Lecture 6 – Simulation-based methods for nonlinear models

Economics 8379
George Washington University

Instructor: Prof. Ben Williams

Simulation-based estimation methods

MSL example

MSM example

How do we generate random numbers anyway?

Importance sampling

Conclusion

traditional MLE and GMM

- MLE:

$$\mathcal{L}(\theta) = \log f^{\theta}_{y_1,\ldots,y_n|x_1,\ldots,x_n} = \sum_{i=1}^{n} \log(f^{\theta}_{y_i|x_i})$$

  - Computation of the MLE involves evaluation the likelihood (and possibly it's derivatives) iteratively for many values of θ.
  - This is difficult when $f^{\theta}_{y_i|x_i}$ is difficult to compute.

traditional MLE and GMM

- GMM based on moments $E(w_i m(y_i, x_i, \theta)) = 0$:
  - The GMM estimator minimizes

$$\left(\sum_{i=1}^{n} w_i m(y_i, x_i, \theta)\right)' W \left(\sum_{i=1}^{n} w_i m(y_i, x_i, \theta)\right)$$

  - Computation involves evaluating this objective function (and possibly it's derivatives) iteratively for many values of θ.
  - This is difficult when $m(y_i, x_i, \theta)$ is difficult to compute.

MLE vs GMM

- Two reasons $f_{y_i|x_i}$ or $m(y_i, x_i, \theta)$ can be difficult to compute:
  - latent variable: $f^{\theta}_{y_i|x_i} = \int f^{\theta}_{y_i|x_i,u} f_u du$
  - y_i is determined conditional on x_i and unobserved shock(s) via an economic model which may involve dynamic optimization, solution of a nash equilibrium, etc.

Some examples

- Multinomial probit:

$$\ell(\beta) = \sum_{i=1}^{n} \sum_{j=1}^{J} \mathbf{1}(y_i = j) \log(Pr(y_i = j \mid X_i))$$

where

$$Pr(y_i = j \mid X_i) = Pr(X_{ij}'\beta + \varepsilon_{ij} \geq \max_{l \neq j} X_{il}'\beta + \varepsilon_{il})$$

and $\varepsilon_i = (\varepsilon_{i1}, \ldots, \varepsilon_{iJ}) \sim N(0, \Sigma)$

Some examples

- Random coefficients logit: same form for likelihood with

$$Pr(y_i = j \mid X_i) = \int \frac{\exp(X_{ij}'\beta_i)}{\sum_{l=1}^{J} \exp(X_{il}'\beta)} f(\beta_i) d\beta_i$$

and $\beta_i \sim N(\bar{\beta}, \Sigma_\beta)$

Some examples

- Random coefficients logit: same form for likelihood with

$$Pr(y_i = j \mid X_i) = \int \frac{\exp(X'_{ij}\beta_i)}{\sum_{l=1}^{J} \exp(X'_{il}\beta)} f(\beta_i) d\beta_i$$

  and $\beta_i \sim N(\bar{\beta}, \Sigma_\beta)$
- Both of these can allow for a choice-invariant regressor with a choice-specific coefficient as well ($\gamma'_j w_i$).

Some examples

- Dynamic discrete choice models.
  - Given state variables $\{x_{it}, \varepsilon_{it}\}$ agent i chooses control variables $\{y_{it}\}$ to maximize $E\left(\sum_{t=0}^{\infty} \beta^t (u(x_{it}, y_{it}, \theta) + \varepsilon_{it}) \mid x_{i0}, \varepsilon_{i0}\right)$

