Coupling Designs for Experiments with Complex Treatments

Max Cytrynbaum

Yale University

Joint work with Fredrik Sävje

Introduction

Consider assigning $D \sim F$ in a randomized experiment.

To improve efficiency, can use stratified randomization.

Matched Pairs. Let $D \in \{0, 1\}$ and $F = \text{Bernoulli}(1/2)$. Match units into pairs $X_i \approx X_j$, assign $D_i = 1$ and $D_j = 0$ or vice-versa.

What if $D \sim F = \text{Unif}[0, u]$, e.g. effect of cash grant on $Y_i(d)$.

Could let $D_i \sim F$ iid, $F = \text{Unif}[0, u]$ and regress $Y \sim 1 + D$.

Estimates linear approx of dose-response $d \to n^{-1} \sum_i Y_i(d)$.

Can no longer stratify, treatment space $\mathcal{D} = [0, u]$ has $+\infty$ levels.

Stratification and Match Quality

Discretize, then stratify? Let $D \sim F_k = \text{Unif}\{1/k, 2/k, \dots\}$.

Matched k-tuples. If $k = 20$, match units into 20-tuples using X_i. Randomly assign one unit to each treatment level $d = l/k$.

Problems. Dose-response $d \to n^{-1} \sum_i Y_i(d)$ is not identified.

For $k = 2$, already hard to find $X_i \approx X_j$ for moderate $\dim(X)$.

Match quality will be much worse for $k = 20$.

For $k = 100$ levels, match into 100-tuples...

Generalizing Matched Pairs

Antithetic Variates

Can we construct a matched pairs design for any $D \sim F$?

Matched pairs. Match $X_i \approx X_j$, let $D_i, D_j \sim \text{Bernoulli}(1/2)$ marginally with $\text{Corr}(D_i, D_j) = -1$.

Idea. Match $X_i \approx X_j$, draw $D_i, D_j \sim F$ with $\text{Corr}(D_i, D_j) \ll 0$.

Antithetic Variates. Draw $U \sim \text{Unif}[0,1]$ and set $D_i^* = F^{-1}(U)$ and $D_j^* = F^{-1}(1-U)$.

Then $D_i^*, D_j^* \sim F$, but $\text{Corr}(D_i^*, D_j^*) \ll 0$.

Hammersley and Mauldon (1956): For any $Y(\cdot)$ monotone,

$$\text{Corr}(Y(D_i^*), Y(D_j^*)) = \min_{G_i = G_j = F} \text{Corr}_G(Y(D_i), Y(D_j)).$$

Antithetic Matched Pairs

$$\mathrm{Corr}(Y(D_i^*), Y(D_j^*)) = \min_{G_i = G_j = F} \mathrm{Corr}_G(Y(D_i), Y(D_j)).$$

Consider simple estimator $\widehat{\theta} = (1/2)(Y_i(D_i) + Y_j(D_j))$.

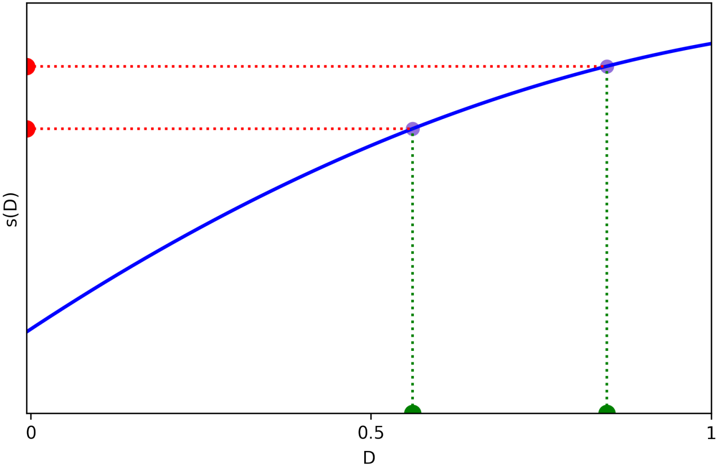
Suppose units well matched, $Y_i(\cdot) \approx Y_j(\cdot) = Y(\cdot)$. Then

$$\frac{\mathrm{Var}_G(\widehat{\theta})}{\mathrm{Var}_{iid}(\widehat{\theta})} - 1 = \mathrm{Corr}_G(Y(D_i), Y(D_j)).$$

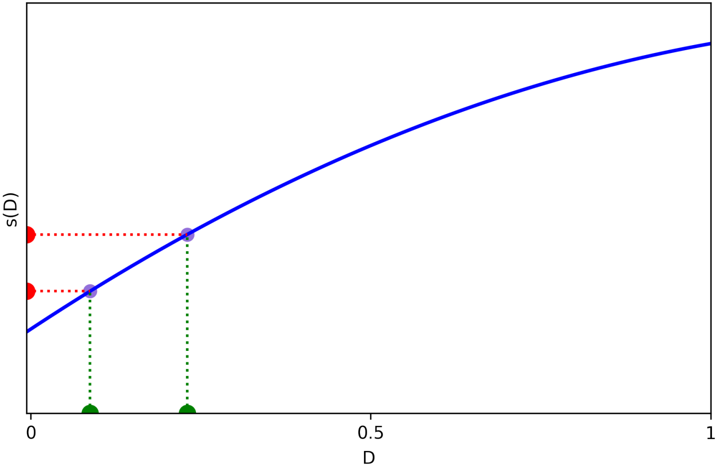
Result. We defined matched pairs for any univariate $D \sim F$.

Ingredients: high quality matches + negative correlation.

