Latin Hyper-Rectangle Sampling for Computer Experiments

David Mease
Department of Marketing and Decision Sciences
San Jose State University

Derek Bingham
Department of Statistics and Actuarial Science
Simon Fraser University

(see www.davemease.com for a copy of the paper)

Statement of the Problem

Given: $g(\cdot)$ (Computer Code)

$$X = (X_1, ..., X_d) \sim F$$

Want: $\mu = \mathsf{E}[g(X)] = \int_{R^d} g(x) f(x) dx = \mathsf{E}[g(F^{-1}(U))] = \int_{[0,1]^d} g^*(u) du$

Example:

$$g(x) = \frac{2\pi x_3 (x_4 - x_6)}{\log(\frac{x_2}{x_1})[1 + \frac{2 x_7 x_3}{\log(x_2/x_1) x_1^2 x_8} + \frac{x_3}{x_5}]}$$

$X_1 =$ radius of borehole, .05 to .15m
$X_2 =$ radius of influence, 100 to 50,000m
$X_3 =$ transmissivity of upper aquifer, 63,070 to 11,5600m^2/yr
$X_4 =$ potentiometric head of upper aquifer, 990 to 1,110m
$X_5 =$ transmissivity of lower aquifer, 63.1 to 116m^2/yr
$X_6 =$ potentiometric head of lower aquifer, 700 to 820m
$X_7 =$ length of borehole, 1,120 to 1,680m
$X_8 =$ hydraulic conductivity of borehole, 9,855 to 12,045m/yr

There are several important general classes of problems that can be approached through computer experiments. Some of the major ones are:

(1) Prediction: Given t, predict y .

(2) Sensitivity analysis: Identify the important and the negligible input variables.

(3) Uncertainty analysis: Determine how uncertainty about t affects y Equivalently, determine the variability in y caused by random variability in t.

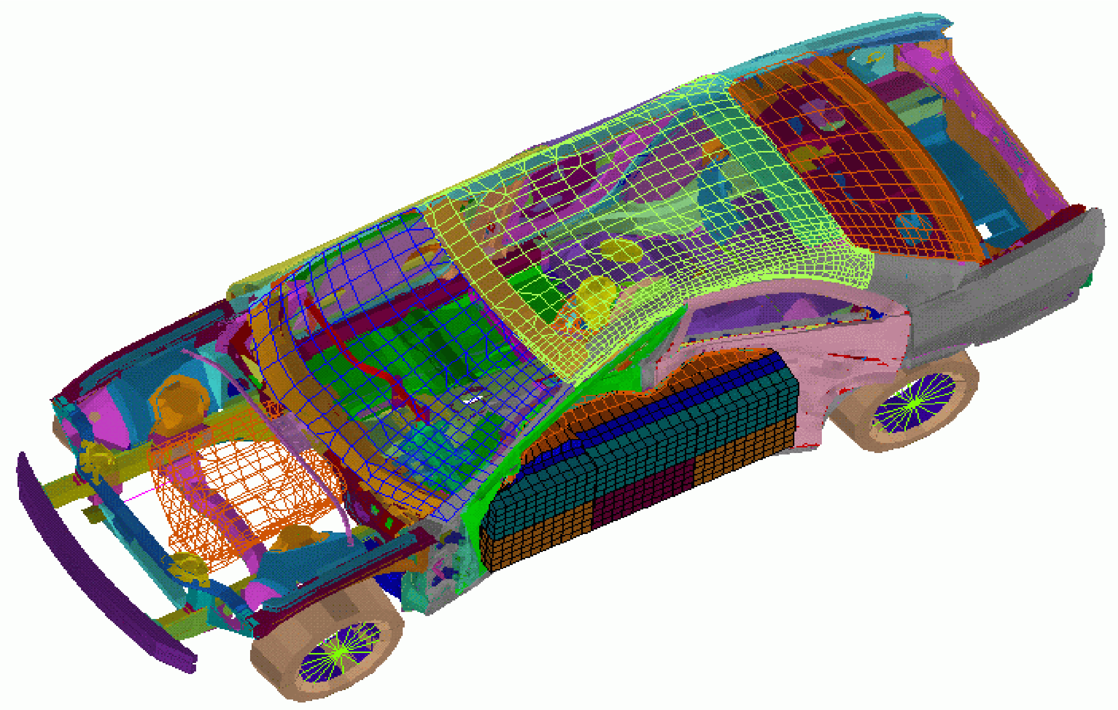
(4) Optimization: Find the t at which y is "best" in some sense.

(5) Root finding: Find a t that yields a specified y

(6) Integration of output: Find the average y that results when t is randomly drawn from a known input distribution.

Examples of Applications

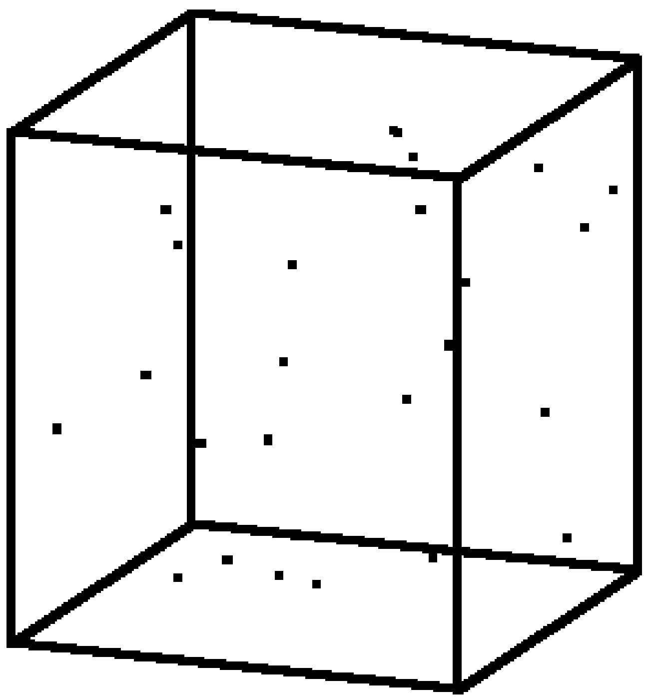
1) Automobile Industry
2) Fluid Flow
3) Integrated Circuits
4) Thermal Energy Storage
5) Combustion
6) Aeronautics
7) Structural Reliability
8) Semiconductors
9) Nuclear Fusion



