Monte Carlo Methods

Introduction

Generating Samples

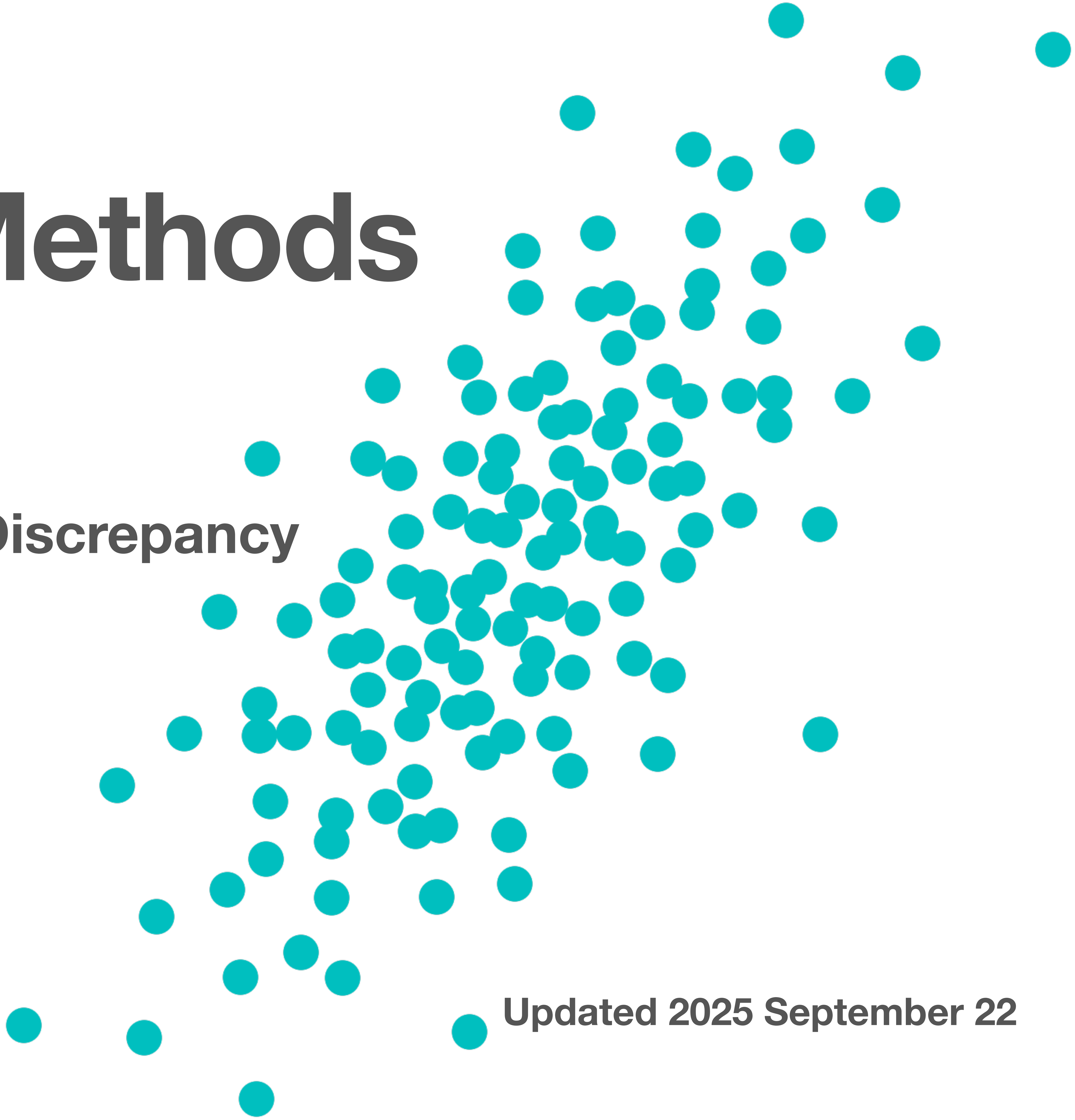
Markov Chain Monte Carlo + Discrepancy

Improving Efficiency

Selected Topics

Git website and repository

Canvas



Fred Hickernell, Fall 2025

Updated 2025 September 22

Markov Chain Monte Carlo (MCMC) + Discrepancy

Owen, Chapters 11– 12

Assignment 3 due ??