Some examples

- Dynamic discrete choice models.
  - Given state variables $\{x_{it}, \varepsilon_{it}\}$ agent *i* chooses control variables $\{y_{it}\}$ to maximize $E\left(\sum_{t=0}^{\infty} \beta^t (u(x_{it}, y_{it}, \theta) + \varepsilon_{it}) \mid x_{i0}, \varepsilon_{i0}\right)$
  - There is a Bellman equation solution and if y_{it} is binary, Rust (1987) provides conditions under which

$$Pr(y_{it} = 1 \mid x_{it}, \theta) = \frac{\exp(u(x_{it}, 1, \theta) + \beta EV(x_{it}, 1, \theta))}{\exp(u(x_{it}, 0, \theta) + \beta EV(x_{it}, 0, \theta)) + \exp(u(x_{it}, 1, \theta) + \beta EV(x_{it}, 1, \theta))}$$

where

$$EV(x, y, \theta) = \int \log\left(\sum_{y'=0,1} \exp(u(x', y', \theta) + \beta EV(x', y', \theta))\right) p(dx' \mid x, y, \theta)$$

Some examples

- Dynamic discrete choice models.
  - Given state variables $\{x_{it}, \varepsilon_{it}\}$ agent *i* chooses control variables $\{y_{it}\}$ to maximize $E\left(\sum_{t=0}^{\infty} \beta^t (u(x_{it}, y_{it}, \theta) + \varepsilon_{it}) \mid x_{i0}, \varepsilon_{i0}\right)$
  - There is a Bellman equation solution and if y_{it} is binary, Rust (1987) provides conditions under which

$$Pr(y_{it} = 1 \mid x_{it}, \theta) = \frac{\exp(u(x_{it}, 1, \theta) + \beta EV(x_{it}, 1, \theta))}{\exp(u(x_{it}, 0, \theta) + \beta EV(x_{it}, 0, \theta)) + \exp(u(x_{it}, 1, \theta) + \beta EV(x_{it}, 1, \theta))}$$

where

$$EV(x, y, \theta) = \int \log\left(\sum_{y'=0,1} \exp(u(x', y', \theta) + \beta EV(x', y', \theta)\right) p(dx' \mid x, y, \theta)$$

  - We need to solve for the *expected value function*, EV, to evaluate the likelihood.

Some examples

- Dynamic discrete choice models.
  - Solving for the expected value function involves an approximation.
  - The method of simulated moments is an alternative.

Maximum Simulated Likelihood

- Suppose that $f(y_i \mid X_i, \theta) = \int g(y_i \mid X_i, u, \theta)\psi(u)du$.
- simulate $u_{i1}, \ldots, u_{iS} \sim_{i.i.d.} \psi(\cdot)$ for each i and replace $\ell_i(\theta) = \log(f(y_i \mid X_i, \theta))$ with

$$\hat{\ell}_i(\theta) = \log\left(S^{-1}\sum_{s=1}^{S} g(y_i \mid X_i, u_{is}, \theta)\right)$$

- then $\hat{\theta}_{MSL} = \arg\max_\theta \sum_{i=1}^{n} \hat{\ell}_i(\theta)$.

Maximum Simulated Likelihood

- Only consistent and asymptotically normal if $\sqrt{n}/S \to 0$. Take S as a multiple of the sample size if feasible.
- do not draw new simulation sample in each iteration of the optimization routine!
- Sometimes this naive simulation can be improved by importance sampling and other variance-reduction techniques. See 12.7 in CT.

Method of Simulated Moments

- Suppose we want to estimate θ based on the moment condition: $E(w_i m(y_i, x_i, \theta_0)) = 0$
- where computing $m(y_i, x_i, \theta) = \int h(y_i, x_i, u, \theta)\psi(u)du$ requires simulation
- The MSM estimator is computed by following these steps:
  - draw u_{is}, $s = 1, \ldots, S$ independently from ψ for each i
  - and compute $\hat{m}(y_i, x_i, \theta) = S^{-1} \sum_{s=1}^{S} h(y_i, x_i, u_{is}, \theta)$
  - minimize

$$\left(\sum_{i=1}^{n} w_i \hat{m}(y_i, x_i, \theta) \right)' W \left(\sum_{i=1}^{n} w_i \hat{m}(y_i, x_i, \theta) \right)$$

