_i = X_i\beta \in \mathbb{R}^m$$

where $y = y(a) \in \{1, ..., m\}$ is the decision, $X \in \mathbb{R}^{m \times q}$ are exogenous variables and $\beta \in \mathbb{R}^q$ is a vector of parameters.
We assume normally distributed dependent errors $\varepsilon \sim \mathcal{N}(0, \Sigma)$ and hence get a multivariate normal CDF Φ in m dimensions which cannot be computed analytically:

$$\Phi(\eta_i) = \int_{\{\varepsilon \leq \eta_i\}} \exp\left(-\frac{1}{2}\varepsilon^T \Sigma^{-1} \varepsilon\right) d\varepsilon \ .$$

The Mixed Logit Model

Again: Want to find choice probabilities for a decision maker a facing m alternatives

$$\Pr(y(a) = i \mid X, Z, U(a)) = \frac{\exp(\eta_i)}{\sum_{j=1}^{m} \exp(\eta_j)}.$$

But now we set

$$\eta = X\beta + ZU \in \mathbb{R}^m$$

with exogenous variables $X, Z \in \mathbb{R}^{m \times q}$, a fixed parameter vector $\beta \in \mathbb{R}^q$ and a random parameter vector $U(a) \in \mathbb{R}^q$, $U \sim h(u|\Sigma)$. This leads to

$$Pr(y(a) = i | X, Z) = \int_{\mathbb{R}^q} \Pr(y(a) = i \mid X, Z, u) h(u|\Sigma) du \ .$$

Generalized Linear Mixed Models

We can define a common framework for these models. The *generalized linear mixed model (GLMM)* is specified by

- a distribution from an exponential family $Y|u \sim f(y|u, \theta), Y \in \mathbb{R}^m$,
- a linear mixed predictor $\eta := X\beta + ZU \in \mathbb{R}^m$,
- a function $\tilde{g}$ relating mean to the parameter of the exponential family $\theta = \tilde{g}(\mu) \in \mathbb{R}^m$,
- a link function g s.t. $\mathbf{E}[Y|u] = \mu = g^{-1}(\eta) \in \mathbb{R}^m$,
- a distribution for the random effects $U \sim h(u|\Sigma), U \in \mathbb{R}^q$.

Generalized Linear Mixed Models

Computing the maximum likelihood requires the approximation of

$$L(\beta, \Sigma|\mathbf{y}) = \int_{\mathbb{R}^q} \exp(\mathbf{y} \cdot \tilde{g}(g^{-1}(\eta)) - b(\tilde{g}(g^{-1}(\eta)))h(u|\Sigma)du \ .$$

This cannot be computed analytically!
Also, g^{-1} might already include the approximation of an integral.
$\Rightarrow$ Which quadrature rule to use?

Quadrature rules

A short overview over some approximation techniques

1D Integration

Numerical integration in one dimension is well studied and the best approximation rules in terms of rate, exactness, nestedness and computability are known.
Consider the approximation of an integral on $[0,1]$

$$Q_N(f) := \sum_{n=1}^{N} w_n f(x_n) \approx \int_0^1 f(x)dx =: \mathcal{I}(f) \ .$$

We evaluate $Q_N(f)$ in terms of its convergence rate $g(N) \to 0$ for the error

$$\varepsilon_N := |Q_N(f) - \mathcal{I}(f)| = O(g(N)) \ .$$

1D Integration

The convergence rates require f to be in a certain function space, assuring regularity and boundary conditions. For $f \in C^r([0,1])$, often used quadrature rules are:

Rule	Nodes and weights	Rate
Trapez	$x_n = \frac{n-\frac{1}{2}}{N}, w_n = \frac{1}{N}$	$O(N^{-2})$
Gauß-Legendre	$x_n =$zeros of Legendre polynomials, w_n depending on x_n	$O(N^{-r})$
Clenshaw-Curtis	$x_n = \frac{n\pi}{N}$, w_n depending on x_n	$O(N^{-r})$

From one dimension to many dimensions

Now consider

$$\mathcal{I} := \int_{[0,1]^d} f(x)dx \approx Q_N^d(f)$$

Simplest approach to extend known rules:

$$(Q_N \otimes \cdots \otimes Q_N)(f) = \sum_{n_1=1}^{N} \cdots \sum_{n_d=1}^{N} w_{n_1} \cdot \ldots \cdot w_{n_d} f(x_{n_1}, \cdots, x_{n_d})$$

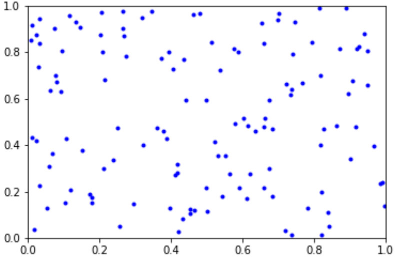
$\Rightarrow M = N^d$ grid points.
The *product rule* achieves the same rate as the univariate rule