Project Selection due Oct 1

MATH 565 Monte Carlo Methods, Fred Hickernell, Fall 2025

Rate your confidence heading into Test 1

- Go to menti.com
- Use code 4923 2283

How do we generate samples when the distribution is complicated?

How do we generate samples when the distribution is complicated?

- Inverse cumulative distribution function applied to uniform samples works for simple cases

How do we generate samples when the distribution is complicated?

- Inverse cumulative distribution function applied to uniform samples works for simple cases
- Correlated multivariate Gaussian distributions are possible through affine transformations of univariate Gaussians

How do we generate samples when the distribution is complicated?

- Inverse cumulative distribution function applied to uniform samples works for simple cases
- Correlated multivariate Gaussian distributions are possible through affine transformations of univariate Gaussians
- Acceptance-rejection sampling

How do we generate samples when the distribution is complicated?

- Inverse cumulative distribution function applied to uniform samples works for simple cases
- Correlated multivariate Gaussian distributions are possible through affine transformations of univariate Gaussians
- Acceptance-rejection sampling
  - Propose from a simple density that you know bounds

How do we generate samples when the distribution is complicated?

- Inverse cumulative distribution function applied to uniform samples works for simple cases
- Correlated multivariate Gaussian distributions are possible through affine transformations of univariate Gaussians
- Acceptance-rejection sampling
  - Propose from a simple density that you know bounds
  - An unnormalized target density

How do we generate samples when the distribution is complicated?

- Inverse cumulative distribution function applied to uniform samples works for simple cases
- Correlated multivariate Gaussian distributions are possible through affine transformations of univariate Gaussians
- Acceptance-rejection sampling
  - Propose from a simple density that you know bounds
  - An unnormalized target density
  - May reject a lot of samples

How do we generate samples when the distribution is complicated?

- Inverse cumulative distribution function applied to uniform samples works for simple cases
- Correlated multivariate Gaussian distributions are possible through affine transformations of univariate Gaussians
- Acceptance-rejection sampling
  - Propose from a simple density that you know bounds
  - An unnormalized target density
  - May reject a lot of samples
- Markov chain Monte Carlo (MCMC) moves samples toward areas of higher target density

Metropolis–Hastings algorithm

Suppose that

- You really want to sample X with known (unnormalized) target density ϱ
- You have a proposal density $\varrho_{\text{new|old}}$ to select the next point starting from where you are

```
Given X[0]
For i = 0 to n−1
  Generate Z ~ ϱ_new|old( · | X_i), U ~ 𝒰[0,1]
  If U ≤ min(1, ϱ(Z)ϱ_new|old(X_i|Z) / ϱ(X_i)ϱ_new|old(Z|X_i)) = min(1, joint density of first Z then X_i / joint density of first X_i then Z)
    X[i+1]←Z
  Else
    X[i+1]←X[i]
```

$$U \le \min\left(1, \frac{\varrho(Z)\varrho_{\text{new|old}}(X_i|Z)}{\varrho(X_i)\varrho_{\text{new|old}}(Z|X_i)}\right) = \min\left(1, \frac{\text{joint density of first } Z \text{ then } X_i}{\text{joint density of first } X_i \text{ then } Z}\right)$$

Note that you always accept either the new or the old

Metropolis algorithm

If $\varrho_{\text{new}|\text{old}}(x\,|\,z) = \varrho_{\text{new}|\text{old}}(z\,|\,x)$, then this can be simplified

Metropolis sampling proceeds as follows

```
Given X[0]
For i = 0 to n-1
  Generate Z ~ ϱ_new|old(·|X_i), U ~ 𝒰[0,1]
  If U ≤ min(1, ϱ(Z)/ϱ(X_i))
     X[i+1]←Z
  Else
     X[i+1]←X[i]
```

$$Z \sim \varrho_{\text{new}|\text{old}}(\cdot\,|\,X_i), \quad U \sim \mathcal{U}[0,1], \quad U \leq \min\left(1, \frac{\varrho(Z)}{\varrho(X_i)}\right)$$