Method of Simulated Moments

- Because $E(\hat{m}(y_i, x_i, \theta) \mid y_i, x_i) = m(y_i, x_i, \theta)$ (unbiased simulation), if the usual GMM conditions are satisfied then the MSM estimator is a consistent, asymptotically normal estimator

Method of Simulated Moments

- Because $E(\hat{m}(y_i, x_i, \theta) \mid y_i, x_i) = m(y_i, x_i, \theta)$ (unbiased simulation), if the usual GMM conditions are satisfied then the MSM estimator is a consistent, asymptotically normal estimator
- if, in addition, $S \to \infty$, then the estimator is asymptotically equivalent to the GMM estimator
- for finite S, the asymptotic variance is inflated by a factor of $1 + S^{-1}$, though this can be improved, e.g. by importance sampling

Method of Simulated Moments

- Variance estimation requires either simulation or bootstrap
- Gourieroux and Monfort (1991) provide more general conditions under which $S \to \infty$ is not necessary
- Pakes and Pollard (1989) provide some examples.

Indirect inference

- Ingredients:
  - economic model, e.g., $y_i = G(X_i, u_i; \beta)$ for $i = 1, \ldots, n$ and $u_i \sim_{iid} F_u$
  - auxiliary model, e.g., a likelihood: $\ell_n(\theta) = \sum_{i=1}^{n} \log(f(y_i \mid X_i, \theta))$
  - an auxiliary estimate, e.g., $\hat{\theta} = \arg\max_\theta \ell_n(\theta)$

Indirect inference

- For any value of β,
  - simulate $\{y_i^m(\beta)\}$ from the economic model for m=1,...,M
  - obtain $\tilde{\theta}(\beta)$ by maximizing

$$\sum_{m=1}^{M}\sum_{i=1}^{n}\log(f(y_i^m(\beta) \mid X_i, \theta)$$

- Alternatively, get M different estimates, $\tilde{\theta}_1(\beta), \ldots, \tilde{\theta}_M(\beta)$ and use $\tilde{\theta}(\beta) = M^{-1}\sum_{m=1}^{M}\tilde{\theta}_m(\beta)$

Indirect inference

- The indirect inference estimator of β is given by

$$\hat{\beta} = \arg\min_{\beta} D(\hat{\theta}, \tilde{\theta}(\beta))$$

  - D is a metric function; Smith (2008) suggests Wald, LR, LM metrics
  - consistent and asymptotically normal for M fixed, $n \to \infty$
  - variance inflate by $1 + M^{-1}$
  - very easy to implement despite the lack of efficiency

Indirect inference

- The following are typical conditions required for indirect inference:
  - the economic model is correctly specified and well-behaved
  - the auxiliary likelihood function is well-behaved in the limit, despite the fact that it is misspecified
  - binding function
    - $\ell_n(\theta) \to_p \ell(\theta; \beta, F_u)$ when the data is generated by the economic model with parameters β and distribution F_u
    - define $\theta(\beta) = \arg\max_\theta \ell(\theta; \beta, F_u)$
    - $\theta_0 = \theta(\beta_0)$ is the pseudo-true value
    - $\theta(\beta)$ is the binding function

Indirect inference

- under appropriate regularity conditions, $\hat{\theta} \rightarrow_p \theta_0$ and $\tilde{\theta}(\beta) \rightarrow_p \theta(\beta)$

Indirect inference

- under appropriate regularity conditions, $\hat{\theta} \rightarrow_p \theta_0$ and $\tilde{\theta}(\beta) \rightarrow_p \theta(\beta)$
- thus the identification condition: is β_0 the unique solution to $\theta_0 = \theta(\beta)$?
- requires $dim(\theta) \geq dim(\beta)$
- simulation avoids needing to know the binding function

Simulation-based estimation methods

MSL example

MSM example

How do we generate random numbers anyway?