Review of Monte Carlo Methods

1) Regular Monte Carlo (Simple Random Sampling)

Sample $X[1], ..., X[n]$ *iid* from f



$$\widehat{\mu} = \frac{1}{n} \sum_{i=1}^{n} g(X[i])$$

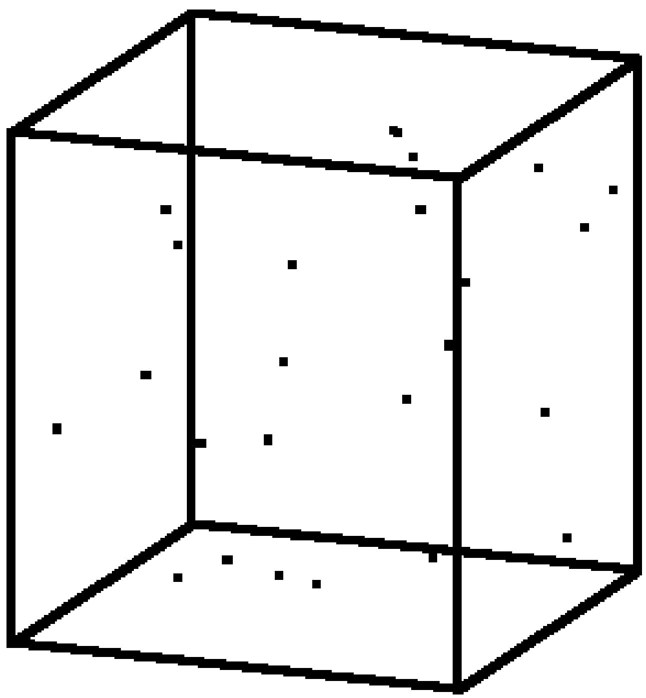
$$n = 27$$

$$d = 3$$

Review of Monte Carlo Methods

2) Importance Sampling

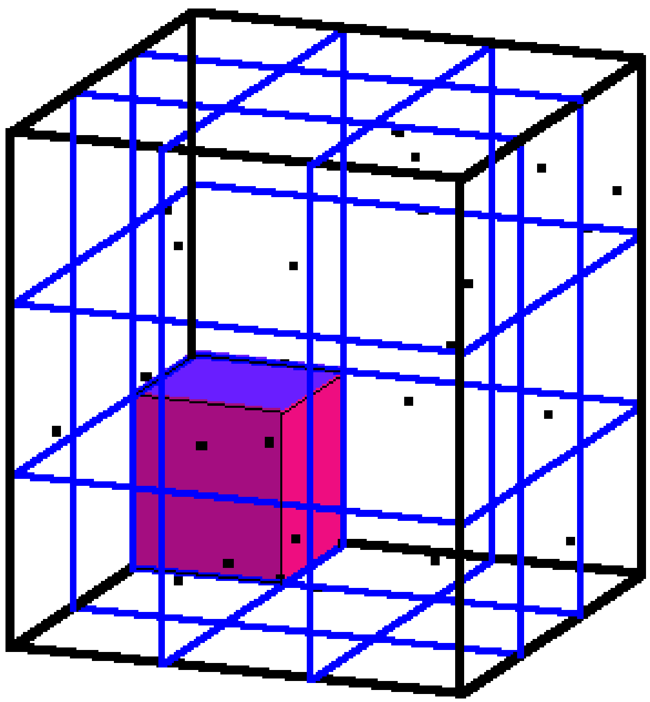
Sample $X[1], \ldots, X[n]$ *iid* from h



$$\hat{\mu} = \frac{1}{n} \sum_{i=1}^{n} g(X[i]) \frac{f(X[i])}{h(X[i])}$$

Review of Monte Carlo Methods

3) Stratified Random Sampling



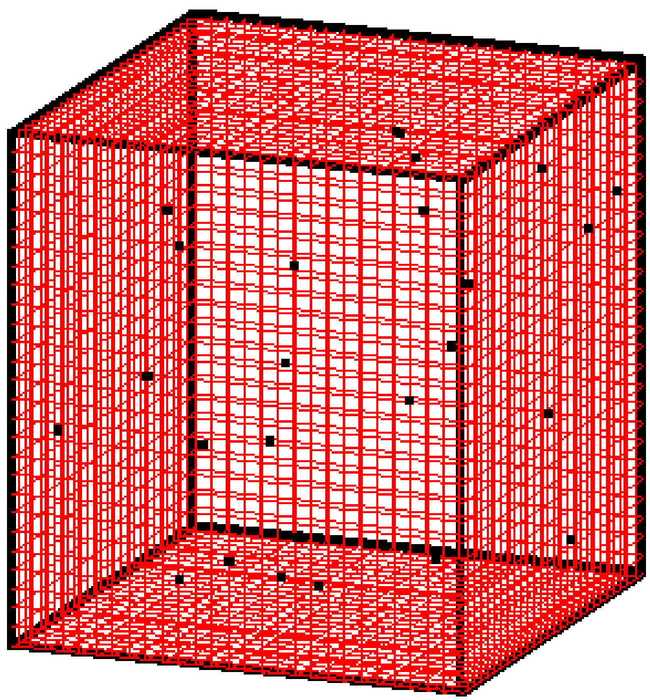
$$\hat{\mu} = \frac{1}{n} \sum_{i=1}^{n} g(X[i])$$