Advantages of and Challenges in MCMC

- Samples from complicated, unnormalized densities
- Easy to implement
- Proposal density and starting point need tuning to ensure that
  - Explore all the sample space and not miss out
  - Exploit sampling in regions where the density is large

How do we measure quality of MCMC

Short answer: We cannot easily

Longer answer: Given a symmetric, positive definite kernel, $K : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$, the discrepancy, sometimes called the maximum mean discrepancy between two empirical distributions is defined by

$$D^2\left(\{\boldsymbol{x}_i\}_{i=0}^{m-1}, \{\boldsymbol{z}_i\}_{i=0}^{n-1}; K\right) = \frac{1}{m^2}\sum_{i,j=0}^{m-1} K(\boldsymbol{x}_i, \boldsymbol{x}_j) - \frac{2}{mn}\sum_{i,j=0}^{m-1,n-1} K(\boldsymbol{x}_i, \boldsymbol{z}_j) + \frac{1}{n^2}\sum_{i,j=0}^{n-1} K(\boldsymbol{z}_i, \boldsymbol{z}_j)$$

Requires $\mathcal{O}\left(\max(m,n)^2\right)$ operations

What is discrepancy?

The discrepancy measures

- The difference between two probability distributions
- The worst-case error in approximating an expectation/integral $\mu = \mathbb{E}[f(\boldsymbol{X})]$ by a sample mean when f is in the unit ball of a Hilbert space with reproducing kernel K
- The root-mean square error in approximating an expectation/integral $\mu = \mathbb{E}[f(\boldsymbol{X})]$ by a sample mean when f is an instance of a Gaussian process with covariance kernel K

Discrepancies from kernels were popularized by [Hickernell 1998] and [Gretton et al. 2012]

Symmetric, Positive Definite Kernels

- A square matrix, K, is symmetric and positive definite iff

$$\mathsf{K}^\mathsf{T} = \mathsf{K} \quad \text{and} \quad \boldsymbol{c}^T \mathsf{K} \boldsymbol{c} > 0 \text{ for all } \boldsymbol{c} \neq \boldsymbol{0}$$

- A kernel, $K : \mathscr{X} \times \mathscr{X} \to \mathbb{R}$, is symmetric and positive definite iff

$$\mathsf{K} := \big(K(\boldsymbol{x}_i, \boldsymbol{x}_j)\big)_{i,j=0}^{n-1} \qquad \text{is symmetric and positive definite}$$

for all $n \in \mathbb{N}$ and $\{\boldsymbol{x}_i\}_{i=0}^{n-1}$ with distinct elements, e.g.

$$K(\boldsymbol{t}, \boldsymbol{x}) = \exp\big(-\|\boldsymbol{t} - \boldsymbol{x}\|^2 / h^2\big), \quad \mathscr{X} = \mathbb{R}^d, \qquad \text{squared exponential}$$

$$K(\boldsymbol{t}, \boldsymbol{x}) = \prod_{\ell=1}^{d} \left[1 + \frac{\gamma_\ell^2}{2}\big(|t_\ell - 1/2| + |x_\ell - 1/2| - |t_\ell - x_\ell|\big)\right], \quad \mathscr{X} = [0,1]^d$$

centered discrepancy

Discrepancy from kernels

Suppose that

- $K : \mathscr{X} \times \mathscr{X} \to \mathbb{R}$ is a symmetric, positive definite kernel,
- $\mathscr{X}$ is the sample space for random variables/vectors $\boldsymbol{X}$ and $\boldsymbol{Z}$
- $\boldsymbol{X}$ has CDF $F_{\boldsymbol{X}}$ and $\boldsymbol{Z}$ has CDF $F_{\boldsymbol{Z}}$