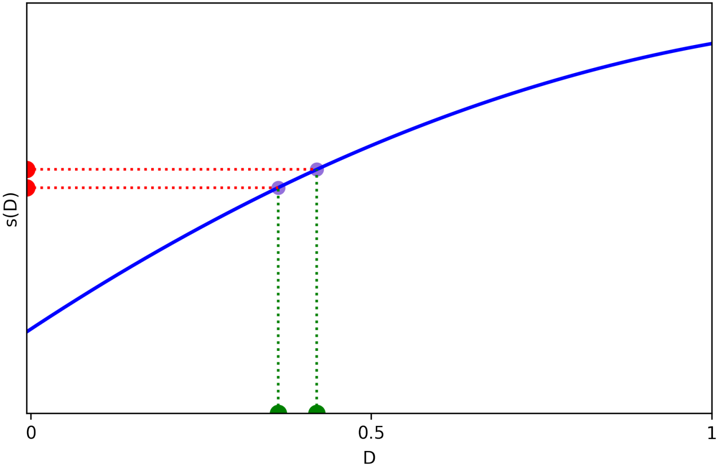
Independent Randomization



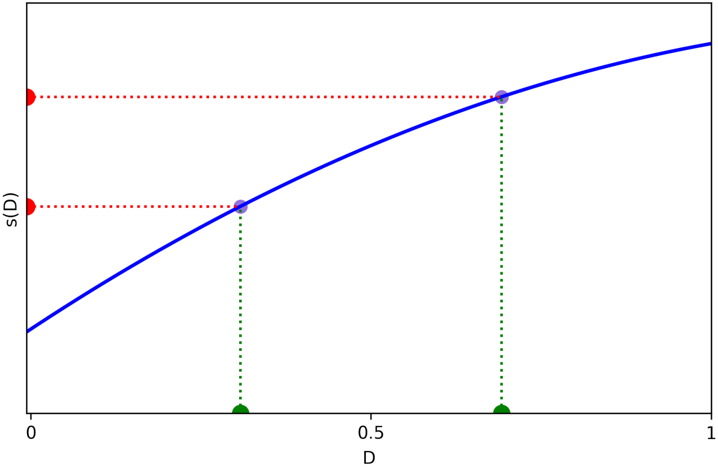
Independent Randomization



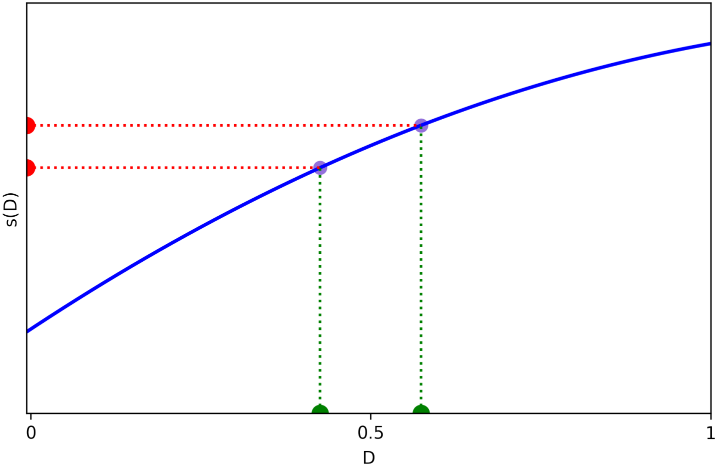
Independent Randomization



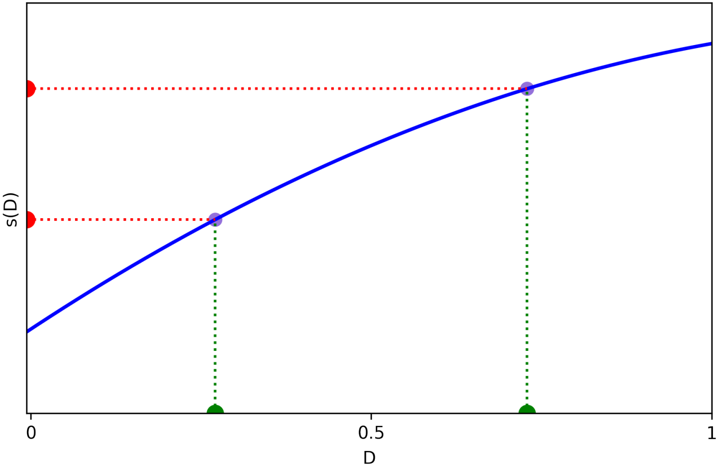
Antithetic Variates



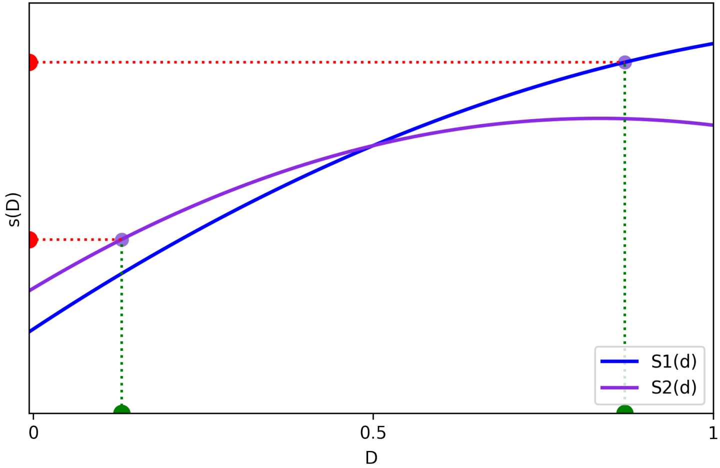
Antithetic Variates



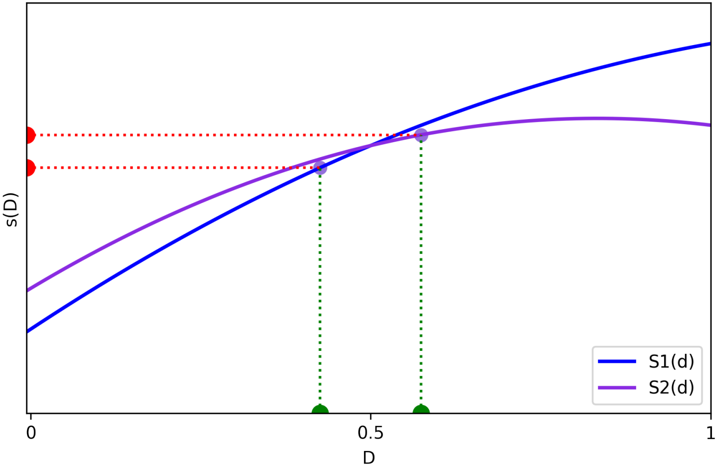
Antithetic Variates



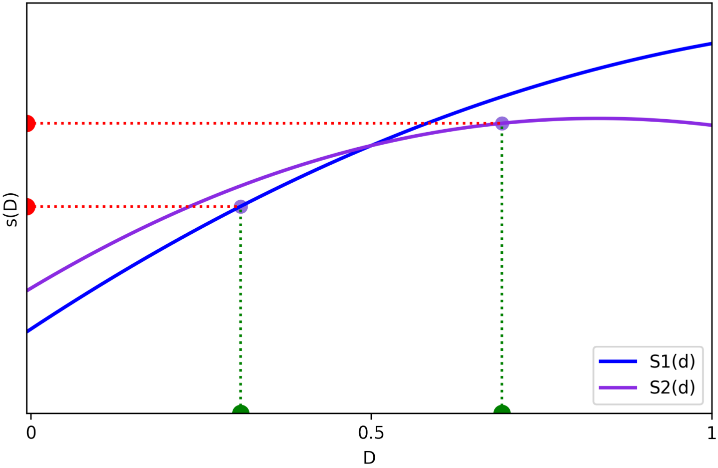
Antithetics + High Match Quality



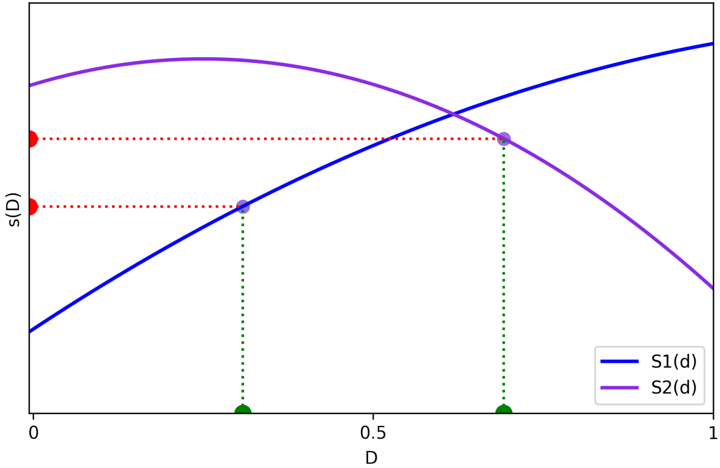
Antithetics + High Match Quality



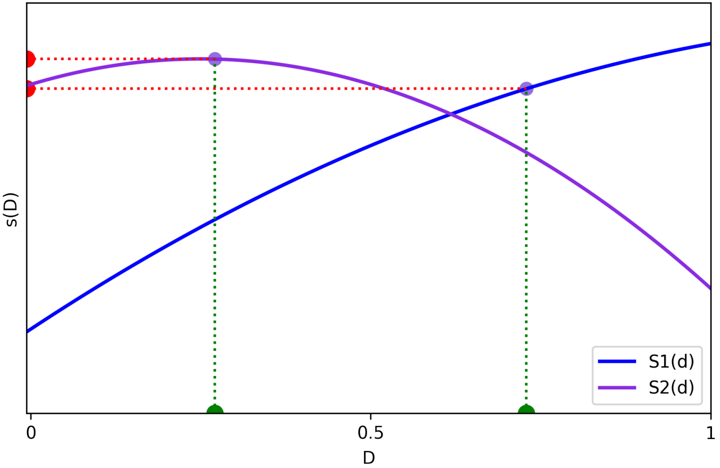
Antithetics + High Match Quality



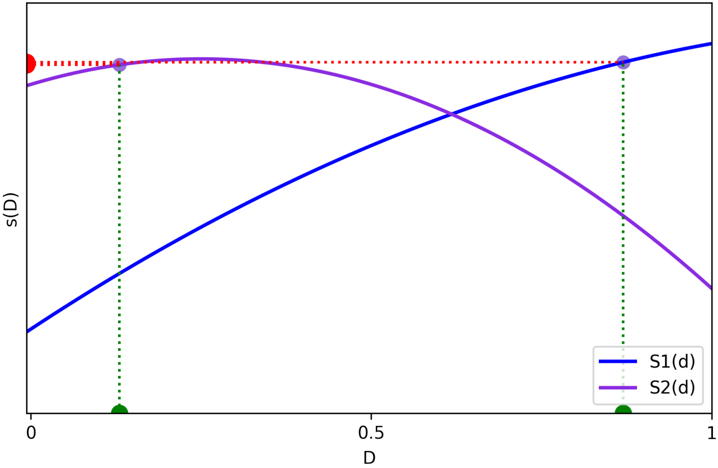
Antithetics + Low Match Quality



Antithetics + Low Match Quality



Antithetics + Low Match Quality



Coupling Designs

$$\frac{\text{Var}_G(\widehat{\theta})}{\text{Var}_{iid}(\widehat{\theta})} - 1 = \text{Corr}_G(Y_i(D_i), Y_j(D_j)), \qquad Y_i(\cdot) \approx Y_j(\cdot).$$

This doesn't require $d \in \mathcal{D}$ univariate, even finite-dim.

Suggests extending this to **complex experiments**, e.g. with irregular treatment space $\mathcal{D} \subseteq \mathbb{R}^m$.

E.g. $\mathcal{D} = [0, u] \times \{1, \ldots, t\}$ cash grant, training frequency.

Definition. (Coupling Designs)

1. Match units into groups $|g| = k$ using covariates X_i.
2. Draw treatments $(D_i)_{i=1}^k \sim G$ with $G_i = F$ such that

$$\text{Cov}_G(\phi(D_i), \phi(D_j)) \ll 0 \qquad \phi : \mathcal{D} \to \mathbb{R}.$$

Complex Treatments

Cash Grants. Let $\mathcal{D} = [0, u] \times \{1, \ldots, t\}$ and $F = \text{Unif}\,\mathcal{D}$.

Assign $D_i \sim F$, outcome $Y_i = Y_i(D_i)$. Goal is precision of OLS:

$$Y \sim 1 + D_1 + D_2 + D_1 D_2$$

For multivariate D, assign $(D_i)_{i=1}^k \sim G$ "spread out" over $\mathcal{D}$.

$$\text{Cov}_G(\phi(D_i), \phi(D_j)) \ll 0 \qquad \phi : \mathcal{D} \to \mathbb{R}.$$

Intuition. Assigning similar units to dissimilar treatments prevents spurious *in-sample* correlations between D_i and X_i.

Covariate Balance. Coupling designs extend this to complex $\mathcal{D}$.

Contributions

Introduce new family of **coupling designs**. Enable efficient randomization in experiments with complex treatments.

Idea. Draw $D_i \sim F$ using coupling $(D_i)_{i=1}^k \sim G$ s.t. highly dispersed over $\mathcal{D}$, within homogeneous matched k-tuples.

Method. Construct couplings G using classic Monte Carlo techniques (random lattices) + tools from OT (Brenier map).

Theory. In family of causal problems, we prove

$$\text{Efficiency Gain} = \text{Dispersion} \times \text{Match Quality}$$

Develop spectral theory of "principal directions" of coupling G, use to study efficiency and robustness.

Related Literature

Stratified Randomization. Bai et al. (2022), Wang et al. (2021), Cytrynbaum (2024a, 2024b), Koo and Pashley (2025).

Monte Carlo. Stein (1987), Owen (1994), Craiu and Meng (2005), L'Ecuyer and Lemieux (2010), Owen (2013).

Optimal Transport. Brenier (1991), Mérigot (2011).

Estimators and Estimands

Estimators and Estimands

We study $\widehat{\theta} = E_n[s_i(D_i)]$ with $s_i(\cdot)$ fixed. Finite-pop estimand

$$\theta_n = n^{-1} \sum_i E_F[s_i(D)].$$

Example. Horvitz-Thompson $\widehat{\theta} = E_n[Y_i(D_i)H(D_i)]$,

$$s_i(d) = Y_i(d)H(d), \quad H(d) = \frac{d - E_F[D]}{\mathrm{Var}_F(D)}.$$

Estimand is BLP of dose-response $\bar{Y}_n(d) = n^{-1} \sum_i Y_i(d)$.

$$\theta_n = \operatorname*{argmin}_{\theta} \min_{\alpha} E_F[(\bar{Y}_n(D) - \alpha - \theta' D)^2].$$

Example. For OLS regression $Y \sim 1 + D$, have

$$\widehat{\theta} - \theta_n = E_n[s_i(D_i)] + O_p(n^{-1}) \qquad s_i(d) = e_i(d)H(d)$$

Discrete Choice

Two-Sided Market. Restaurants on delivery platform with features $d_j = (\text{cuisine, price, rating})$. Decision $Y_i(d) \in \{0,1\}$.

Treatment space $\mathcal{D} = \{d_1, \ldots, d_N\} \subseteq \mathbb{R}^m$ non-manipulable.

Model $P(Y_i = 1 \mid D_i) = L(\beta' D_i)$, estimate Logit MLE $\widehat{\beta}$

$$\widehat{\beta} - \beta_n = E_n[s_i(D_i)] + O_p(n^{-1})$$

Influence function $s_i(d) = [Y_i(d) - L(\beta_n' d)] \cdot J_n^{-1} d$.

Designs exploit smoothness of $s_i(\cdot)$ for $d \in \mathcal{D}$.

Estimand. $\mathrm{KL}(p \| q)$ divergence of Bernoulli(p) and Bernoulli(q):

$$\beta_n = \underset{\beta}{\operatorname{argmin}} \; E_F\left[\mathrm{KL}\left(\bar{Y}_n(D) \parallel L(\beta' D)\right)\right].$$

Coupling Construction

Coupling Construction

Construct exchangeable coupling G s.t. $(D_i)_{i=1}^k \sim G$ highly dispersed over $\mathcal{D}$, with fixed marginal $G_i \sim F$.

Must accommodate complex treatment spaces like:

$$\mathcal{D} = [0, u] \times \{1, \ldots, t\}, \qquad \mathcal{D} = \{d_1, \ldots, d_N\} \subseteq \mathbb{R}^m.$$

Monte Carlo + OT.

1. Draw highly dispersed $(U_i)_{i=1}^k \sim G_U$ with $U_i \sim \mathrm{Unif}[0, 1]^m$.
2. Set $D_i = T(U_i)$ for a geometry-preserving **transport map**

$$T : [0, 1]^m \to \mathcal{D}, \qquad T(U) \sim F.$$

Univariate Couplings

For $k \geq 2$ want $(U_i)_{i=1}^k \sim G_U$ with $\mathrm{Corr}_G(U_i, U_j) \ll 0$.

Nonparametric. For $k \geq 2$, options include

- Latin hypercube sampling (LHS)
- Rotation sampling (RS)

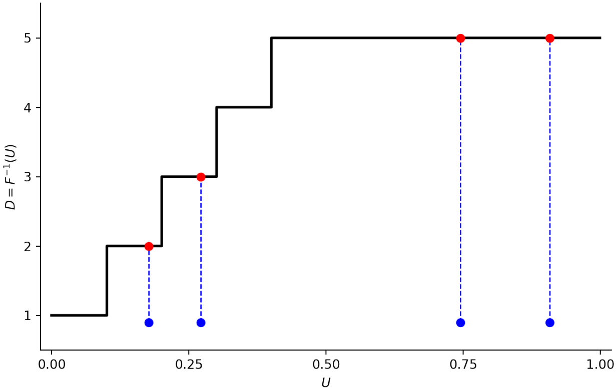
Parametric. (Gaussian Copula) Draw $Z \sim \mathcal{N}(0, \Sigma)$ with $\Sigma_{ii} = 1$

$$\Sigma_{ij} = -(k-1)^{-1}, \quad i \neq j.$$

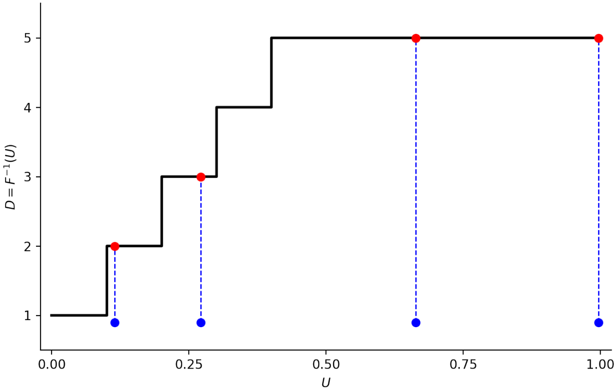
Transform $U_i = \Phi(Z_i)$, then $\mathrm{Corr}_G(U_i, U_j) < 0$.

Transport Map. If $D_i = F^{-1}(U_i)$ then $D_i \sim F$, so $T = F^{-1}$.

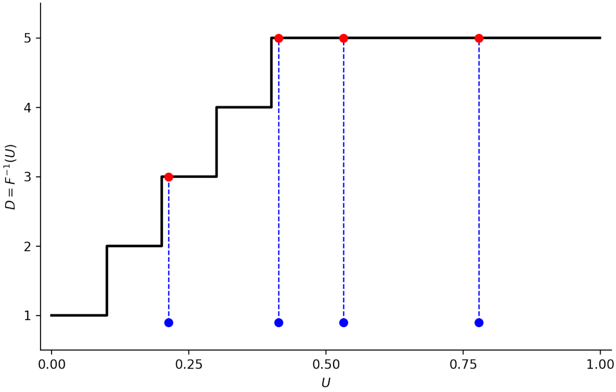
Latin Hypercube for 5-arm Distribution, $k = 4$



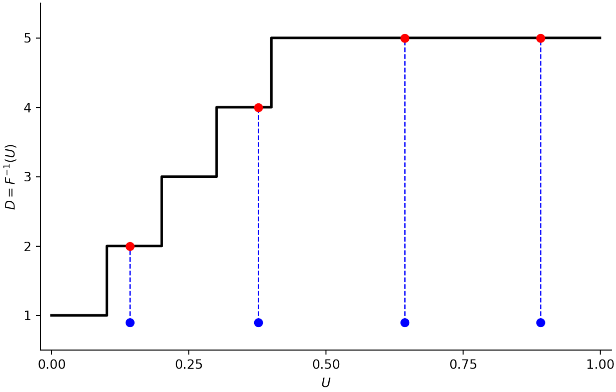
Latin Hypercube for 5-arm Distribution, $k = 4$



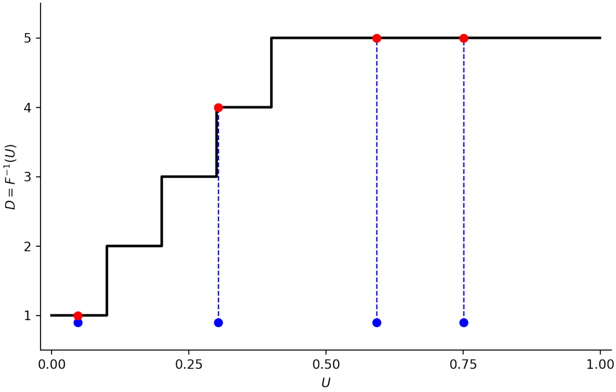
Latin Hypercube for 5-arm Distribution, $k = 4$