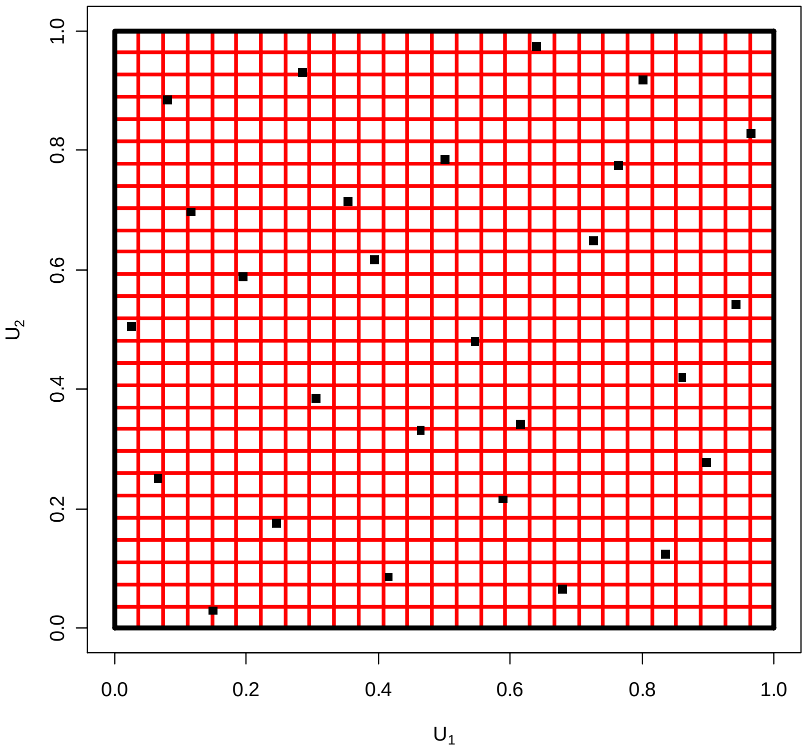
$$n = 27 = 3^3$$

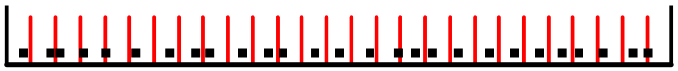
Review of Monte Carlo Methods

4) Latin Hypercube Sampling (LHS)