Then discrepancy measures the difference between two distributions:

$$D^2\left(F_{\boldsymbol{X}}, F_{\boldsymbol{Z}}; K\right) = \mathbb{E}_{\boldsymbol{X},\boldsymbol{X}' \sim F_{\boldsymbol{X}}} K(\boldsymbol{X}, \boldsymbol{X}') - 2\mathbb{E}_{\boldsymbol{X}\sim F_{\boldsymbol{X}}, \boldsymbol{Z}\sim F_{\boldsymbol{Z}}} K(\boldsymbol{X}, \boldsymbol{Z}) + \mathbb{E}_{\boldsymbol{Z},\boldsymbol{Z}' \sim F_{\boldsymbol{Z}}} K(\boldsymbol{Z}, \boldsymbol{Z}')$$

$$= \int_{\mathscr{X}\times\mathscr{X}} K(\boldsymbol{x}, \boldsymbol{x}')\, \mathrm{d}F_{\boldsymbol{X}}(\boldsymbol{x})\mathrm{d}F_{\boldsymbol{X}}(\boldsymbol{x}') - 2\int_{\mathscr{X}\times\mathscr{X}} K(\boldsymbol{x}, \boldsymbol{z})\, \mathrm{d}F_{\boldsymbol{X}}(\boldsymbol{x})\mathrm{d}F_{\boldsymbol{Z}}(\boldsymbol{z})$$

$$+ \int_{\mathscr{X}\times\mathscr{X}} K(\boldsymbol{z}, \boldsymbol{z}')\, \mathrm{d}F_{\boldsymbol{Z}}(\boldsymbol{z})\mathrm{d}F_{\boldsymbol{Z}}(\boldsymbol{z}') \geq 0$$

Discrepancy from kernels

Then discrepancy measures the difference between two distributions:

$$
\begin{aligned}
D^2\big(F_X, F_Z; K\big) &= \int_{\mathcal{X}\times\mathcal{X}} K(\boldsymbol{x},\boldsymbol{x}')\,\mathrm{d}F_X(\boldsymbol{x})\mathrm{d}F_X(\boldsymbol{x}') - 2\int_{\mathcal{X}\times\mathcal{X}} K(\boldsymbol{x},\boldsymbol{z})\,\mathrm{d}F_X(\boldsymbol{x})\mathrm{d}F_Z(\boldsymbol{z}) \\
&\qquad + \int_{\mathcal{X}\times\mathcal{X}} K(\boldsymbol{z},\boldsymbol{z}')\,\mathrm{d}F_Z(\boldsymbol{z})\mathrm{d}F_Z(\boldsymbol{z}') \\
&= \int_{\mathcal{X}\times\mathcal{X}} K(\boldsymbol{x},\boldsymbol{x}')\,\mathrm{d}[F_X(\boldsymbol{x}) - F_Z(\boldsymbol{x})\mathrm{d}[F_X(\boldsymbol{x}') - F_Z(\boldsymbol{x}')] \\
&= \int_{\mathcal{X}\times\mathcal{X}} K(\boldsymbol{x},\boldsymbol{x}')\,\varrho_X(\boldsymbol{x})\varrho_X(\boldsymbol{x}')\,\mathrm{d}\boldsymbol{x}\mathrm{d}\boldsymbol{x}' - 2\int_{\mathcal{X}\times\mathcal{X}} K(\boldsymbol{x},\boldsymbol{z})\,\varrho_X(\boldsymbol{x})\varrho_Z(\boldsymbol{z})\,\mathrm{d}\boldsymbol{x}\mathrm{d}\mathrm{z} \\
&\qquad + \int_{\mathcal{X}\times\mathcal{X}} K(\boldsymbol{z},\boldsymbol{z}')\,\varrho_Z(\boldsymbol{z})\varrho_Z(\boldsymbol{z}')\,\mathrm{d}\boldsymbol{z}\mathrm{d}\boldsymbol{z}' \geq 0
\end{aligned}
$$

Discrepancy with empirical distributions

The formula for the squared discrepancy works if one or both of the distributions is the empirical distribution of a sample.