Importance sampling

Conclusion

Random coefficient logit model

- Consider the binary outcome model

$$Y_i = \mathbf{1}(\beta_0 + \beta_{1i} X_i + \varepsilon_i \geq 0)$$

  where
  - ε_i is iid logistic
  - $\beta_{1i} = \beta_1 + \sigma_{\beta_1} u_i$ where u_i is iid $N(0,1)$
- Then

$$Pr(Y_i = 1 \mid X_i) = \int \frac{\exp(\beta_0 + \beta_1 X_i + \sigma_{\beta_1} X_i u)}{1 + \exp(\beta_0 + \beta_1 X_i + \sigma_{\beta_1} X_i u)} \phi(u) du$$

MLE for the RL model

- Let $\theta = (\beta_0, \beta_1, \sigma_{\beta_1})$.
- The log likelihood is

$$\mathcal{L}(\theta) = \sum_{i=1}^{n} \log \left(\int \pi(Y_i \mid X_i, u, \theta) \phi(u) du \right)$$

where

$$\pi(y \mid x, u, \theta) = \begin{cases} \frac{\exp(\beta_0 + \beta_1 x + \sigma_{\beta_1} xu)}{1 + \exp(\beta_0 + \beta_1 x + \sigma_{\beta_1} xu)} & \text{if } y = 1 \\ \frac{1}{1 + \exp(\beta_0 + \beta_1 x + \sigma_{\beta_1} xu)} & \text{if } y = 0 \end{cases}$$

MSL for the RL model

- First, simulate $u_{i1}, \ldots, u_{iS} \sim_{i.i.d.} \psi(\cdot)$ for each i
- Let

$$\hat{\ell}_i(\theta) = \log\left(S^{-1}\sum_{s=1}^{S}\pi(Y_i \mid X_i, u_{is}, \theta)\right)$$

- Then $\hat{\theta}_{MSL} = \arg\max_\theta \sum_{i=1}^{n} \hat{\ell}_i(\theta)$.

MSL for the RL model

- MSL replaces choice probabilities,

$$\pi(y_i \mid X_i, \theta) = \int \pi(Y_i \mid X_i, u, \theta)\phi(u)du,$$

- with simulated choice probabilities:

$$\hat{\pi}(y_i \mid X_i, \theta) = S^{-1} \sum_{s=1}^{S} \pi(Y_i \mid X_i, u_{is}, \theta)$$

MSL for the RL model

- The estimated asymptotic variance can be derived using equation (12.21) in CT:

$$\hat{V} = \left(\sum_{i=1}^{n} \left(\hat{h}_i(\hat{\theta}) \hat{h}_i(\hat{\theta})' \right) \right)^{-1}$$

$$\hat{h}_i(\hat{\theta}) = \frac{(-1)^{1-Y_i} \sum_{s=1}^{S} W_{is} \pi(1 \mid X_i, u_{is}, \hat{\theta}) \pi(0 \mid X_i, u_{is}, \hat{\theta})}{\sum_{s=1}^{S} \pi(Y_i \mid X_i, u_{is}, \hat{\theta})}$$

where $W_i = (1, X_i, X_i u_{is})'$

demonstration of MSL for the RL model

- A simple Matlab code snippet:

```
function RLsimlik=RLsimlik(Y,X,U,theta)


RLsimlik=sum(log(mean(cp(Y,X,U,theta),2)));

function cp=cp(Y,X,U,theta)
n=length(Y);
S=size(U,2);
bet0=theta(1);
bet1=theta(2);
sig=theta(3);

cp=exp(   bet0*ones(n,S)+bet1*X*ones(1,S)+sig*(X*ones(1,S)).*U   ).^(Y*ones(1,S))...
    ./(ones(n,S)+exp(   bet0*ones(n,S)+bet1*X*ones(1,S)+sig*(X*ones(1,S)).*U   ))
```

demonstration of MSL for the RL model

- Then the estimator is computed, e.g., by

```
maxer=fmincon(@(x) -RLsimlik(Y,X,U,x),[0 1 .5],[],[],[],[],[-Inf -Inf 0],[]);
```