$$\hat{\mu} = \frac{1}{n} \sum_{i=1}^{n} g(X[i])$$





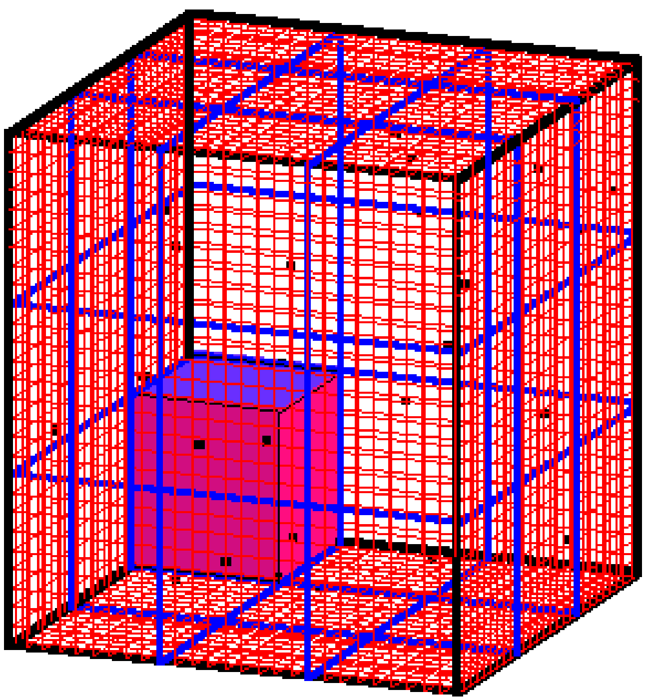


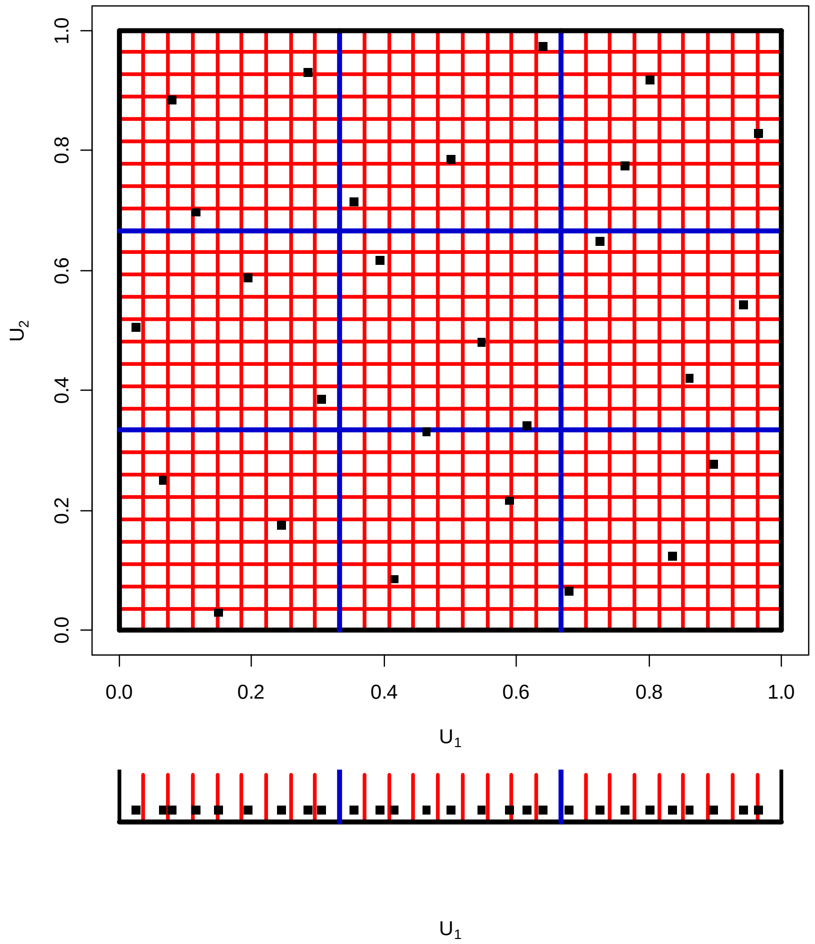
Review of Monte Carlo Methods

5) Stratified Latin Hypercube Sampling

$$\hat{\mu} = \frac{1}{n} \sum_{i=1}^{n} g(X[i])$$

$$n = 27 = 3^3$$



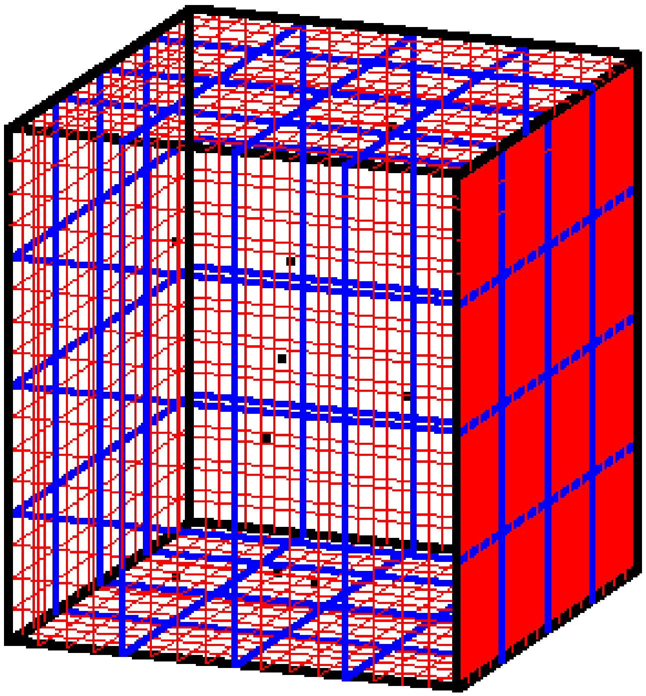


Review of Monte Carlo Methods

6) OA-Based Latin Hypercube Sampling

$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} g(X[i])$$

$$n = 16 = 4^2$$



$$OA(16,3,4,2)$$

```
D####D####D
D####E####E
D####F####F
D####G####G
E####D####E
E####E####D
E####F####G
E####G####F
F####D####F
F####E####G
F####F####D
F####G####E
G####D####G
G####E####F
G####F####E
G####G####D
```