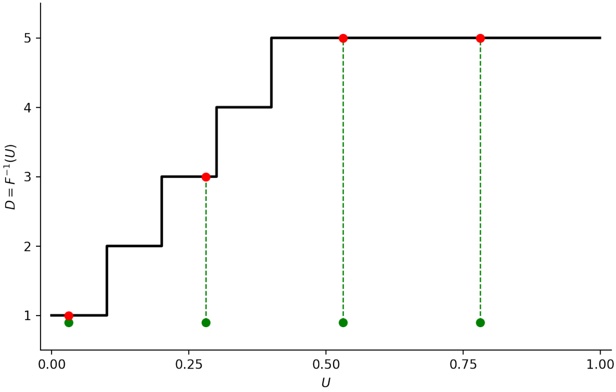
Latin Hypercube for 5-arm Distribution, $k = 4$



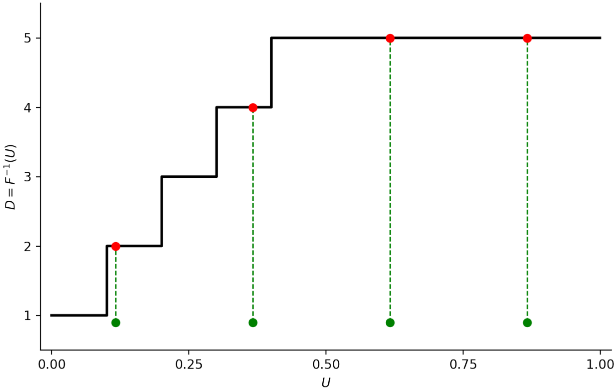
Latin Hypercube for 5-arm Distribution, $k = 4$



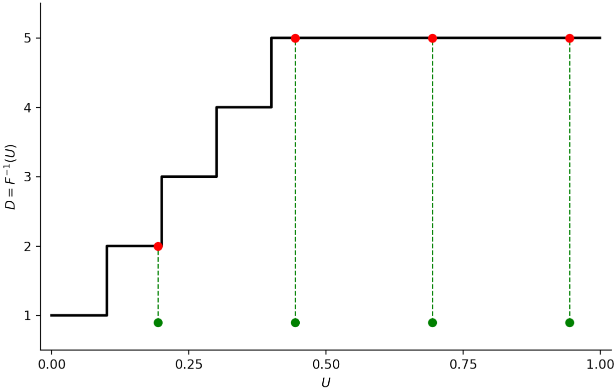
Rotation Sampling for 5-arm Distribution, $k = 4$



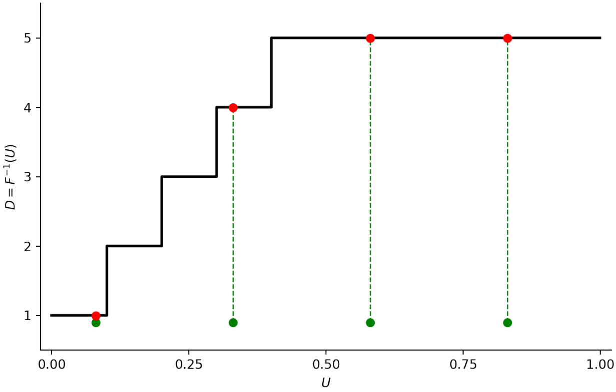
Rotation Sampling for 5-arm Distribution, $k = 4$



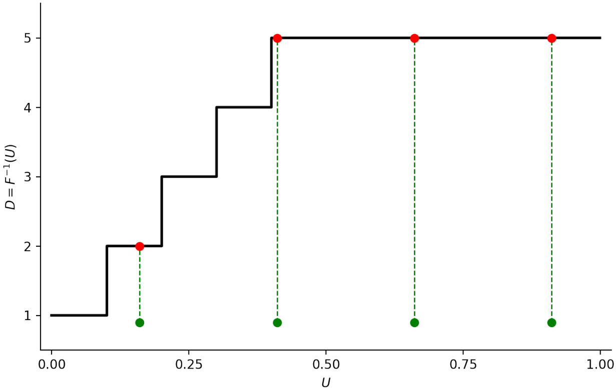
Rotation Sampling for 5-arm Distribution, $k = 4$



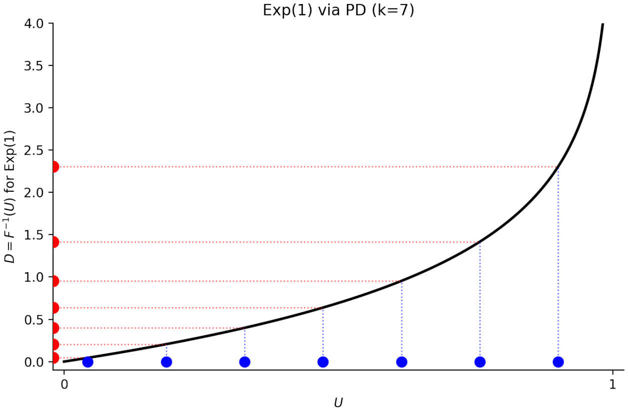
Rotation Sampling for 5-arm Distribution, $k = 4$



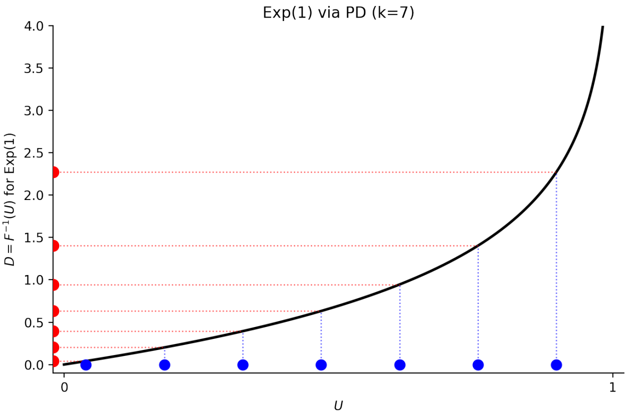
Rotation Sampling for 5-arm Distribution, $k = 4$



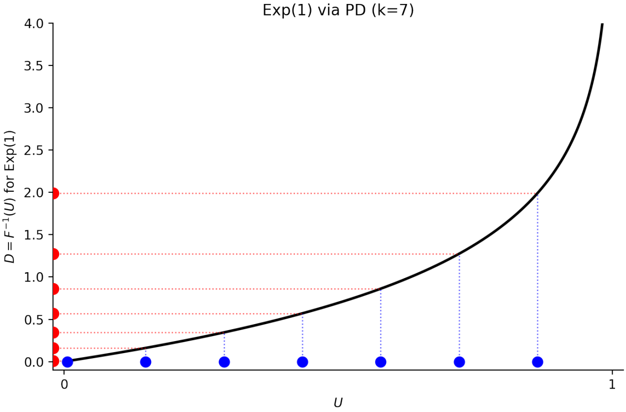
Rotation Sampling for Exp(1) with k=7



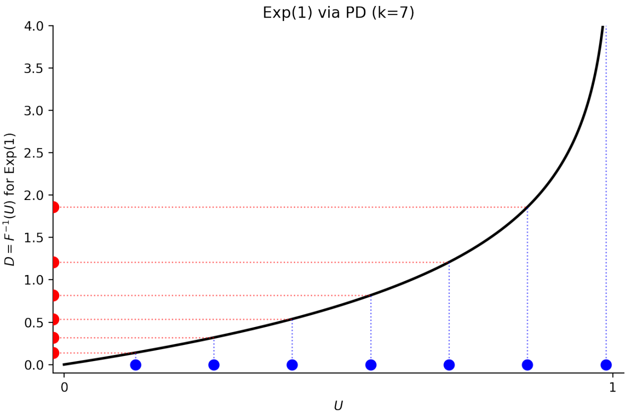
Rotation Sampling for Exp(1) with k=7



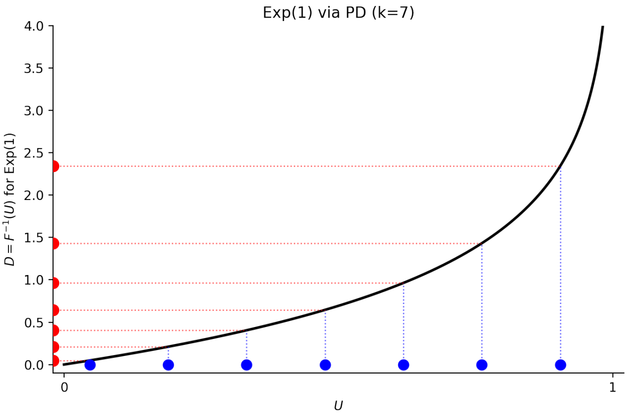
Rotation Sampling for Exp(1) with k=7



Rotation Sampling for Exp(1) with k=7



Rotation Sampling for Exp(1) with k=7



Multivariate Couplings

Options for $(U_i)_{i=1}^k \sim G_U$ with $U_i \sim \mathrm{Unif}[0,1]^m$ include:

- Multivariate Latin hypercube (McKay et al., 1979)
- Scrambled digital nets (Owen, 1995)
- Randomly shifted lattice (Sloan and Joe, 1994)

Multivariate Transport Maps

Brenier Map. For general $\mathcal{D}$, let $T : [0,1]^m \to \mathcal{D}$ solve:

$$\inf_T \int |U - T(U)|^2 dU \quad \text{s.t.} \quad T(U) \sim F$$

Set $D_i = T(U_i)$ to preserve high dispersion of $(U_i)_{i=1}^k \subseteq [0,1]^m$.

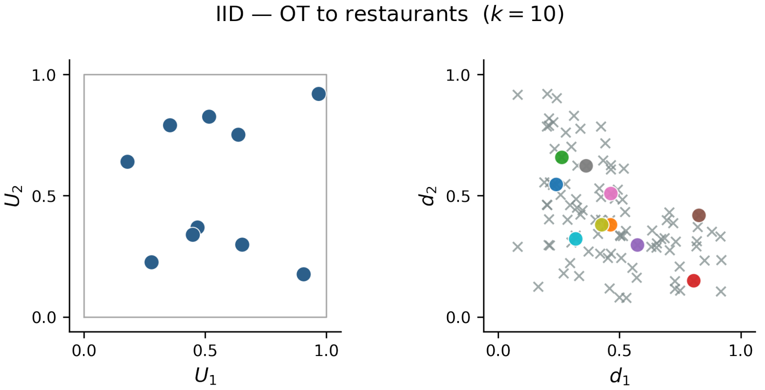
If $F = \otimes_{j=1}^m F_j$, Brenier map is:

$$T^*(U_i) = \left(F_1^{-1}(U_{i1}), \ \ldots, \ F_m^{-1}(U_{im})\right).$$

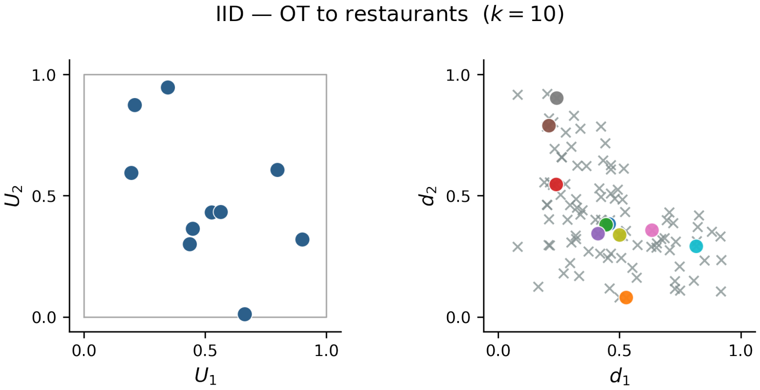
Recall example with irregular $\mathcal{D} = \{d_1, \ldots, d_N\} \subseteq \mathbb{R}^5$.

For $\mathcal{D} \subseteq \mathbb{R}^m$ discrete, this is semi-discrete OT, efficiently solvable.