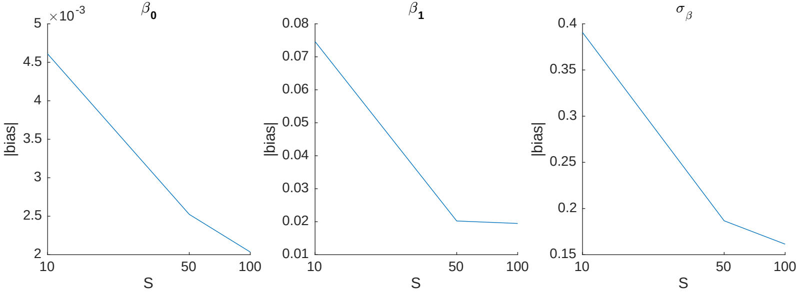
demonstration of MSL for the RL model

- Then the estimator is computed, e.g., by

```
maxer=fmincon(@(x) -RLsimlik(Y,X,U,x),[0 1 .5],[],[],[],[],[-Inf -Inf 0],[]);
```

demonstration of MSL for the RL model

The following graph shows bias as a function of S.



A nonseparable model

- Suppose $Y_i = g(X_i, V_i \mid \beta)$.
  - for example, $g(x, v \mid \beta) = x'\beta + v$
- Further suppose that V_i is independent of X_i and $V_i \sim N(0, \sigma_V^2)$.

GMM for the nonseparable model

- Let $\theta = (\beta, \sigma_V)$ and

$$m(y_i, x_i, \theta) = \begin{pmatrix} y_i - E(y_i \mid x_i, \theta) \\ x_i(y_i - E(y_i \mid x_i, \theta)) \\ y_i^2 - E(y_i^2 \mid x_i, \theta) \end{pmatrix}$$

- Then the moment conditions $E(m(y_i, x_i, \theta)) = 0$ can be used to estimate θ.

GMM for the nonseparable model

- If $g(x, v \mid \beta) = x'\beta + v$ then $E(y_i \mid x_i, \theta) = x_i'\beta$ and $E(y_i^2 \mid x_i, \theta) = (x_i'\beta)^2 + \sigma_V^2$ and the GMM estimator is equivalent to OLS.
- More generally,

$$E(y_i \mid x_i, \theta) = \int g(x, \sigma_V u \mid \beta)\phi(u)du$$
$$E(y_i^2 \mid x_i, \theta) = \int g(x, \sigma_V u \mid \beta)^2\phi(u)du$$

and these may not have a closed form solution.

MSM for the nonseparable model

- First, simulate $u_{i1}, \ldots, u_{iS} \sim_{i.i.d.} \psi(\cdot)$ for each i
- Let

$$\hat{g}_1(x_i, \theta) = S^{-1} \sum_{s=1}^{S} g(x_i, \sigma_V u_{is} \mid \beta)$$

$$\hat{g}_2(x_i, \theta) = S^{-1} \sum_{s=1}^{S} g(x_i, \sigma_V u_{is} \mid \beta)^2,$$

  $\hat{m}_1(y_i, x_i, \theta) = y_i - \hat{g}_1(x_i, \theta)$, $\hat{m}_2(y_i, x_i, \theta) = x_i(y_i - \hat{g}_1(x_i, \theta))$, and $\hat{m}_3(y_i, x_i, \theta) = y_i - \hat{g}_2(x_i, \theta)$.
- Then define

$$Q_N(\theta) = \left(n^{-1} \sum_{i=1}^{n} \hat{m}(y_i, x_i, \theta) \right)' \left(n^{-1} \sum_{i=1}^{n} \hat{m}(y_i, x_i, \theta) \right)$$

  where $\hat{m} = (\hat{m}_1, \hat{m}_2, \hat{m}_3)'$.