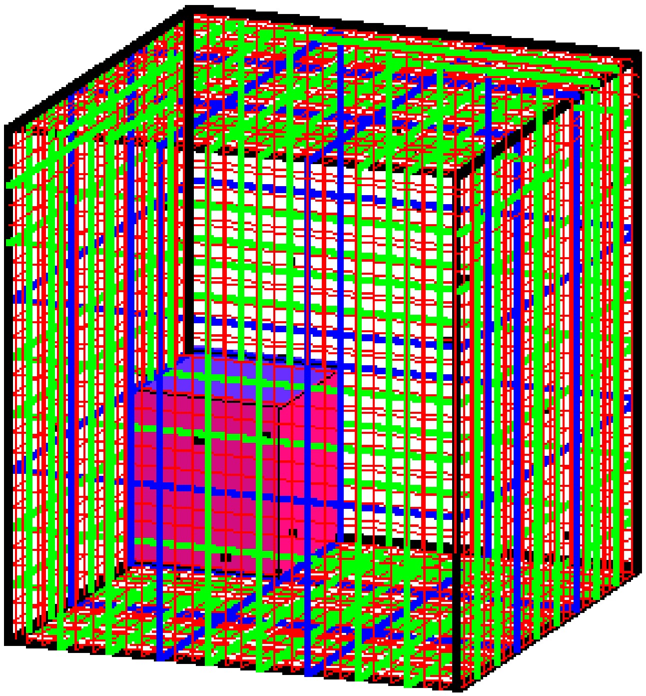
Review of Monte Carlo Methods

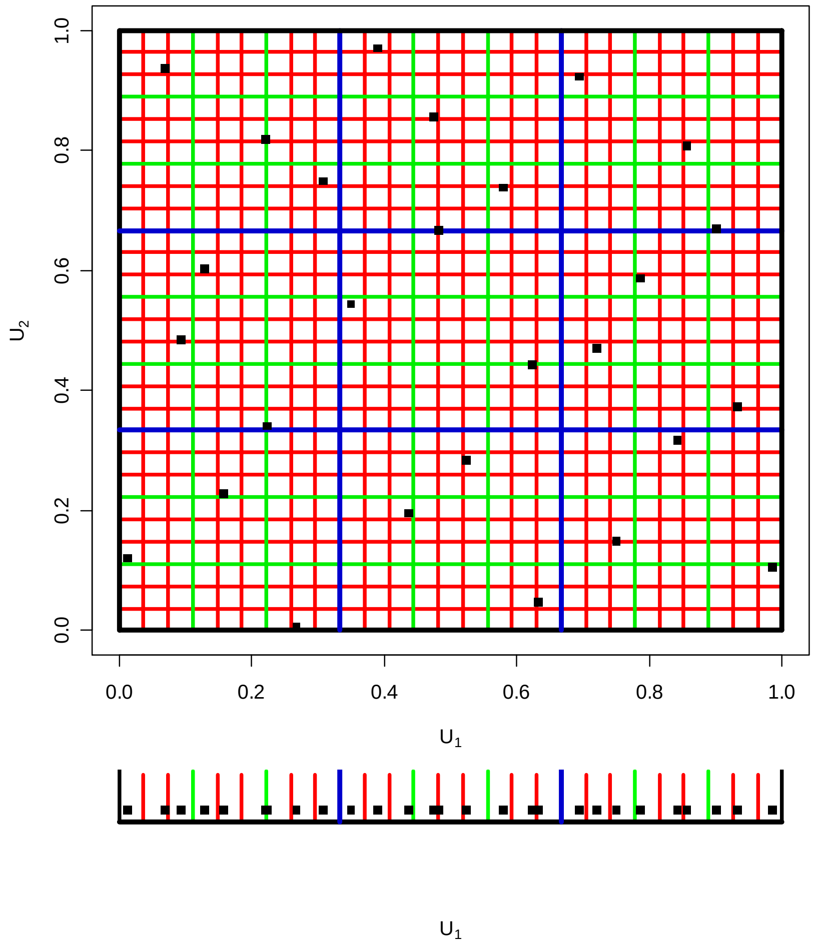
7) Scrambled Nets

$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} g(X[i])$$

$$n = 27$$

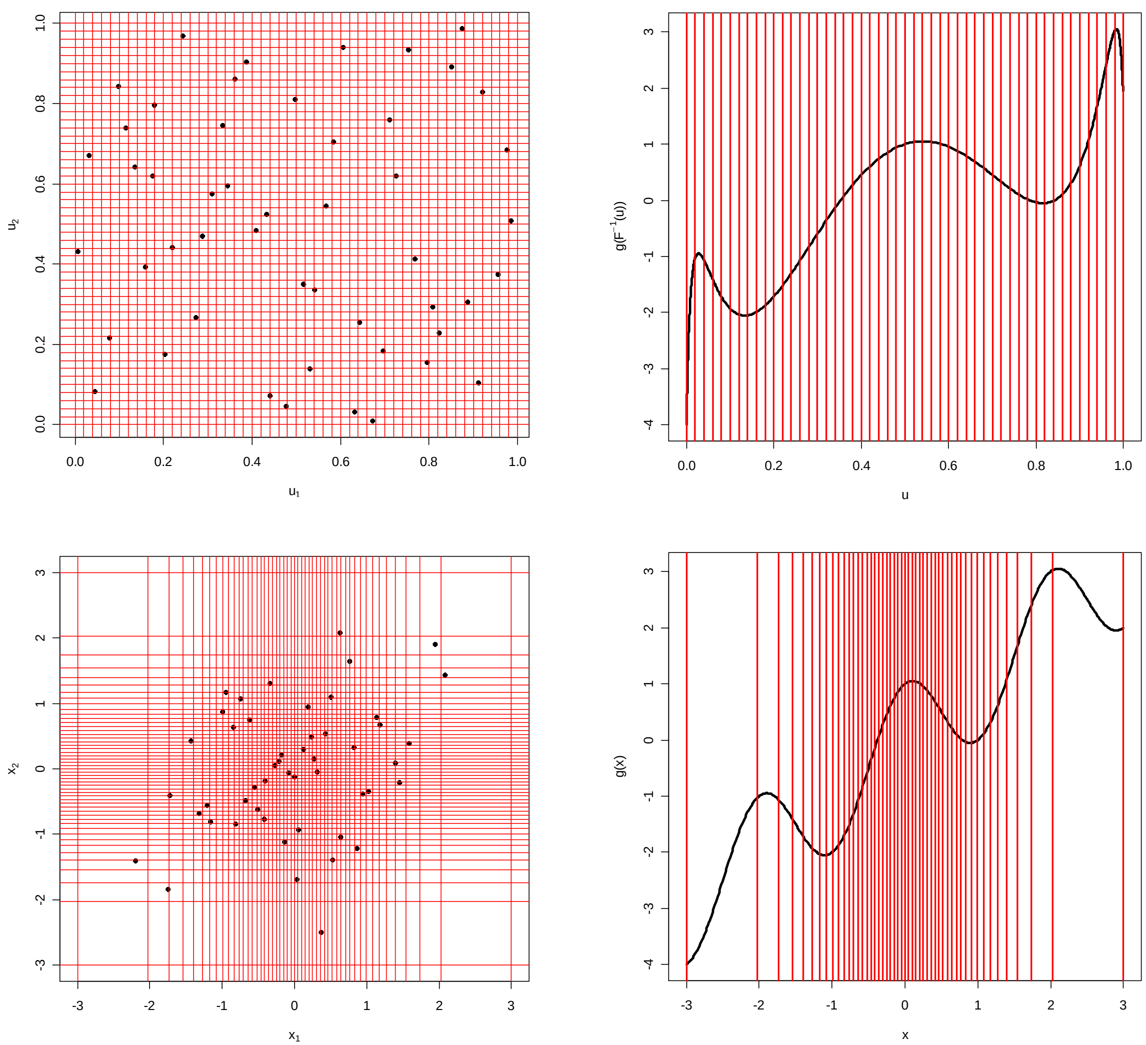
$$(t, m, d) = (0, 3, 3) \text{ net in base } 3$$



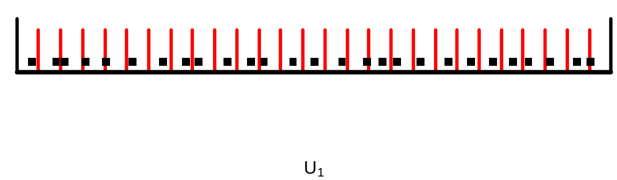


Why Hyper-Rectangles?