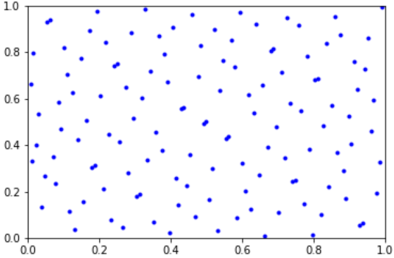
$$O(N^{-s}) = O(M^{-\frac{s}{d}})$$

but needs much more points. This is called the *curse of dimensionality*.

Monte Carlo and Quasi Monte Carlo Integration

MC chooses the points randomly in the domain and achieves only a probabilistic rate $Var(Q_N(f)) = Var(f)N^{-\frac{1}{2}}$. It requires $f \in L_2([0,1]^d)$.
QMC takes points from a *low-discrepancy sequence*, e.g. the Halton- or Sobol-sequence. It requires $f \in H^1([0,1]^d)$ and achieves the rate $O(\log(N)^d N^{-1})$.





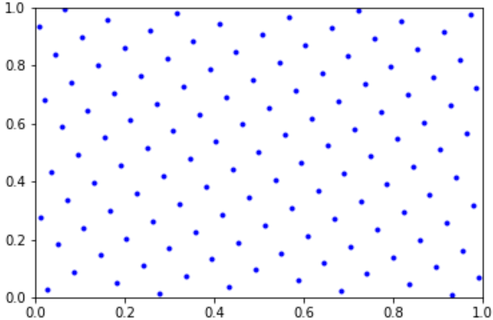
Frolov points

Frolov quadrature is a lattice rule:

$$Q_N^d(f) = \frac{1}{N} \sum_{\mathbf{k} \in \mathbb{Z}^d} f(\mathbf{A}_N \mathbf{k})$$

It is universal and optimal, but only for functions with zero boundary:

$$f \in H_0^r([0,1]^d) \Rightarrow \varepsilon_N = O(log(N)^{\frac{d-1}{2}} N^{-r})$$



Sparse grids

The *sparse grid* approach try to fix the product rule by choosing only the subgrids which supply the most information.
We compute the difference quadrature rules

$$\Delta_l(f) := Q_l(f) - Q_{l-1}(f)$$

(we set $Q_l(f) = \sum_{n=1}^{2^l} ...$) and use the identity

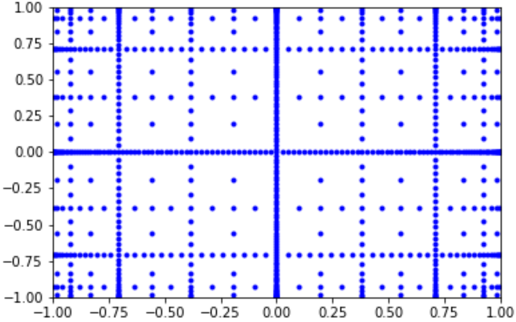
$$\sum_{k=1}^{l} \Delta_k(f) = Q_l(f) \ .$$

We sum over $k_1 + \cdots + k_d \leq l$ instead of $\max\{k_1, ..., k_d\} \leq l$:

$$Q_l^d(f) = \sum_{k_1+\cdots+k_d \leq l} \Delta_{k_1} \otimes \cdots \otimes \Delta_{k_d}(f) \ .$$

Sparse grids

This leads to $O(N \log(N)^{d-1})$ instead of $O(N^d)$ points.
For $f \in H^r_{\text{mix}}$ we can achieve a rate of $O(N^{-r} \log(N)^{(d-1)r})$.