MSM for the nonseparable model

- CT call $\hat{m}(y_i, x_i, \theta)$ the frequency simulator. This estimator:
  - is unbiased, e.g.,

$$
\begin{aligned}
E\left(\hat{m}_1(y_i, x_i, \theta)\right) &= E(y_i) - E\left(S^{-1}\sum_{s=1}^{S} E\left(g(x_i, \sigma_V u_{is} \mid \beta) \mid x_i, \theta\right)\right) \\
&= E(y_i) - E\left(\int g(x_i, \sigma_V u \mid \beta)\phi(u)du\right) \\
&= E(y_i - E(y_i \mid x_i, \theta))
\end{aligned}
$$

  - has variance $Var(\hat{m}) = \left(1 + \frac{1}{S}\right) Var(m)$

MSM for the nonseparable model

- The estimated asymptotic variance will just be $1 + S^{-1}$ times the estimated asymptotic variance of the conventional GMM estimator.
    - This entails estimating $\frac{\partial}{\partial\theta} m(y_i, x_i, \theta)$, which requires another simulation.
- The bootstrap is commonly used instead.

demonstration of MSM for the ns model

- A simple Matlab code snippet:

iams ▸ allfiles ▸ everything ▸ teaching ▸ current semester ▸ econ8379_spring2018 ▸ february 21

Editor – /Users/bwilliams/allfiles/everything/teaching/current semester/econ8379_spring2018/fe

hazards.m | RLsimlik.m | RLMSL.m | OLSsimmom.m | OLSMSM.m | RLsimV.m

```
function OLSsimmom=OLSsimmom(Y,X,U,theta)

n=length(Y);
S=size(U,2);
bet0=theta(1);
bet1=theta(2);
sig=theta(3);

mom(1)=sum(mean(Y*ones(1,S)-(bet0*ones(n,S)+bet1*X*ones(1,S)+sig*U),2));
mom(2)=X'*mean(Y*ones(1,S)-(bet0*ones(n,S)+bet1*X*ones(1,S)+sig*U),2);
mom(3)=sum(mean(Y.^2*ones(1,S)-(bet0*ones(n,S)+bet1*X*ones(1,S)+sig*U).^2,2));
mom;
OLSsimmom=(mom/n)*(mom/n)';

```

Indirect inference for the ns model

- an alternative to MSM is to define the objective function $Q_n(\gamma)$ by following these steps
  1. simulate the data M times from the nonseparable model with parameter $\gamma = (\beta, \sigma_V)$:

$$y_i^m = g(x_i, \sigma_V u_{im} \mid \beta)$$

  2. for each simulated dataset, compute the OLS parameter estimates, $\tilde{\theta}^m(\gamma)$
  3. average these estimates across the M simulations, $\tilde{\theta}(\gamma) = M^{-1} \sum_{m=1}^{M} \tilde{\theta}^m(\gamma)$
  4. measure the distance between these average estimates and the estimate from the real data, call this $Q_n(\gamma) = d(\hat{\theta}, \tilde{\theta}(\gamma))$
- Then let $\hat{\gamma} = \arg\min_\gamma Q_n(\gamma)$

demonstration of Indirect Inference for the ns model

- A simple Matlab code snippet:

```
function OLSindinf=OLSindinf(Y,X,U,theta)

n=length(Y);
S=size(U,2);
bet0=theta(1);
bet1=theta(2);
sig=theta(3);
W=[ones(n,1) X];
% generate data
Yt=bet0*ones(n,S)+bet1*X*ones(1,S)+sig*U;

% estimate OLS
for s=1:S,
    btild(:,s)=inv(W'*W)*W'*Yt(:,s);
    stild=sqrt((Yt(:,s)-W*btild(:,s))'*(Yt(:,s)-W*btild(:,s))/n);
end
bhat=inv(W'*W)*W'*Y;
shat=sqrt((Y-W*bhat)'*(Y-W*bhat)/n);

that=[bhat' shat]';
ttild=[btild' stild]';
OLSindinf=(that-ttild)'*(that-ttild);
```