$$D^2\left(F_X, F_{\{z_i\}_{i=0}^{n-1}}; K\right) = D^2\left(F_X, \{z_i\}_{i=0}^{n-1}; K\right)$$

$$= \int_{\mathcal{X}\times\mathcal{X}} K(\boldsymbol{x}, \boldsymbol{x}')\, \mathrm{d}F_X(\boldsymbol{x})\mathrm{d}F_X(\boldsymbol{x}') - \frac{2}{n}\sum_{i=0}^{n-1}\int_{\mathcal{X}} K(\boldsymbol{x}, z_i)\, \mathrm{d}F_X(\boldsymbol{x})$$

$$+ \frac{1}{n^2}\sum_{i,j=0}^{n-1} K(z_i, z_j)$$

$$D^2\left(\{\boldsymbol{x}_i\}_{i=0}^{m-1}, \{z_i\}_{i=0}^{n-1}; K\right) = \frac{1}{m^2}\sum_{i,j=0}^{m-1} K(\boldsymbol{x}_i, \boldsymbol{x}_j) - \frac{2}{mn}\sum_{i,j=0}^{m-1,n-1} K(\boldsymbol{x}_i, z_j) + \frac{1}{n^2}\sum_{i,j=0}^{n-1} K(z_i, z_j)$$

Unbiased estimates of squared discrepancy

The squared discrepancy is non-negative, but we can construct unbiased and possibly negative estimates of $D^2\left(F_X, F_Z; K\right)$ based on IID samples $\{\boldsymbol{x}_i\}_{i=0}^{m-1}$ and/or $\{z_i\}_{i=0}^{n-1}$

$$D^2_{\text{unb}}\left(F_X, \{z_i\}_{i=0}^{n-1}; K\right) = \int_{\mathcal{X}\times\mathcal{X}} K(\boldsymbol{x}, \boldsymbol{x}')\, \mathrm{d}\varrho_X(\boldsymbol{x})\mathrm{d}\varrho_X(\boldsymbol{x}') - \frac{2}{n}\sum_{i=0}^{n-1}\int_{\mathcal{X}} K(\boldsymbol{x}, z_i)\, \mathrm{d}\varrho_X(\boldsymbol{x})$$

$$+\frac{1}{n(n-1)}\sum_{\substack{i,j=0\\ i\neq j}}^{n-1} K(z_i, z_j)$$

$$D^2_{\text{unb}}\left(\{\boldsymbol{x}_i\}_{i=0}^{m-1}, \{z_i\}_{i=0}^{n-1}; K\right) = \frac{1}{m(m-1)}\sum_{\substack{i,j=0\\ i\neq j}}^{m-1} K(\boldsymbol{x}_i, \boldsymbol{x}_i) - \frac{2}{mn}\sum_{i,j=0}^{m-1,n-1} K(\boldsymbol{x}_i, z_j)$$

$$+\frac{1}{n(n-1)}\sum_{\substack{i,j=0\\ i\neq j}}^{n-1} K(z_i, z_j)$$

Discrepancy as worst case error

Let $K : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ be a symmetric, positive definite kernel. Then

- $\mathcal{H}$:= completion of

  $$\{c_0 K(\cdot, \boldsymbol{x}_0) + \cdots + c_{n-1} K(\cdot, \boldsymbol{x}_{n-1}) : n \in \mathbb{N},\, c_i \in \mathbb{R},\, \boldsymbol{x}_i \in \mathcal{X}\}$$

  is a Hilbert space (vector space + inner product) of functions defined on $\mathcal{X}$ with reproducing kernel K

- Which means $f(\boldsymbol{x}) = \langle K(\cdot, \boldsymbol{x}), f \rangle_{\mathcal{H}}$ for all $f \in \mathcal{H}$, $\boldsymbol{x} \in \mathcal{X}$

- Suppose that $\int_{\mathcal{X}\times\mathcal{X}} K(\boldsymbol{x}, \boldsymbol{x}')\, \mathrm{d}F_{\boldsymbol{X}}(\boldsymbol{x}) \mathrm{d}F_{\boldsymbol{X}}(\boldsymbol{x}')$ is defined and finite, which means that $f \mapsto \int_{\mathcal{X}} f(\boldsymbol{x})\, \mathrm{d}F_{\boldsymbol{X}}(\boldsymbol{x})$ is a bounded linear functional on $\mathcal{H}$

- This means that cubature error, $f \mapsto \int_{\mathcal{X}} f(\boldsymbol{x})\, \mathrm{d}F_{\boldsymbol{X}}(\boldsymbol{x}) - n^{-1} \sum_{i=0}^{n-1} f(\boldsymbol{z}_i)$ is also a bounded linear functional on $\mathcal{H}$