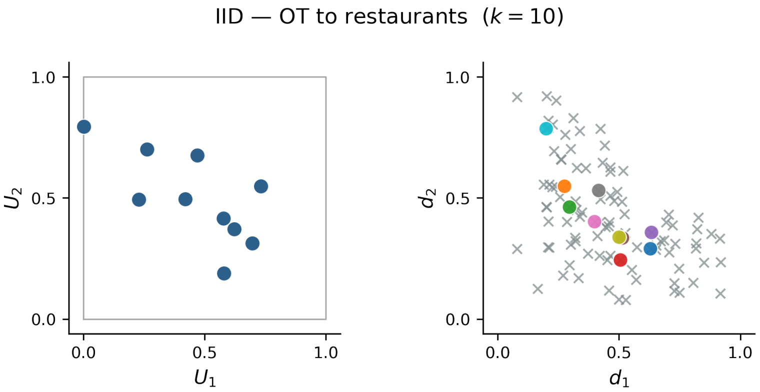
IID coupling: discrete OT map



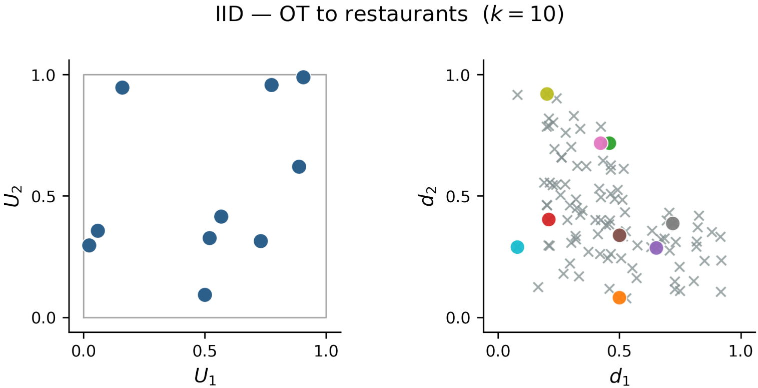
IID coupling: discrete OT map



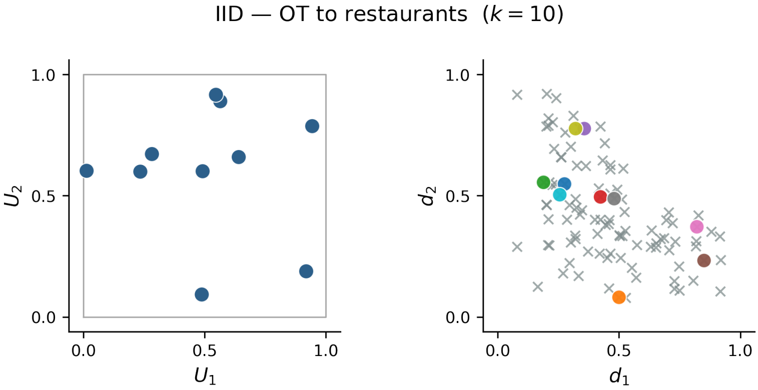
IID coupling: discrete OT map



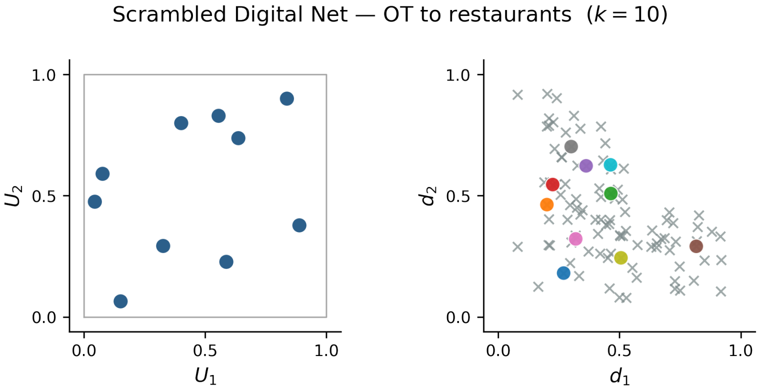
IID coupling: discrete OT map



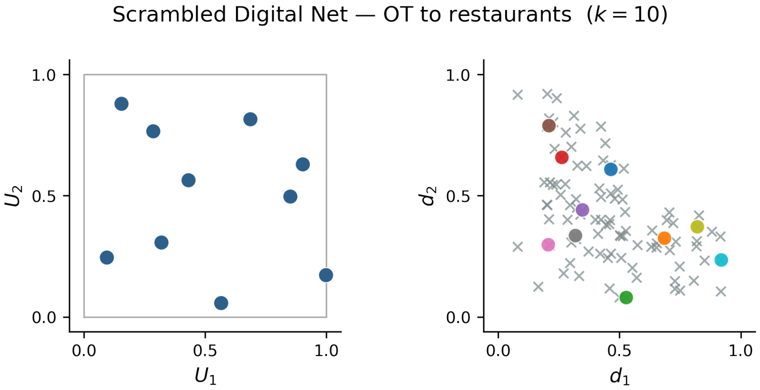
IID coupling: discrete OT map



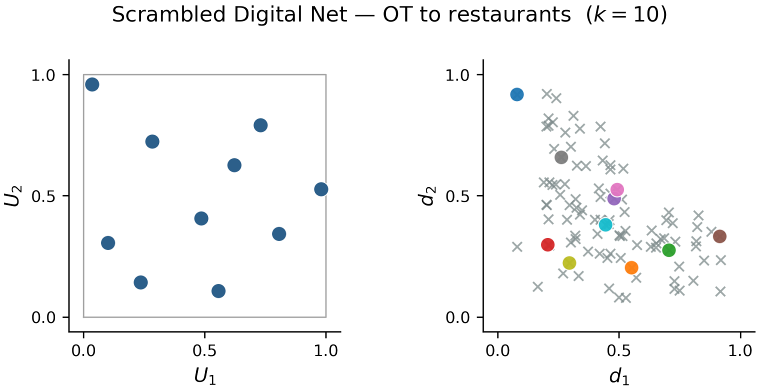
Scrambled Digital Net (Sobol): discrete OT map



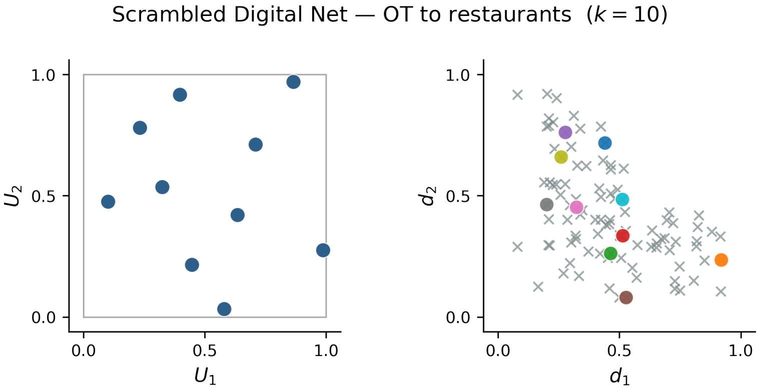
Scrambled Digital Net (Sobol): discrete OT map



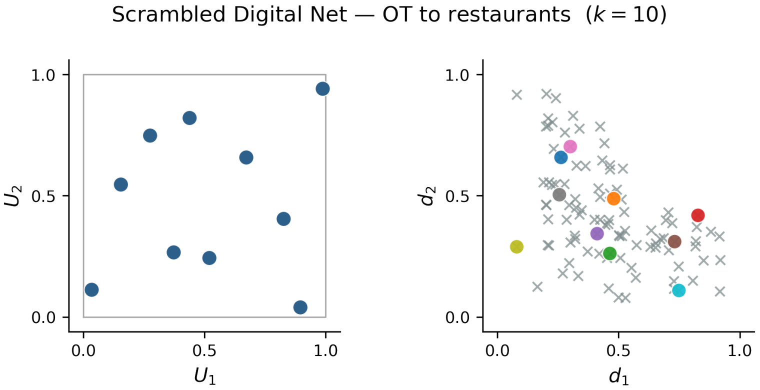
Scrambled Digital Net (Sobol): discrete OT map



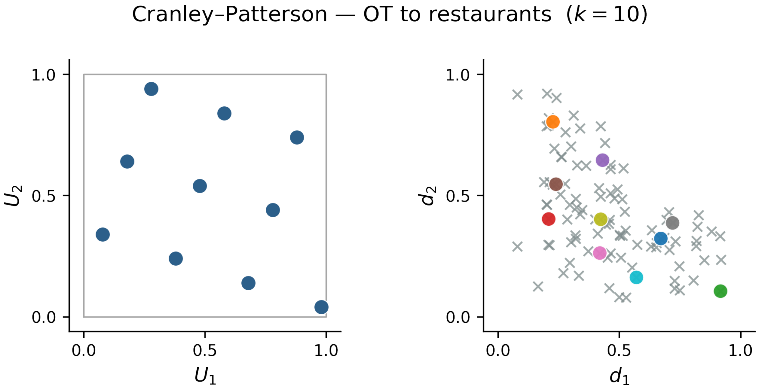
Scrambled Digital Net (Sobol): discrete OT map



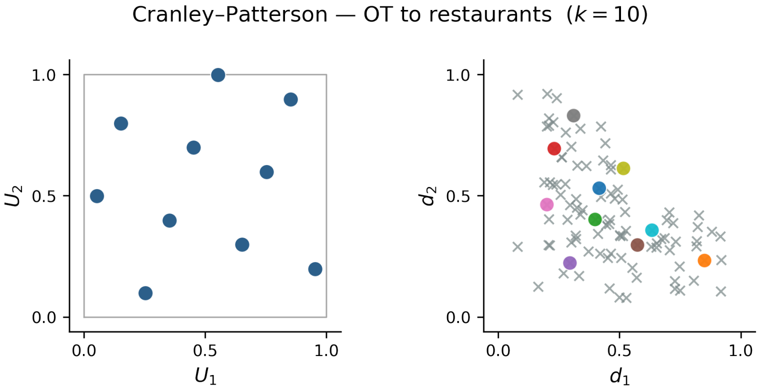
Scrambled Digital Net (Sobol): discrete OT map



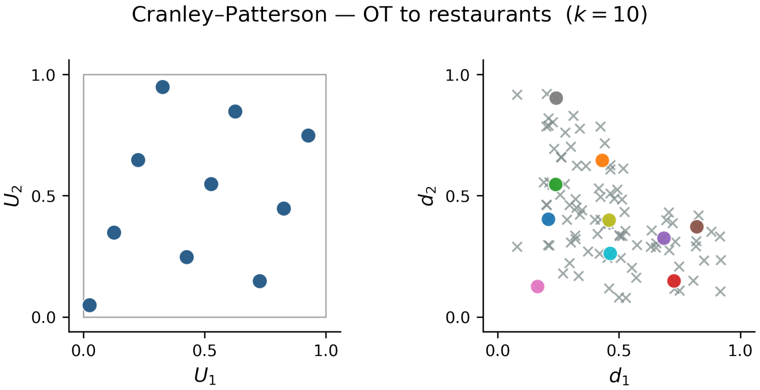
Shifted Lattice coupling: discrete OT map



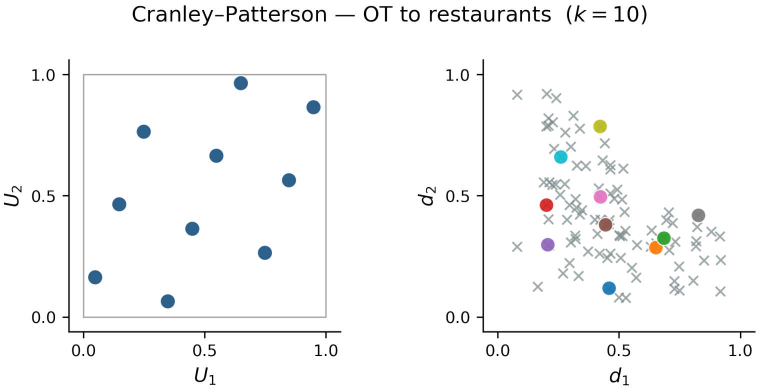
Shifted Lattice coupling: discrete OT map



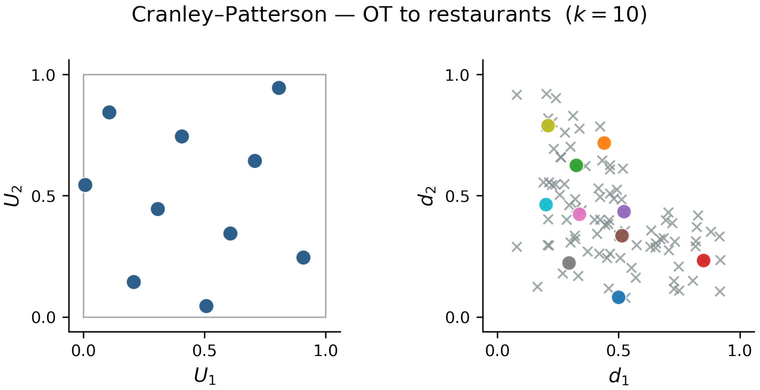
Shifted Lattice coupling: discrete OT map



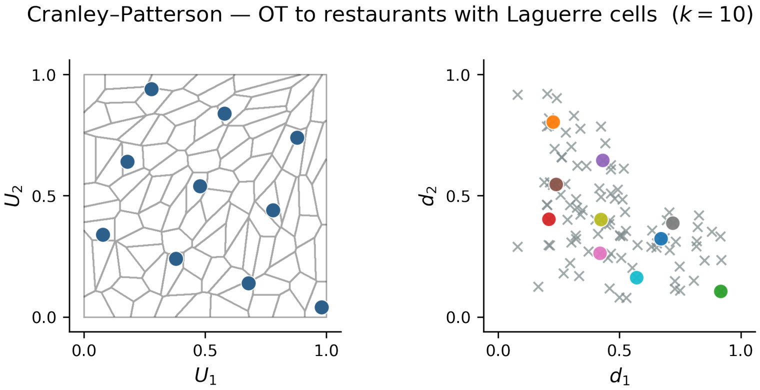
Shifted Lattice coupling: discrete OT map



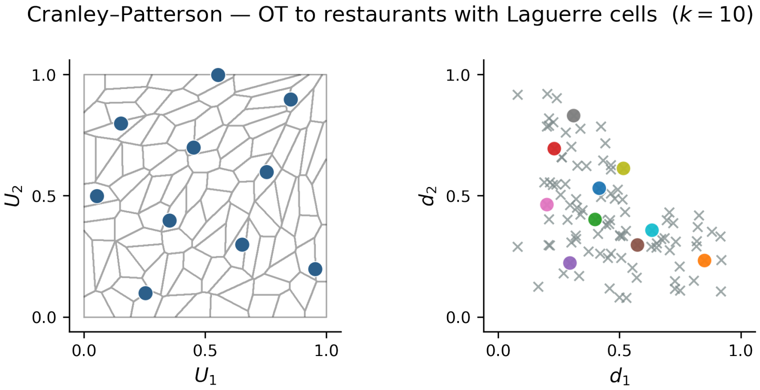
Shifted Lattice coupling: discrete OT map