Further notes on indirect inference

- The asymptotic variance can be derived using standard results for m-estimators. See formula in Gourieroux, Monfort, Renault (1993).
  - Requires asymptotic covariance matrix for $\hat{\theta}$.
  - Also requires an estimate of $\frac{\partial}{\partial\gamma}\theta(\gamma)$.
- Gourieroux, Monfort, Renault (1993) also discuss an optimal weighting matrix when $d(\hat{\theta}, \tilde{\theta}(\gamma)) = (\hat{\theta} - \tilde{\theta}(\gamma))' W(\hat{\theta} - \tilde{\theta}(\gamma))$
- Use of bootstrap is very common.

Further notes on indirect inference

- Smith (2008) says to instead define $\tilde{\theta}(\gamma)$ as the maximizer of the average of the *m* likelihoods.
  - Gourieroux, Monfort, Renault (1993) show that this is asymptotically equivalent in this model (and in fact most models).
- Many people refer to this as a method of simulated moments or simulated method of moments.
- Generally, $\hat{\theta}$, the auxiliary model, does not have to be a huge system of equations but instead can be different combinations of separate estimators.

Simulation-based estimation methods

MSL example

MSM example

How do we generate random numbers anyway?

Importance sampling

Conclusion

Inverse probability integral transform

- Suppose $u_1, \ldots, u_n$ are independent draws from a *Uniform*$(0, 1)$ distribution.
- Let F_X denote the cdf of a particular distribution.
- Then let $X_i = F_X^{-1}(u_i)$.

$$
\begin{aligned}
Pr(X_i \leq x) &= Pr(u_i \leq F_X(x)) \\
&= F_X(x)
\end{aligned}
$$

Inverse probability integral transform

- Suppose $u_1, \ldots, u_n$ are independent draws from a $Uniform(0, 1)$ distribution.
- Let F_X denote the cdf of a particular distribution.
- Then let $X_i = F_X^{-1}(u_i)$.

$$
\begin{aligned}
Pr(X_i \leq x) &= Pr(u_i \leq F_X(x)) \\
&= F_X(x)
\end{aligned}
$$

- $\Longrightarrow X_1, \ldots, X_n$ are independent draws from the distribution with cdf F_X.

Inverse probability integral transform

- Suppose $u_1, \ldots, u_n$ are independent draws from a *Uniform*$(0, 1)$ distribution.
- Let F_X denote the cdf of a particular distribution.
- Then let $X_i = F_X^{-1}(u_i)$.

$$\begin{aligned} Pr(X_i \leq x) &= Pr(u_i \leq F_X(x)) \\ &= F_X(x) \end{aligned}$$

- $\implies X_1, \ldots, X_n$ are independent draws from the distribution with cdf F_X.
- **One drawback**: What if we don't know F_X?

Pseudo-random number generators

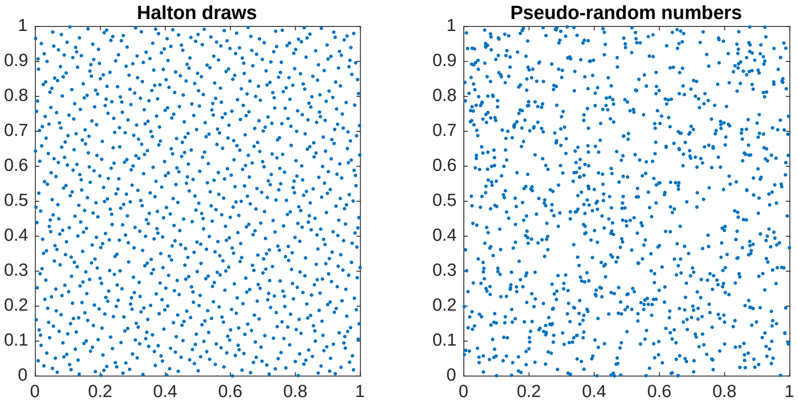
- How do we obtain $u_1, \ldots, u_n$ though?
  - A pseudo-random number generator is a deterministic sequence that mimics properties of a sequence of random variables.
  - Requires a *seed* to start
    - This is useful for replicating results because starting with the same seed produces the same sequence of draws.
  - *Period*: After a certain (large) number of draws, the sequence repeats itself.