Discrepancy as worst case error (cont'd)

- Reproducing property: $f(\boldsymbol{x}) = \langle K(\cdot, \boldsymbol{x}), f \rangle_{\mathscr{H}}$ for all $f \in \mathscr{H}$, $\boldsymbol{x} \in \mathscr{X}$
- Cubature error, $f \mapsto \int_{\mathscr{X}} f(\boldsymbol{x}) \, \mathrm{d}F_{\boldsymbol{X}}(\boldsymbol{x}) - n^{-1} \sum_{i=0}^{n-1} f(\boldsymbol{z}_i)$, is a bounded linear functional on $\mathscr{H}$
- By the Riesz Representation Theorem there exists a $\zeta \in \mathscr{H}$ such that

$$\int_{\mathscr{X}} f(\boldsymbol{x}) \, \mathrm{d}F_{\boldsymbol{X}}(\boldsymbol{x}) - \frac{1}{n} \sum_{i=0}^{n-1} f(\boldsymbol{z}_i) = \langle \zeta, f \rangle_{\mathscr{H}} \quad \forall f \in \mathscr{H}$$

- By the Cauchy-Schwarz inequality

$$\left| \int_{\mathscr{X}} f(\boldsymbol{x}) \, \mathrm{d}F_{\boldsymbol{X}}(\boldsymbol{x}) - \frac{1}{n} \sum_{i=0}^{n-1} f(\boldsymbol{z}_i) \right| = \left| \langle \zeta, f \rangle_{\mathscr{H}} \right| \leq \|\zeta\|_{\mathscr{H}} \|f\|_{\mathscr{H}} \quad \forall f \in \mathscr{H}$$

Discrepancy as worst case error (cont'd)

$$\left| \int_{\mathcal{X}} f(\boldsymbol{x}) \, \mathrm{d}F_X(\boldsymbol{x}) - \frac{1}{n} \sum_{i=0}^{n-1} f(\boldsymbol{x}_i) \right| = \left| \langle \zeta, f \rangle_{\mathcal{H}} \right| \leq \|\zeta\|_{\mathcal{H}} \|f\|_{\mathcal{H}} \quad \forall f \in \mathcal{H}$$

- By the reproducing property and the Riesz Representation Theorem,

$$\zeta(\boldsymbol{x}') = \langle K(\,\cdot\,, \boldsymbol{x}'), \zeta \rangle_{\mathcal{H}} = \int_{\mathcal{X}} K(\boldsymbol{x}, \boldsymbol{x}') \, \mathrm{d}F_X(\boldsymbol{x}) - \frac{1}{n} \sum_{i=0}^{n-1} K(\boldsymbol{z}_i, \boldsymbol{x}')$$

$$\|\zeta|_{\mathcal{H}}^2 = \langle \zeta, \zeta \rangle_{\mathcal{H}} = \int_{\mathcal{X} \times \mathcal{X}} K(\boldsymbol{x}, \boldsymbol{x}') \, \mathrm{d}F_X(\boldsymbol{x}) \mathrm{d}F_X(\boldsymbol{x}') - \frac{2}{n} \sum_{i=0}^{n-1} \int_{\mathcal{X}} K(\boldsymbol{x}, \boldsymbol{z}_i) \, \mathrm{d}F_X(\boldsymbol{x})$$

$$+ \frac{1}{n^2} \sum_{i,j=0}^{n-1} K(\boldsymbol{z}_i, \boldsymbol{z}_j) = D^2\left(F_X, \{\boldsymbol{z}_i\}_{i=0}^{n-1}; K\right)$$

Discrepancy as average case error

Suppose that f is drawn from a Gaussian process with zero mean and covariance kernel $K : \mathscr{X} \times \mathscr{X} \to \mathbb{R}$. (Do not need to specify the sample space for f, which will almost surely have less smoothness than the Hilbert space with reproducing kernel K.)