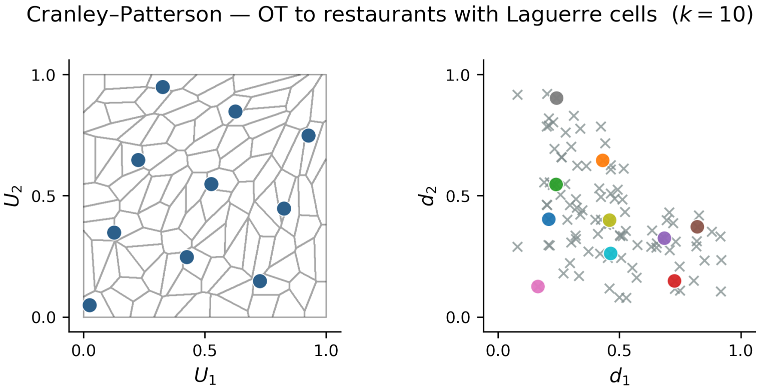
Shifted Lattice coupling: Laguerre cells on $[0, 1]^2$



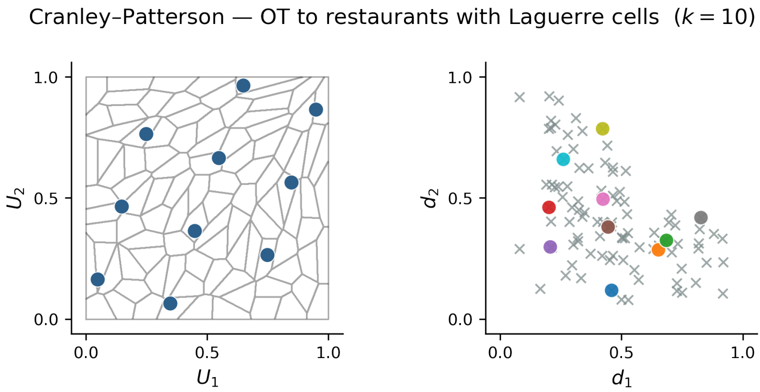
Shifted Lattice coupling: Laguerre cells on $[0,1]^2$



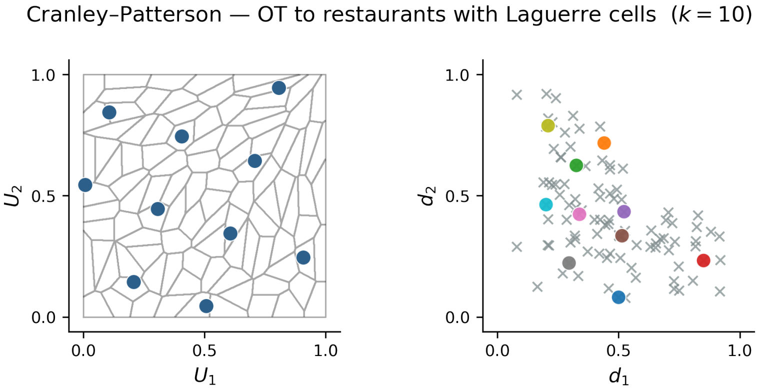
Shifted Lattice coupling: Laguerre cells on $[0, 1]^2$



Shifted Lattice coupling: Laguerre cells on $[0,1]^2$



Shifted Lattice coupling: Laguerre cells on $[0,1]^2$



Recap

Summary. We do the following:

1. Use covariates to match units into tight groups of size k.
2. Draw $(D_i)_{i=1}^k \sim G$ with $G_i = F$ highly dispersed over $\mathcal{D}$.

How does this affect efficiency of $\widehat{\theta} = E_n[s_i(D_i)]$? We show

$$\text{Efficiency Gain} = \text{Dispersion} \times \text{Match Quality}$$

Dispersion and Match Quality

Dispersion

Efficiency Gain = Dispersion × Match Quality

Let $\mathrm{Var}_k(\phi(D_i)) \equiv (k-1)^{-1} \sum_{i \in [k]} (\phi(D_i) - \bar{\phi}_k)^2$.

Definition. (Dispersion) For $(D_i)_{i=1}^k \sim G$ and $\phi : \mathcal{D} \to \mathbb{R}$

$$\mathrm{Disp}_G(\phi) \equiv (k-1)\left(\frac{E_G \, \mathrm{Var}_k(\phi(D_i))}{\mathrm{Var}_F(\phi)} - 1\right).$$

For iid design, $\mathrm{Disp}_G(\phi) = 0$. Have $\mathrm{Disp}_G(\phi) \leq 1$ for any $\phi(\cdot)$.

Proposition. For exchangeable G with $G_i = F$

$$\mathrm{Disp}_G(\phi) = -(k-1)\,\mathrm{Corr}_G(\phi(D_i), \phi(D_j)) \quad i \neq j.$$

Match Quality

How well are units matched? Let $v_{iid}(s) = n^{-1} \sum_i \mathrm{Var}_F(s_i(D))$,

$$\Delta(s) \equiv (2n(k-1))^{-1} \sum_g \sum_{i \neq j \in [k]} \mathrm{Var}_F(s_{ig}(D) - s_{jg}(D))$$

Define **match quality** coefficient

$$Q_k(s) \equiv 1 - \frac{\Delta(s)}{v_{iid}(s)}$$

Example. Suppose $s_i(d) = a_i + b_i \phi(d)$.

$$Q_k(s) = 1 - \frac{\sum_g \sum_{i \neq j \in [k]} (b_{ig} - b_{jg})^2}{2(k-1) \sum_i b_i^2}$$

Efficiency Gain

Model. Let $s_i(d) = a_i + b_i\phi(d)$.

Theorem. (Efficiency) In the univariate model:

$$n\,\mathrm{Var}_G(\widehat{\theta}) = v_{iid}(s) \cdot (1 - \mathrm{Disp}_G(\phi) Q_k(s))$$

$$\text{Efficiency Gain} \equiv \left| \frac{n\,\mathrm{Var}_G(\widehat{\theta}) - v_{iid}(s)}{v_{iid}(s)} \right| = \mathrm{Disp}_G(\phi) \times Q_k(s)$$

For $\widehat{\theta} = E_n[\phi(D_i)]$ efficiency gain $= \mathrm{Disp}_G(\phi)$.

Intuition. For smooth $\phi(\cdot)$, if $D_i \approx D_j$ then $\phi(D_i) \approx \phi(D_j)$.
Dispersed $(D_i)_{i=1}^k$ maximize info about $\phi(\cdot)$.

For $\widehat{\theta} = E_n[s_i(D_i)]$, extra heterogeneity costs $\mathrm{Disp}_G(\phi) \times Q_k(s)$.

Covariate Balance

For $D \in \{0, 1\}$, balanced if $E_n[X_i | D_i = 1] - E_n[X_i | D_i = 0] \approx 0$.

Up to scaling factor, equivalent to $\mathrm{Cov}_n(D_i, X_i) \approx 0$.

Covariate balance for $D \in \mathcal{D}$? Let $\phi : \mathcal{D} \to \mathbb{R}$ and $b : \mathbb{R}^p \to \mathbb{R}$

$$\text{Imbalance:} \quad \rho_n(\phi, b) \equiv \frac{\mathrm{Cov}_n(\phi(D_i), b(X_i))}{(\mathrm{Var}_F(\phi) \cdot \mathrm{Var}_n(b_i))^{1/2}}.$$

Corollary. (Covariate Imblance)

$$E_G[\,\rho_n(\phi, b)^2\,] = n^{-1}\left[1 - \mathrm{Disp}_G(\phi) \cdot Q_k(b)\right].$$

Assigning similar units to highly dissimilar treatments over $\mathcal{D}$ prevents spurious correlations between X and D.

Simulation

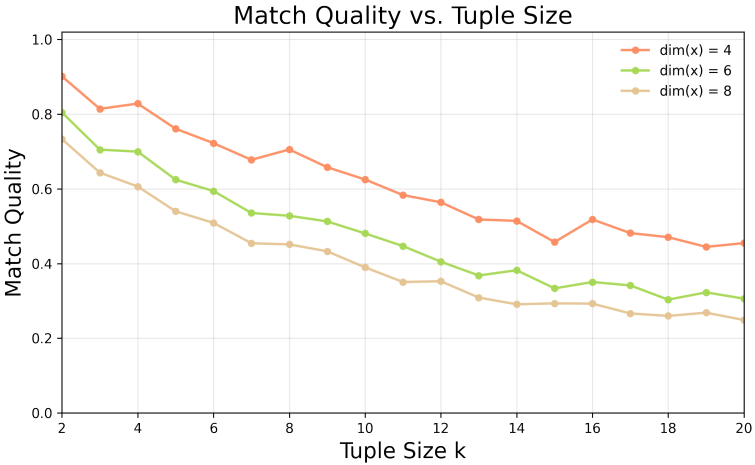
$$\text{Efficiency Gain} = \text{Dispersion} \times \text{Match Quality}$$

Simulation. Let $Y_i(d) = \mathbb{1}' X_i$ and $F = \text{Exp}(1)$. Consider BLP

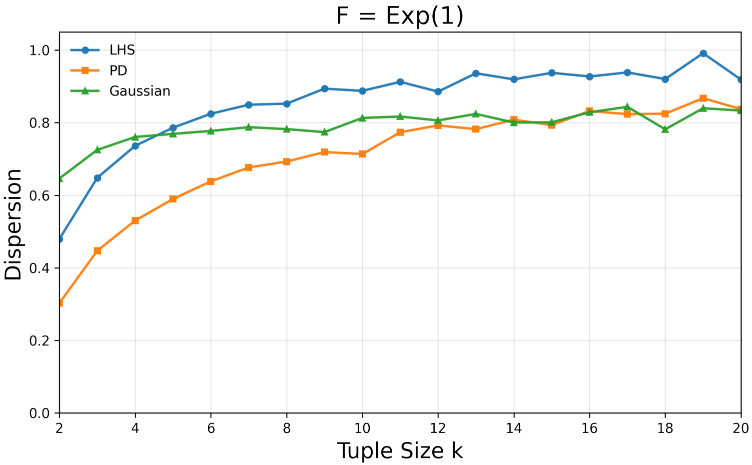
$$s_i(d) = H(d) Y_i(d) = H(d)(\mathbb{1}' X_i)$$

How do match quality, dispersion change with k?

Match Quality

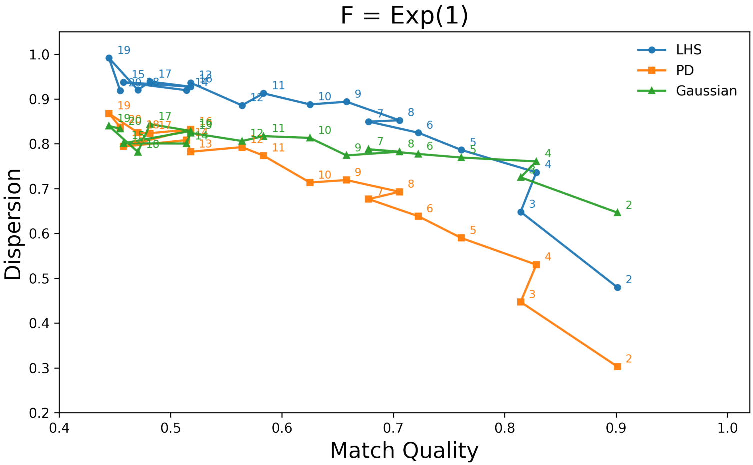


Dispersion

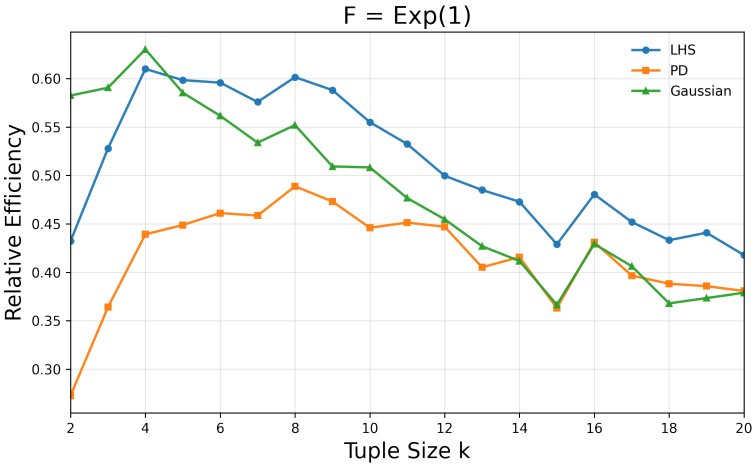


F = Exp(1)
LHS
PD
Gaussian
Dispersion
Tuple Size k
0.0
0.2
0.4
0.6
0.8
1.0
2
4
6
8
10
12
14
16
18
20

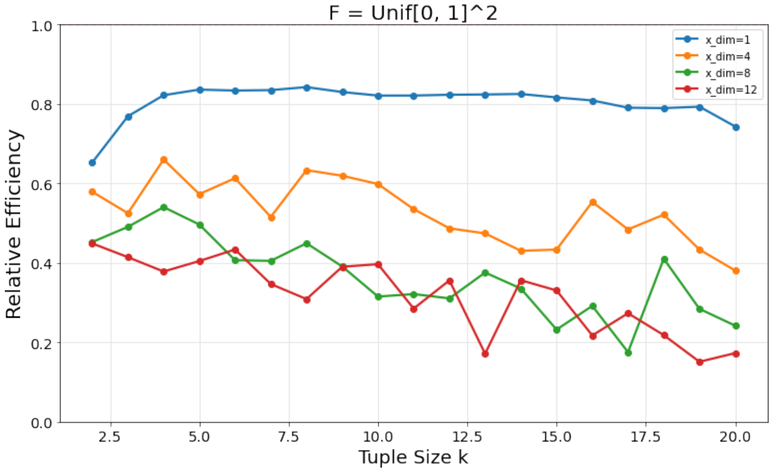
Match Quality vs. Dispersion Frontier



Efficiency Gain $\sim$ Dispersion $\times$ Match Quality



Efficiency Gain, $Y \sim 1 + D_1 + D_2$ and $G = \text{LHS}$



Efficiency Theory

Theory Overview

Theory.