$$\mu = \mathsf{E}[g(X)] = \int_{R^d} g(x) f(x) dx = \mathsf{E}[g(F^{-1}(U))] = \int_{[0,1]^d} g^*(u) du$$

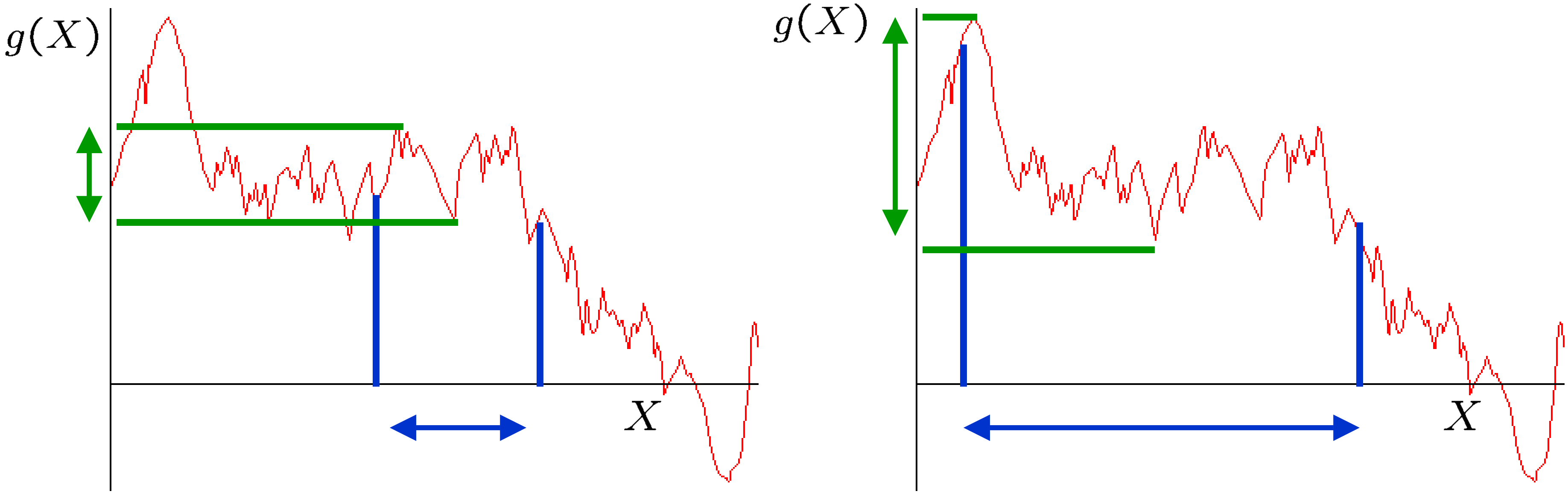


Is LHS "Optimal" for 1 Dimensional Projections?



$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} g(X[i]) \longrightarrow \hat{\mu} = \sum_{i=1}^{n} g(X[i])p_i$$

$$\mathsf{Var}[\hat{\mu}] = \sum_{i=1}^{n} \mathsf{Var}[g(X[i])]p_i^2$$



$$\mathsf{E}_g(\mathsf{Var}_{X[i]}(g(X[i])) =$$

$$\frac{1}{2}\mathsf{E}_g[E_{X[i],Y[i]}(g(X[i]) - g(Y[i]))^2] =$$

$$\frac{1}{2}\mathsf{E}_{X[i],Y[i]}[E_g(g(X[i]) - g(Y[i]))^2] =$$

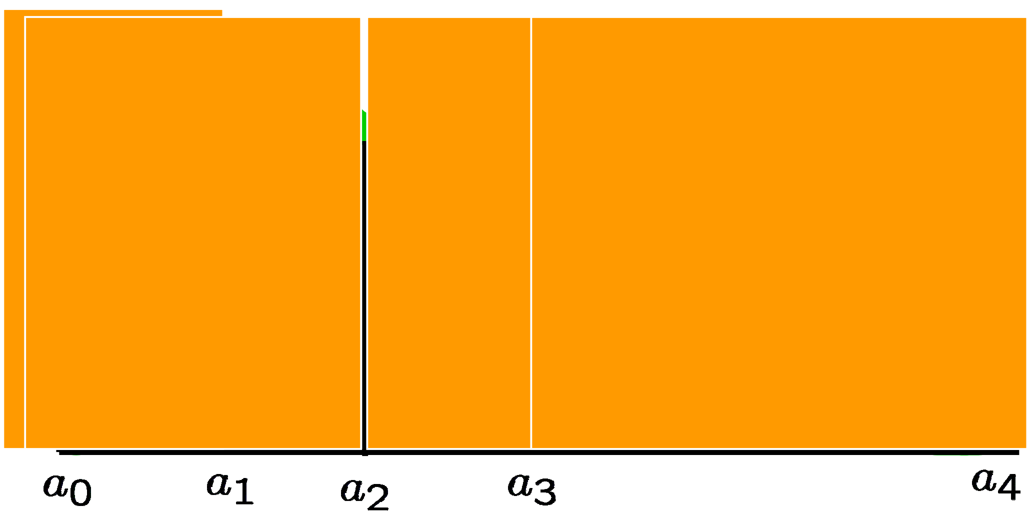
$$\frac{1}{2}\mathsf{E}_{X[i],Y[i]}|X[i] - Y[i]| \equiv \Delta_i$$