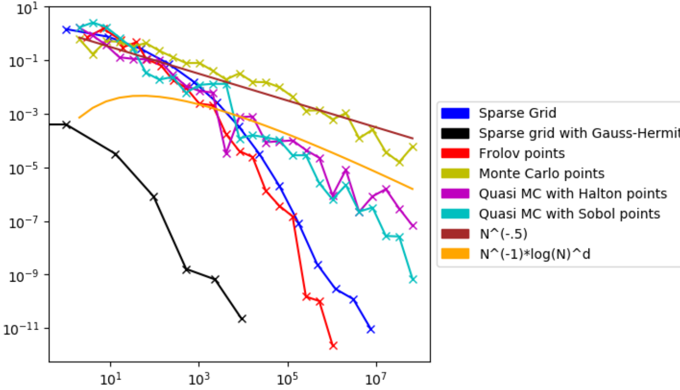


First results

Applying different quadrature rules to some models

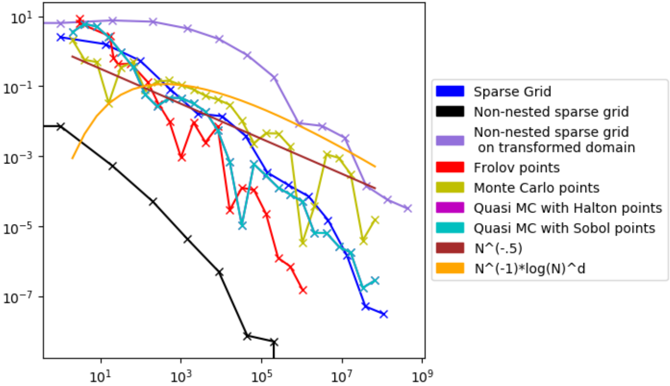
First results

Mixed Logit model with Gaussian random effects for dim = 4

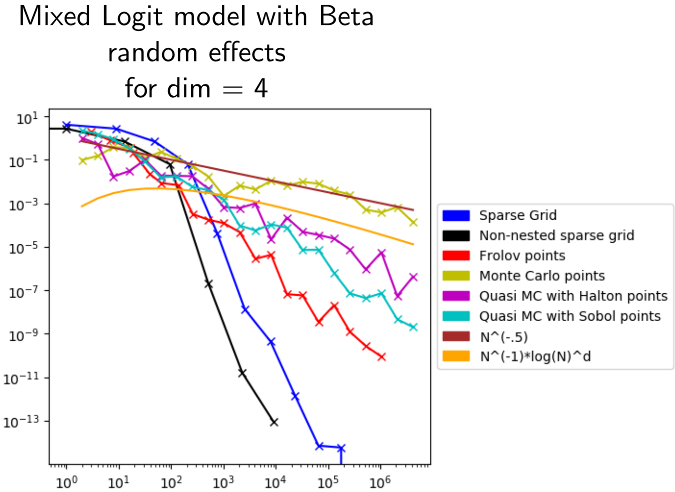


First results

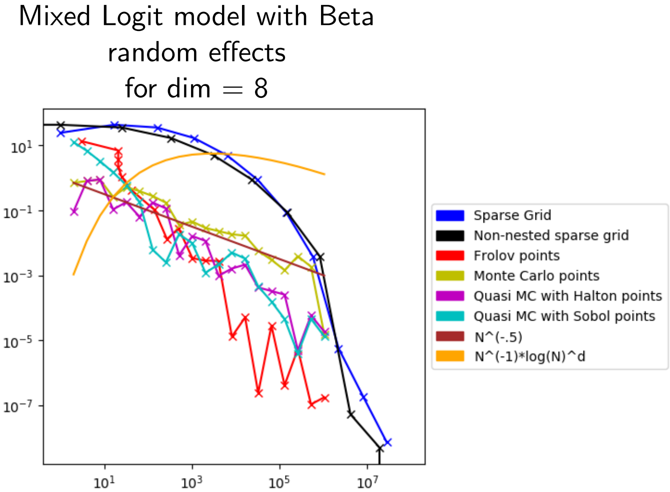
Mixed Logit model with Gaussian random effects for dim = 6



First results

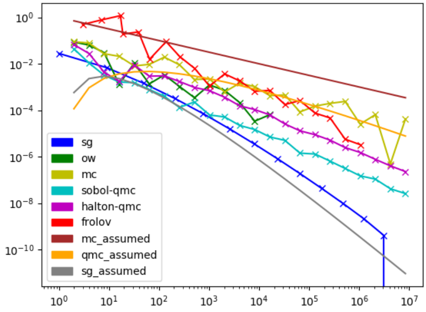


First results

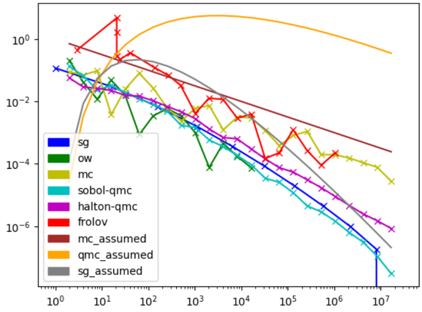


First results

Multinomial probit model
for dim = 4



Multinomial probit model
for dim = 8



Conclusion

Before applying those rules, one needs to check:

- The regularity of the integrand
- Boundary conditions
- The domain (and how transforming the integrand changes the first two points)

Even then there might be problems like

- preasymptotics (when does the rate actually start?)
- high constants in the O-notation.

$$\Rightarrow \text{Work in Progress!}$$

Thank you for your attention!