1. Prove efficiency result for general $s_i(\cdot)$.
2. Compare representative couplings G.
3. Asymptotics and CLT.

Previously, showed efficiency result for $s_i(d) = a_i + b_i\phi(d)$.

Influences $s_i(\cdot)$ had common shape $\phi(\cdot)$, common $\mathrm{Disp}_G(\phi)$.

General case: expand $s_i(\cdot)$ over subspaces w/ $\mathrm{Disp}_G(\cdot)$ constant.

Define subspaces to be **principal directions** of the coupling G, what functions is it "good at randomizing."

Principal Directions

Denote $L_0^2(F) \equiv \{\phi \in L_2(F) : E_F[\phi(D)] = 0\}$.

Coupling Operator. Define $U_G : L_0^2(F) \to L_0^2(F)$

$$(U_G\phi)(d) \equiv E_G[\phi(D_i) \mid D_j = d], \quad i \neq j.$$

Decompose $s_i(\cdot)$ over eigenspaces E_m of U_G, $L_0^2(F) = \oplus_{m\geq 1} E_m$.

Theorem. (Dispersion Basis)

(a) If E an eigenspace of U_G, $\mathrm{Disp}_G(\phi) = \mathrm{Disp}_G(\psi)\ \forall \phi, \psi \in E$.

(b) For $\phi \in L_2(F)$, let $P_m\phi$ projection on E_m,

$$\mathrm{Disp}_G(\phi) = \sum_{m\geq 1} \frac{\mathrm{Var}_F(P_m\phi)}{\mathrm{Var}_F(\phi)} \mathrm{Disp}_G(E_m).$$

Latin Hypercube Eigenspaces

$$\mathrm{Disp}_G(\phi) = \sum_{m \geq 1} \frac{\mathrm{Var}_F(P_m \phi)}{\mathrm{Var}_F(\phi)}\ \mathrm{Disp}_G(E_m).$$

Ex. Let $G = \mathrm{LHS}$, $F = \mathrm{Unif}[0,1]$. Bins $J(l) = [(l-1)/k,\ l/k)$,

$$E_{hist} \equiv \left\{\phi \in L_0^2(F) : \phi(d) = \textstyle\sum_l \alpha_l \cdot \mathbb{1}(d \in J(l))\right\}.$$

Have $L_0^2(F) = E_{hist} \oplus E_{hist}^{\perp}$. We prove these are eigenspaces of U_G.

Show $\mathrm{Disp}_G(E_{hist}) = 1$ and $\mathrm{Disp}_G(E_{hist}^{\perp}) = 0$.

For $G = \mathrm{LHS}$, by dispersion basis theorem:

$$\mathrm{Disp}_G(\phi) = \frac{\mathrm{Var}_F(P_{hist}\phi)}{\mathrm{Var}_F(\phi)}$$

Latin Hypercube Eigenspaces

$$\text{Disp}_G(\phi) = \frac{\text{Var}_F(P_{hist}\phi)}{\text{Var}_F(\phi)}$$

Smoothness. Dispersion large if $\phi(\cdot)$ well approximated by histograms on partition $J(l) = [(l-1)/k,\ l/k)$.

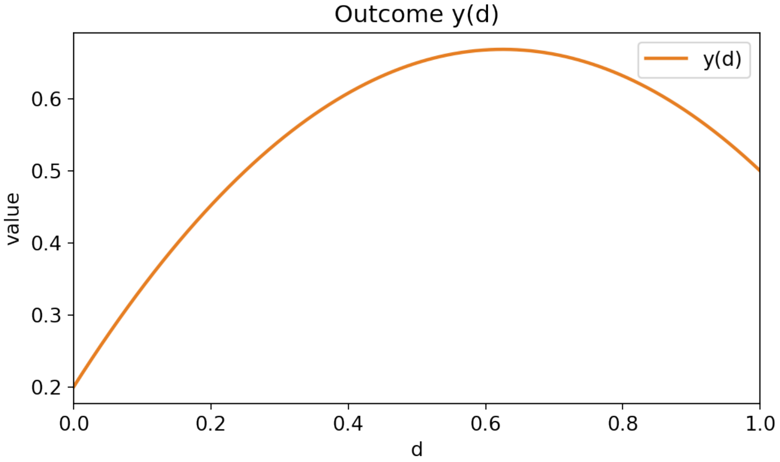
As k grows, partition refines. High dispersion for rougher $\phi(\cdot)$.

Show $\text{Disp}_G(\phi) \to 1$ as $k \to \infty$, uniformly over Lipschitz class.

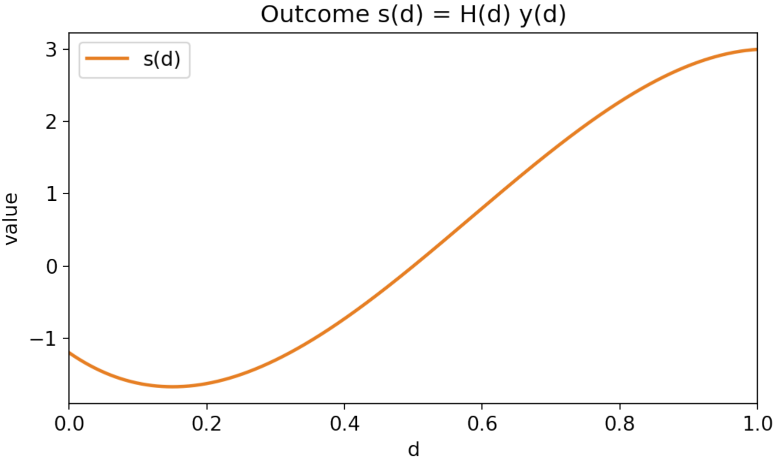
Proposition. For any G, if $L_0^2(F) = E \oplus E^\perp$ eigenspaces of U_G

$$E = \underset{\phi \neq c}{\text{argmax}}\ \text{Disp}_G(\phi), \qquad E^\perp = \underset{\phi \neq c}{\text{argmin}}\ \text{Disp}_G(\phi).$$