Minimize $\text{Var}[\hat{\mu}] = \sum_{i=1}^{n} \text{Var}[g(X[i])] p_i^2$

$$\Rightarrow \sum_{i=1}^{n} \Delta_i p_i^2 = \sum_{i=1}^{n} \int_{a_{i-1}}^{a_i} \int_{a_{i-1}}^{a_i} |x-y| f(x) f(y) dx dy$$

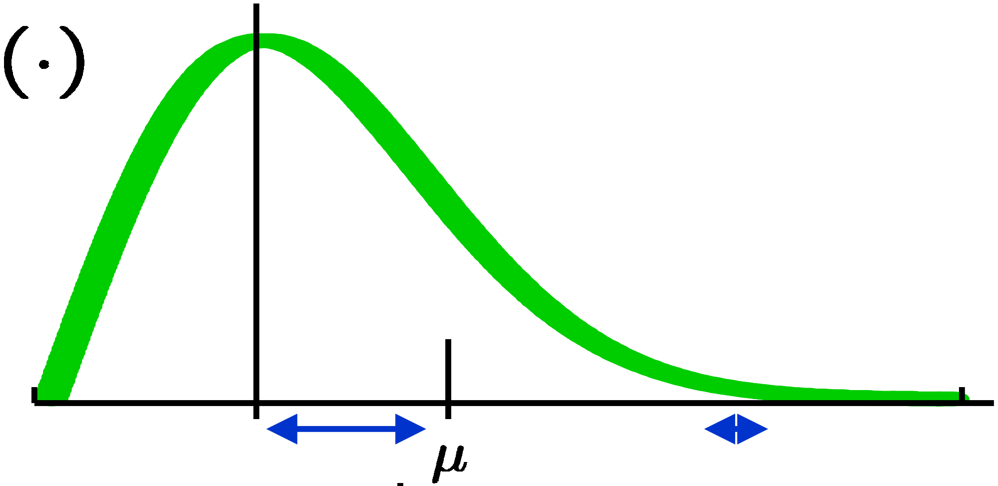
Set derivative=0

$$\Rightarrow a_i = \frac{\int_{a_{i-1}}^{a_{i+1}} x f(x) dx}{\int_{a_{i-1}}^{a_{i+1}} f(x) dx}$$