Quasi-random number generators

- Pseudo-random numbers tend to not fill sample space uniformly.
  - Just like random numbers!
- This can lead to slow ($O(S^{-1/2})$) convergence of Monte Carlo integration.
- Quasi-random numbers are designed to provide better coverage (low discrepancy).

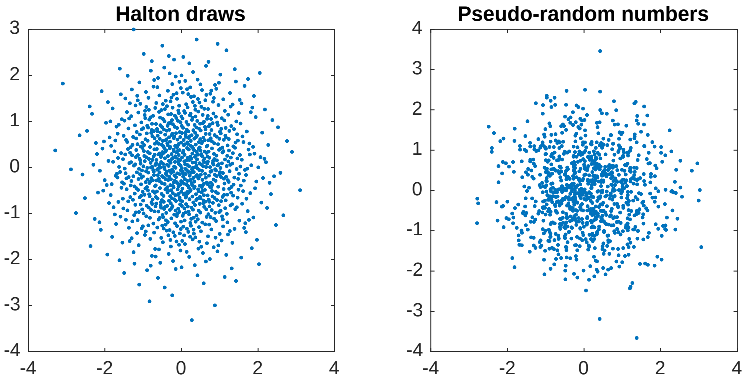
Quasi-random number generators

independent *Uniform*$(0, 1)$



Quasi-random number generators

independent $N(0,1)$



Quasi-random number generators

- Faster convergence but
  - this advantage is lost in high dimensions
  - at the cost (?) of not being independent draws
  - See Train (2000) for results on using Halton sequences in MSL estimation of the mixed logit model.

The basic idea

- Notice that

$$\int h(x)f(x)dx = \int \frac{h(x)f(x)}{g(x)}g(x)dx$$

- So we can either sample from f and compute $S^{-1}\sum_{s=1}^{S} f(x_s)$

The basic idea

- Notice that

$$\int h(x)f(x)dx = \int \frac{h(x)f(x)}{g(x)}g(x)dx$$

- So we can either sample from f and compute $S^{-1}\sum_{s=1}^{S} f(x_s)$
- or sample from g and compute $S^{-1}\sum_{s=1}^{S} \frac{h(x_s)f(x_s)}{g(x_s)}$

The basic idea

- Two reasons to do this:
  - The two simulators have the same mean but the variance of the latter is lower if $\int h(x)^2 f(x)dx > \int \frac{h(x)^2 f(x)}{g(x)} f(x)dx$.
  - If the function h is not smooth on the support of f but it is smooth on the support of g.

Example

Multinomial probit model.

- We need to simulate integrals like

$$\int \mathbf{1}(z_1 > c_1, z_2 > c_2)\phi(z_1, z_2; 0, \Sigma)dz$$

- This can be done by sampling Z_1^* from a truncated normal, $TN(0, 1; c_1/\sigma_{11}, \infty)$, and then sampling Z_2^* from $TN(0, 1; (c_2 - \sigma_{12}Z_1^*)/\sigma_{22}, \infty)$ and then computing

$$S^{-1}\sum_{s=1}^{S}(1 - \Phi(c_1/\sigma_{11}))\,(1 - \Phi((c_2 - \sigma_{12}Z_{1s}^*)/\sigma_{22}))$$

- Note that here h is $\mathbf{1}(z_1 > c_1, z_2 > c_2)$, f is the normal density and g is the truncated normal density.

Parting remarks

- The next time you take a look at a paper using these methods:
  - What is the advantage of the structural method over a “reduced form” method?
  - Do they discuss identification?
  - Do they discuss the simulator they use?
  - How do they compute standard errors?