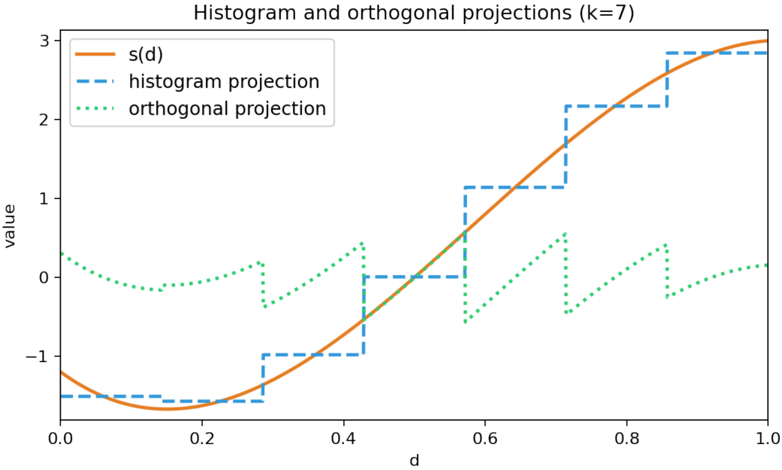
Latin Hypercube Projections



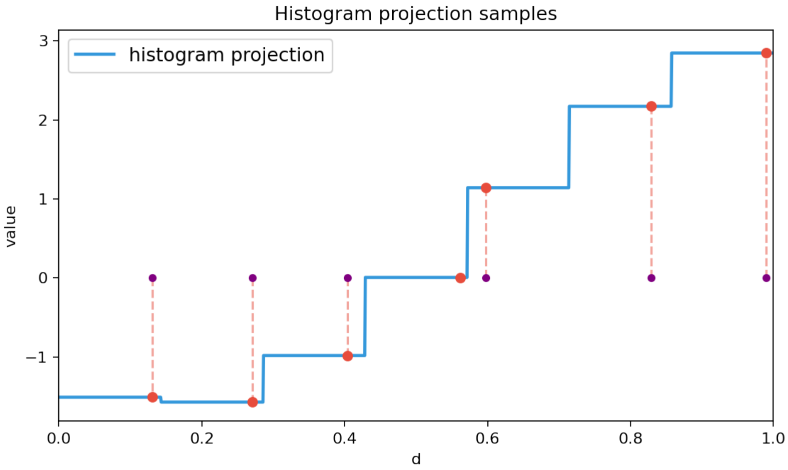
Latin Hypercube Projections



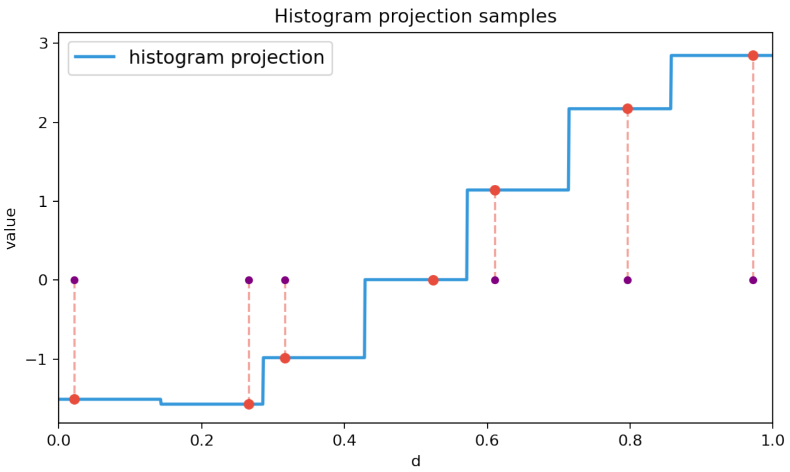
Latin Hypercube Projections



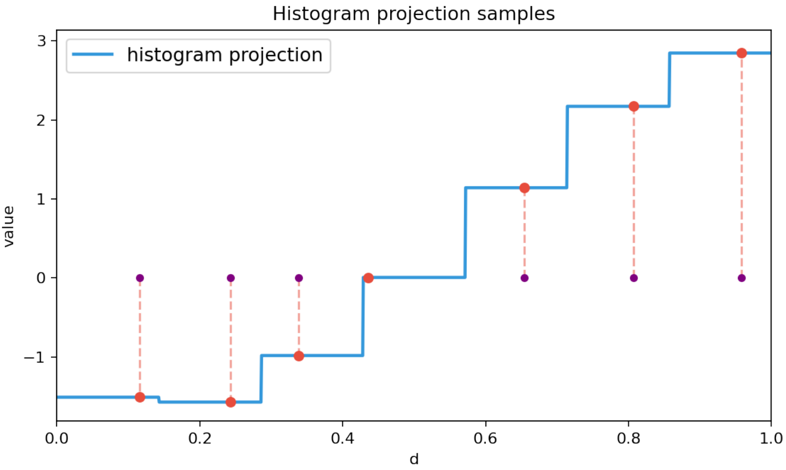
Latin Hypercube Projections



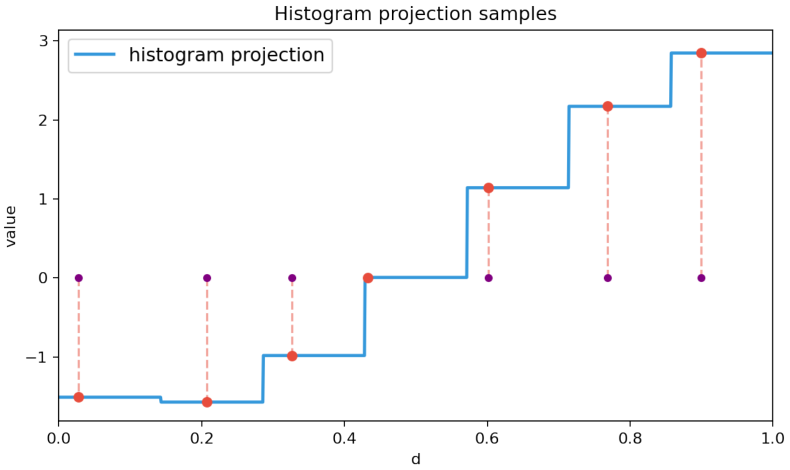
Latin Hypercube Projections



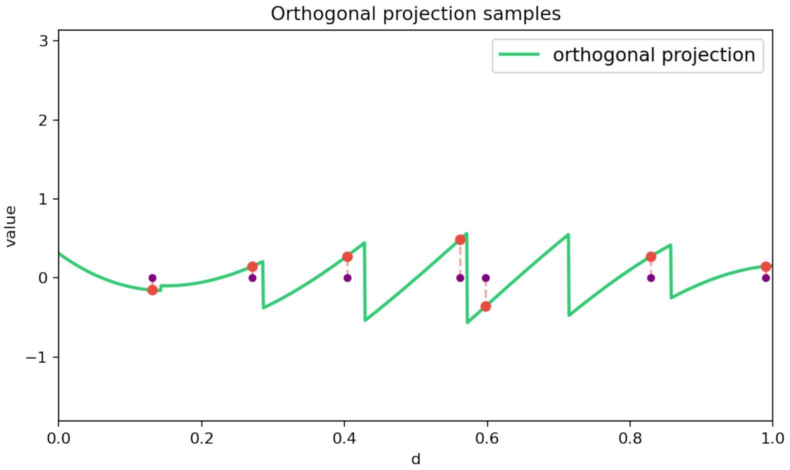
Latin Hypercube Projections



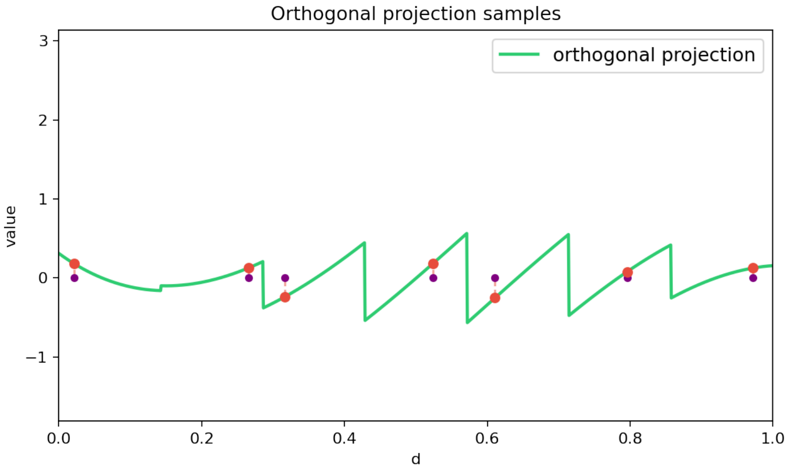
Latin Hypercube Projections



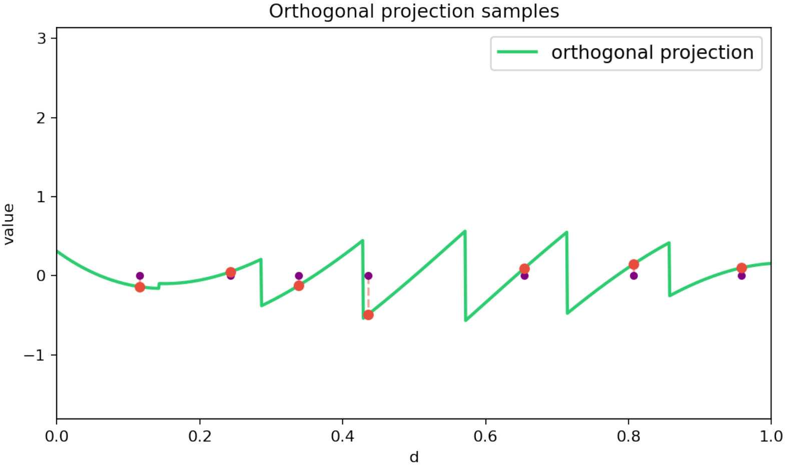
Latin Hypercube Projections



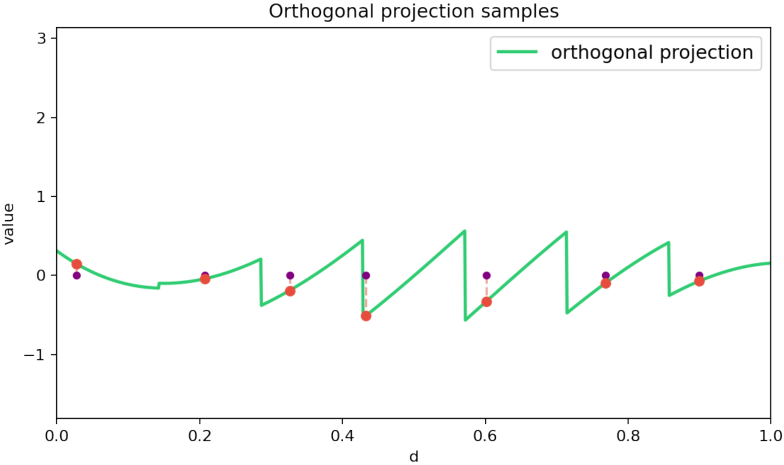
Latin Hypercube Projections



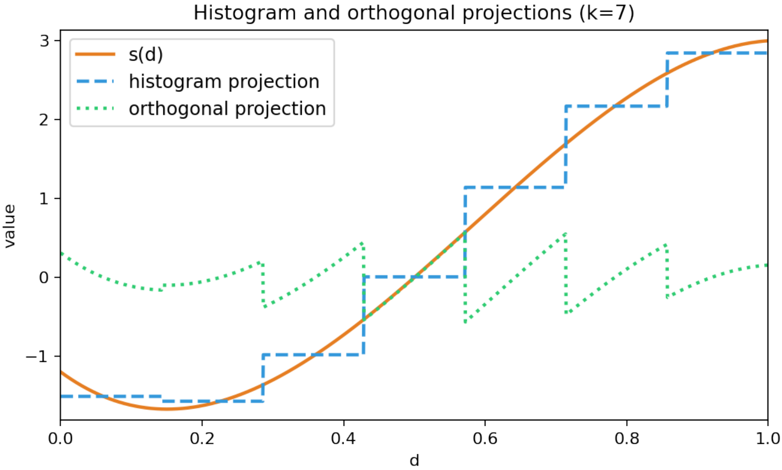
Latin Hypercube Projections



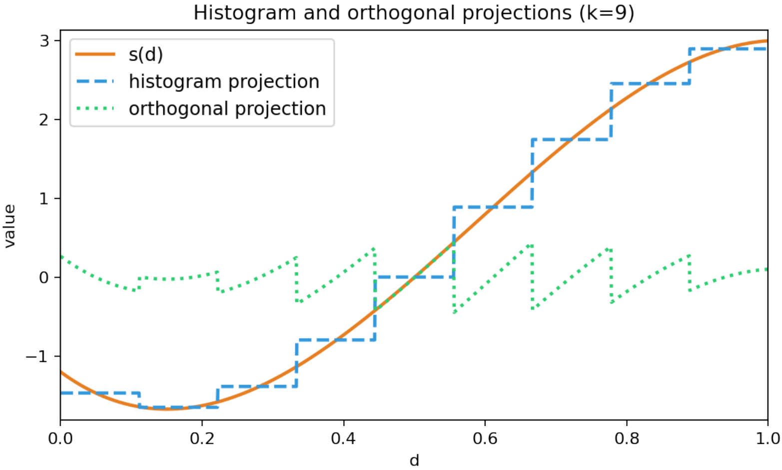
Latin Hypercube Projections



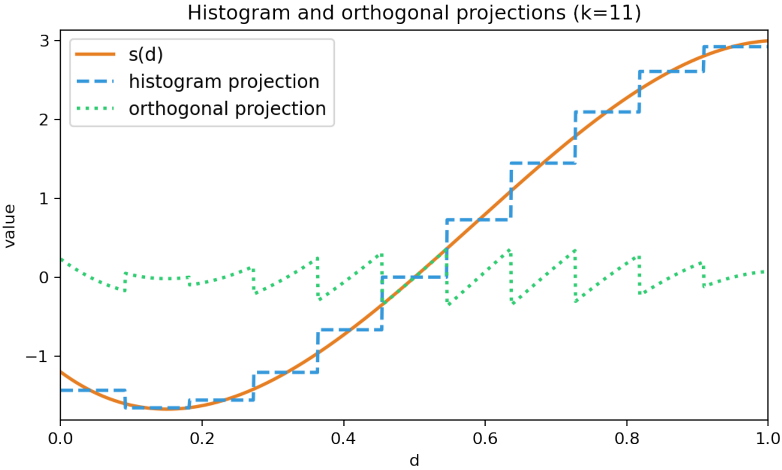
Latin Hypercube Projections



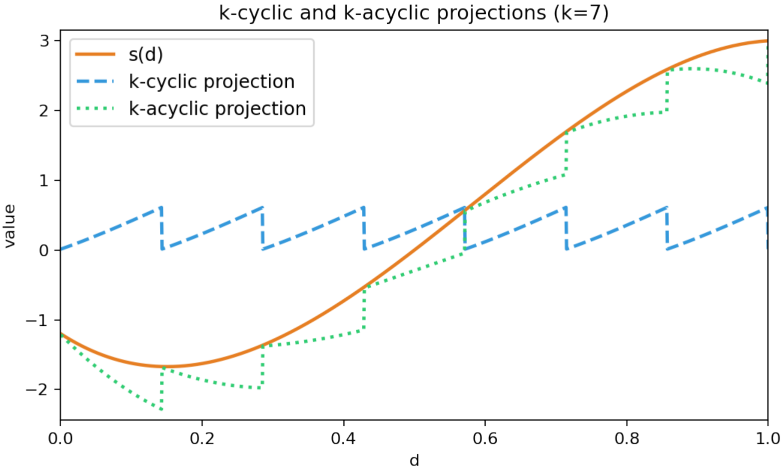
Latin Hypercube Projections



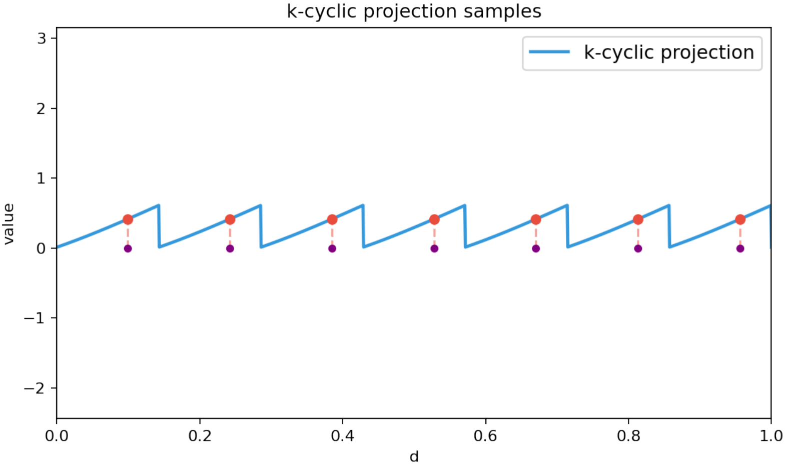
Latin Hypercube Projections



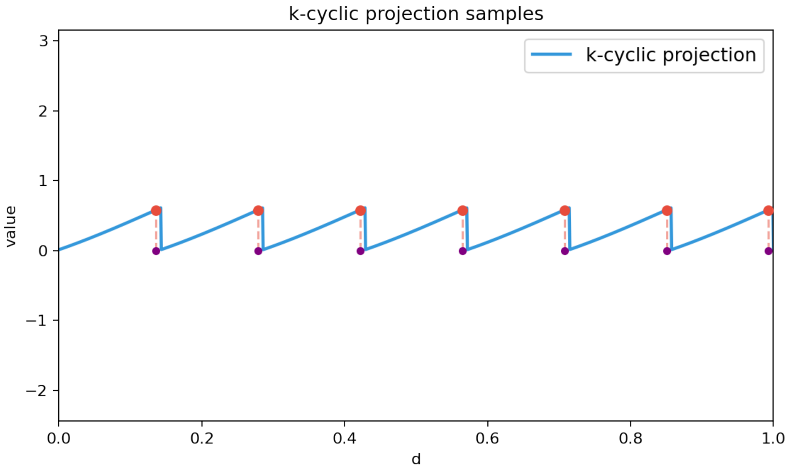
Rotation Sampling Projections



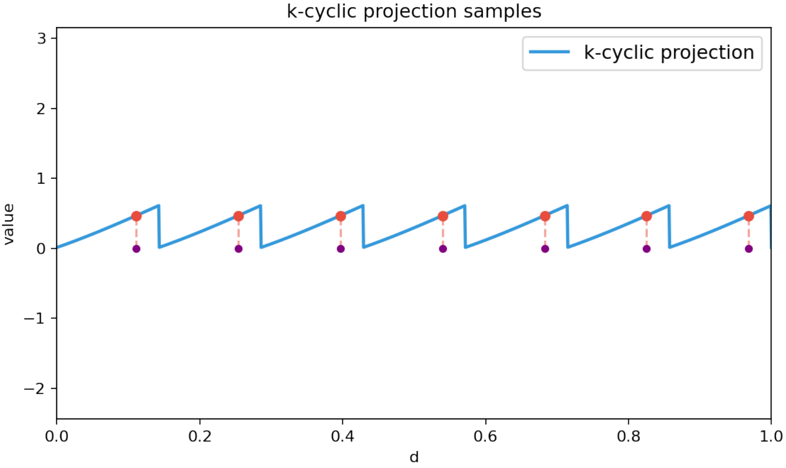
Rotation Sampling Projections



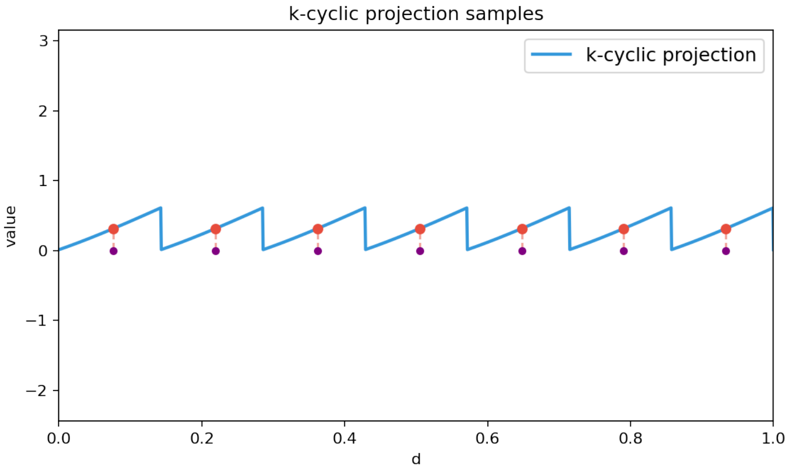
Rotation Sampling Projections



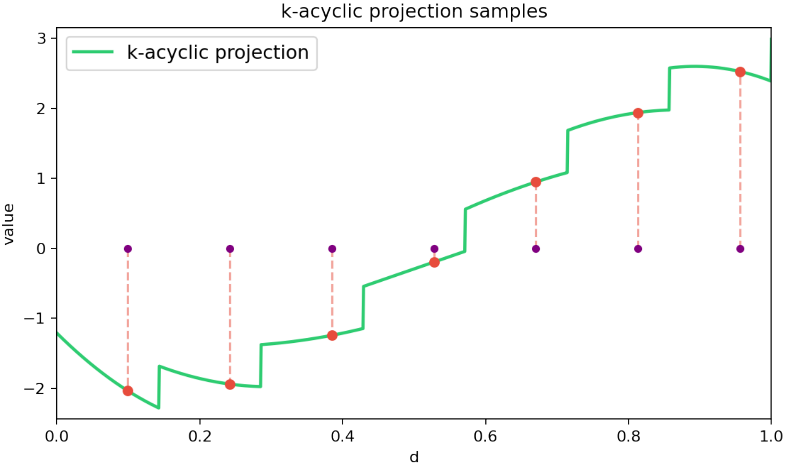
Rotation Sampling Projections



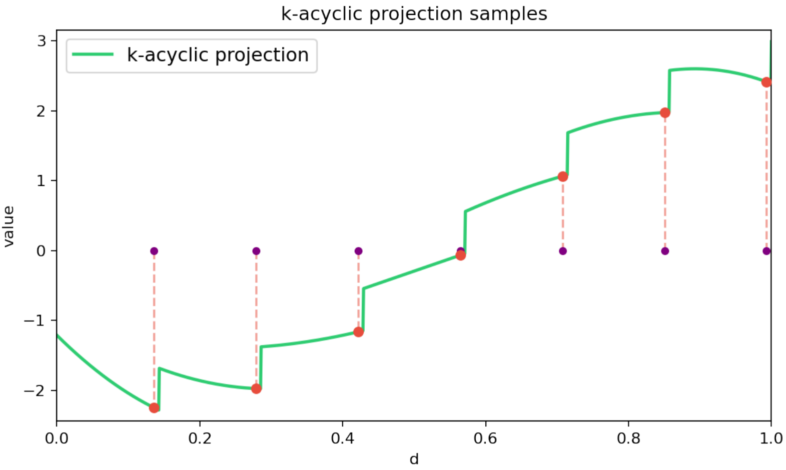
Rotation Sampling Projections



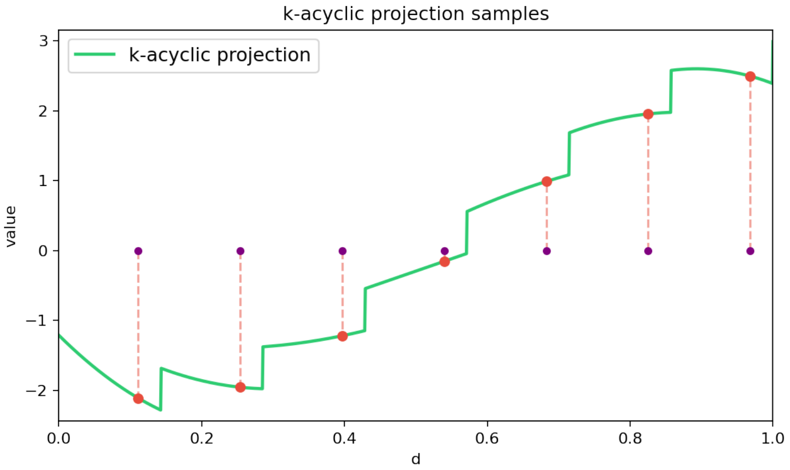
Rotation Sampling Projections



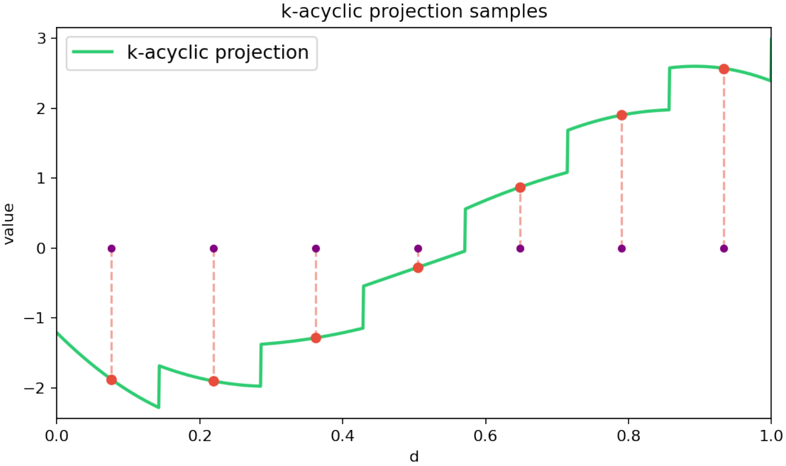
Rotation Sampling Projections



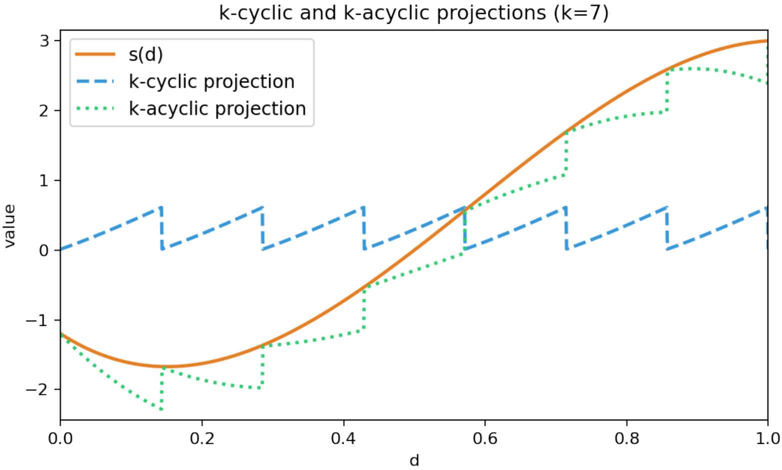
Rotation Sampling Projections



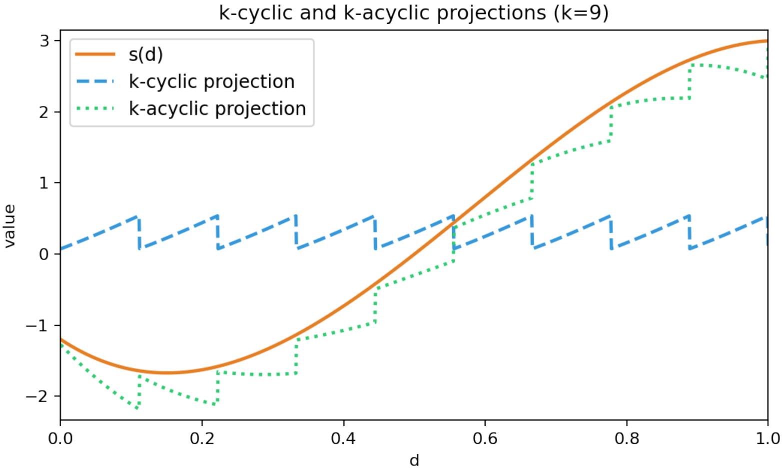
Rotation Sampling Projections



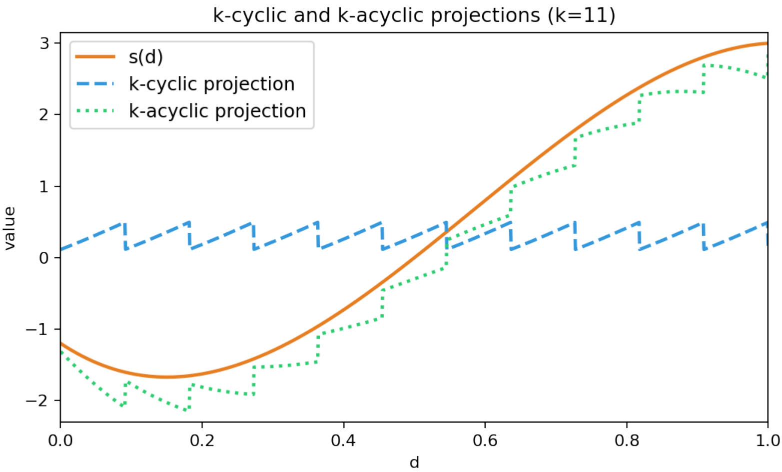
Rotation Sampling Projections



Rotation Sampling Projections



Rotation Sampling Projections



Efficiency Theorem

Theorem. Let $s_i^m = P_m(s_i)$ projection on E_m. Then

$$\text{Efficiency}(G) = \sum_m w_m(s) \cdot \text{Disp}_G(E_m) Q_k(s^m)$$

Weights $w_m(s)$ reflect the shape of $s_i(\cdot)$. In particular

$$w_m(s) = \frac{\sum_i \text{Var}_F(P_m s_i)}{\sum_i \text{Var}_F(s_i)}$$

$w_m(s)$ large $\iff$ $s_i(\cdot)$ well-approx on high dispersion eigenspaces.

For nonparametric $G =$ LHS, RS, smoothness restriction on $s_i(\cdot)$.

For $G =$ Gaussian, linear parametric restriction on $s_i(\cdot)$.

Coupling Comparisons

Latin Hypercube

$$\text{Efficiency}(G) = \sum_m w_m(s) \cdot \text{Disp}_G(m) Q_k(s^m)$$

Corollary. (LHS) Let $s_i^{hist} = P_{hist}(s_i)$ projection on E_{hist}. Then