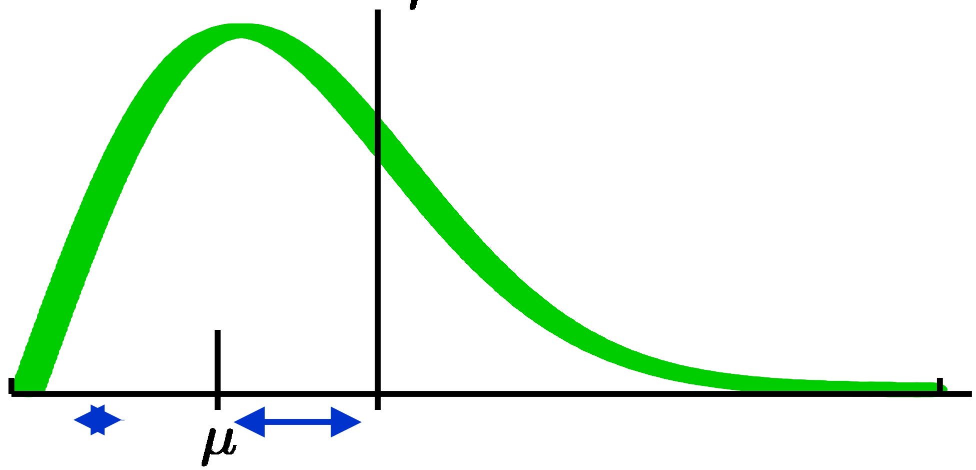


$$p_i = \int_{a_{i-1}}^{a_i} f(x) dx$$

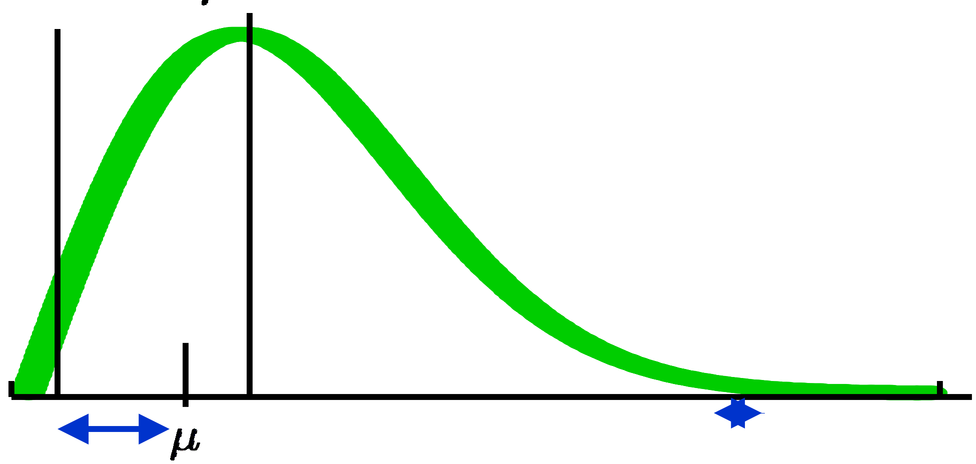
Log-concave density $f(\cdot)$



→ 1) X has IHR



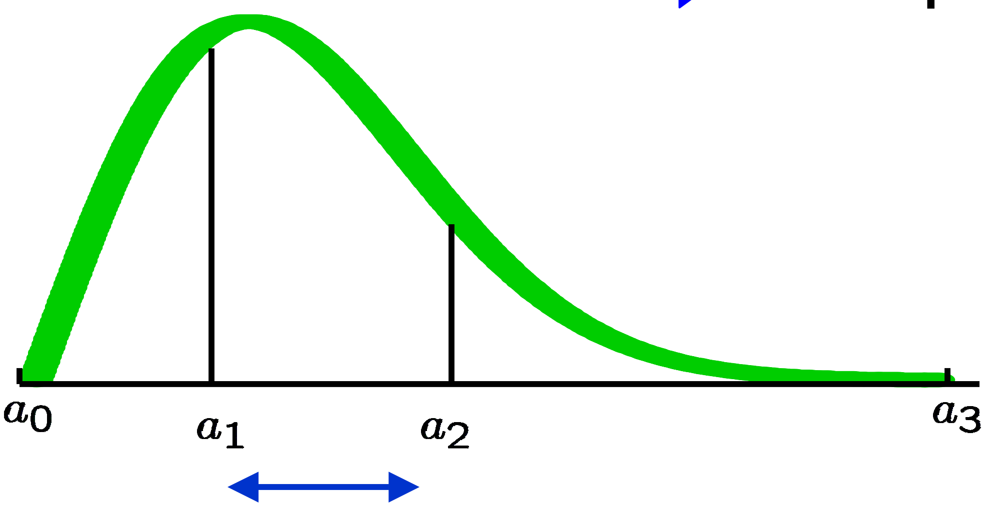
→ 2) $-X$ has IHR

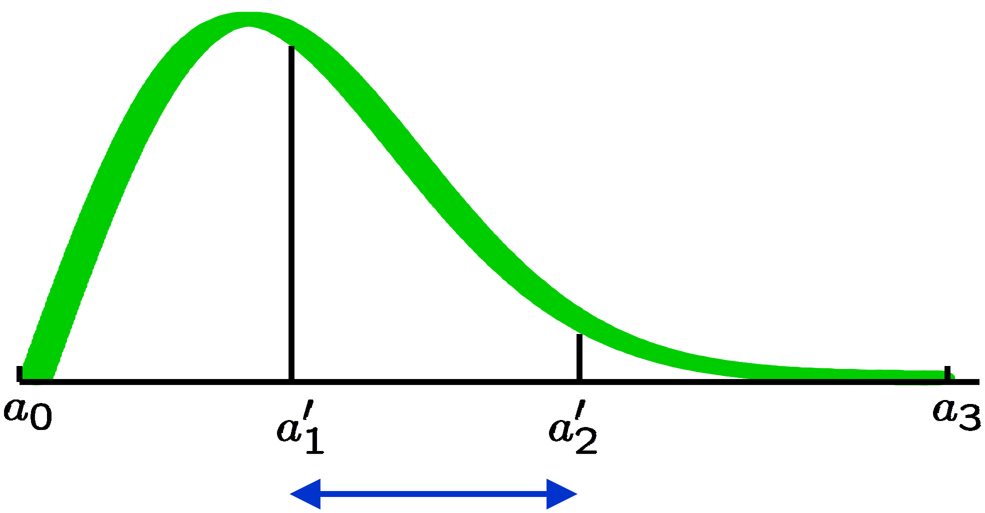


→ 3) X has ILR

Log-concave density $f(\cdot)$

Unique $a_i = \frac{\int_{a_{i-1}}^{a_{i+1}} x f(x) dx}{\int_{a_{i-1}}^{a_{i+1}} f(x) dx}$





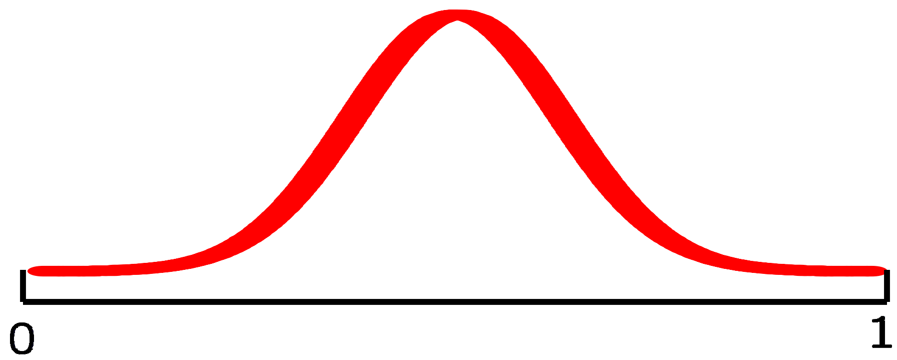
Suppose $a'_2 > a_2$

$\Rightarrow$ $a'_1 > a_1$

and $a'_2 - a'_1 > a_2 - a_1$

$\Rightarrow$ **Contradiction!**

Optimal p_i for Truncated Normal with $n = 20$



(LHS)		(EW)
0.050	0.026	0.002
0.050	0.037	0.005
0.050	0.044	0.010
0.050	0.049	0.018
0.050	0.053	0.031
0.050	0.055	0.048
0.050	0.057	0.069
0.050	0.059	0.090
0.050	0.060	0.108
0.050	0.060	0.118
0.050	0.060	0.118
0.050	0.060	0.108
0.050	0.059	0.090
0.050	0.057	0.069
0.050	0.055	0.048
0.050	0.053	0.031
0.050	0.049	0.018
0.050	0.044	0.010
0.050	0.037	0.005
0.050	0.026	0.002

Variance Comparison for Truncated Normal

Function	LHS Cells	Optimal Cells	Relative Efficiency
$g(x) = x$	1.76×10^{-5}	7.37×10^{-6}	2.39
$g(x) = x^2$	2.54×10^{-5}	9.59×10^{-6}	2.65
$g(x) = x^3$	3.84×10^{-5}	1.23×10^{-5}	3.13
$g(x) = \sin(x)$	1.26×10^{-5}	5.39×10^{-6}	2.33
$g(x) = \sin(3x)$	1.03×10^{-4}	3.02×10^{-5}	3.43