$$\text{Efficiency}(G) = w_{hist}(s) \cdot Q_k(s^{hist}).$$

Design with $G =$ LHS efficient: units well-matched, $s_i(\cdot)$ smooth.

Robustness. $\text{Disp}_G = 0$ on "bad" space $E_{hist}^{\perp}$, like iid design.

Rotation Sampling

$$\text{Efficiency}(G) = \sum_m w_m(s) \cdot \text{Disp}_G(m) Q_k(s^m)$$

For $G = \text{RS}$, show eigenspaces of U_G are $L_0^2(F) = E_{acyc} \oplus E_{cyc}$

$$E_{cyc} \equiv \left\{ \phi : \phi(x \oplus 1/k) = \phi(x),\ \forall x \in [0,1] \right\}.$$

Have $\text{Disp}_G(E_{acyc}) = 1$ and $\text{Disp}_G(E_{cyc}) = -(k-1)$.

Corollary. (Rotation Sampling)

$$\text{Efficiency}(G) = w_a(s) \cdot Q_k(s^a) - w_c(s) \cdot (k-1) \cdot Q_k(s^c).$$

Robustness Comparisons

$$\text{Efficiency}(G) = w_a(s) \cdot Q_k(s^a) - w_c(s) \cdot (k-1) \cdot Q_k(s^c).$$

Lattice rules behave like clustered randomization in worst case.

Heavily penalized for high-frequency cyclic component of $s_i(\cdot)$.

Non-smooth, cyclic $s_i(\cdot)$ unlikely in economics applications.

For univariate $\mathcal{D} \subseteq \mathbb{R}$, we recommend LHS.

For $\mathcal{D} \subseteq \mathbb{R}^m$ with $m > 1$, shifted lattice rules or digital nets.

Comparison with Gaussian Copula

Gaussian copula: high dispersion space $E_L = \{\phi : \phi(d) = a + bd\}$.

Theorem. (Gaussian) Let $s_i^L = P_L(s_i)$. For $G =$ Gaussian,

$$\text{Efficiency}(G) = w_L(s) \cdot Q_k(s^L) + O(k^{-1}).$$

Efficiency requires $s_i(\cdot)$ approx linear. For BLP, if $Y_i(d) = a_i + b_i d$, have $s_i(d) = Y_i(d)H(d)$ quadratic. Similarly,

$$\text{Disp}_G(\phi) = \frac{\text{Var}_F(P_L\phi)}{\text{Var}_F(\phi)} + O(k^{-1})$$

For $\phi \in E_L^\perp$ nonlinear, $\text{Disp}_G(\phi) \to 0$ as $k \to \infty$.

As tuple size k increases, Gaussian copula based design can have **worse dispersion** and worse match quality.

Conclusion

Coupling Designs. New tool to randomize efficiently, balance covariates in experiments with complex treatments, $\mathcal{D} \subseteq \mathbb{R}^m$.

Match units into k-tuples, draw $(D_i)_{i=1}^k \sim G$ highly dispersed over $\mathcal{D} \subseteq \mathbb{R}^m$ using Monte Carlo + OT construction.

$$\text{Efficiency Gain} = \text{Dispersion} \times \text{Match Quality}$$

Theory. Expanded $s_i(\cdot)$ over eigenspaces of coupling operator, representing "principal directions" of G for sampling.

Allowed detailed comparison of $G =$ LHS vs. RS vs. Gaussian.

Asymptotics and inference methods in the paper.