The mean squared error for a deterministic cubature rule is

$$\mathbb{E}_{f \in \mathcal{GP}(0,K)} \left| \int_{\mathcal{X}} f(\boldsymbol{x}) \, \mathrm{d}F_{\boldsymbol{X}}(\boldsymbol{x}) - \frac{1}{n} \sum_{i=0}^{n-1} f(\boldsymbol{z}_i) \right|^2$$

$$= \int_{\mathcal{X} \times \mathcal{X}} \mathbb{E}_{f \in \mathcal{GP}(0,K)} \left[f(\boldsymbol{x}) f(\boldsymbol{x}') \right] \mathrm{d}F_{\boldsymbol{X}}(\boldsymbol{x}) \mathrm{d}F_{\boldsymbol{X}}(\boldsymbol{x}') - \frac{2}{n} \sum_{i=0}^{n-1} \int_{\mathcal{X}} \mathbb{E}_{f \in \mathcal{GP}(0,K)} \left[f(\boldsymbol{x}) f(\boldsymbol{z}_i) \right] \mathrm{d}F_{\boldsymbol{X}}(\boldsymbol{x})$$

$$+ \frac{1}{n^2} \sum_{i,j=0}^{n-1} \int_{\mathcal{X}} \mathbb{E}_{f \in \mathcal{GP}(0,K)} \left[f(\boldsymbol{z}_i) f(\boldsymbol{z}_j) \right]$$

Discrepancy as average case error (cont'd)

The mean squared error for a deterministic cubature rule is

$$\mathbb{E}_{f \in \mathcal{GP}(0,K)} \left| \int_{\mathcal{X}} f(\boldsymbol{x}) \, \mathrm{d}F_{\boldsymbol{X}}(\boldsymbol{x}) - \frac{1}{n} \sum_{i=0}^{n-1} f(\boldsymbol{z}_i) \right|^2$$

$$= \int_{\mathcal{X} \times \mathcal{X}} \mathbb{E}_{f \in \mathcal{GP}(0,K)} \left[f(\boldsymbol{x}) f(\boldsymbol{x}') \right] \mathrm{d}F_{\boldsymbol{X}}(\boldsymbol{x}) \mathrm{d}F_{\boldsymbol{X}}(\boldsymbol{x}') - \frac{2}{n} \sum_{i=0}^{n-1} \int_{\mathcal{X}} \mathbb{E}_{f \in \mathcal{GP}(0,K)} \left[f(\boldsymbol{x}) f(\boldsymbol{z}_i) \right] \mathrm{d}F_{\boldsymbol{X}}(\boldsymbol{x})$$

$$+ \frac{1}{n^2} \sum_{i,j=0}^{n-1} \int_{\mathcal{X}} \mathbb{E}_{f \in \mathcal{GP}(0,K)} \left[f(\boldsymbol{z}_i) f(\boldsymbol{z}_j) \right]$$

$$= \int_{\mathcal{X} \times \mathcal{X}} K(\boldsymbol{x}, \boldsymbol{x}') \, \mathrm{d}F_{\boldsymbol{X}}(\boldsymbol{x}) \mathrm{d}F_{\boldsymbol{X}}(\boldsymbol{x}') - \frac{2}{n} \sum_{i=0}^{n-1} \int_{\mathcal{X}} K(\boldsymbol{x}, \boldsymbol{z}_i) \, \mathrm{d}F_{\boldsymbol{X}}(\boldsymbol{x}) + \frac{1}{n^2} \sum_{i,j=0}^{n-1} \int_{\mathcal{X}} K(\boldsymbol{z}_i, (\boldsymbol{z}_j)$$

$$= D^2 \left(F_{\boldsymbol{X}}, \{\boldsymbol{z}_i\}_{i=0}^{n-1} \right)$$

Observations about discrepancy

- The value of the discrepancy depends on the choice of kernel, K, and the parameters that define it
  - K may be chosen for convenience
  - K may be chosen based on knowledge about properties of f, such as domain, smoothness, periodicity, importance of different coordinates
  - $D(\cdot, \cdot; c^2 K) = |c| D(\cdot, \cdot; K)$
- $D(F_X, \{z\}_{i=0}^{n-1}; K)$ measures the quality of $\{z\}_{i=0}^{n-1}$ for estimating the mean, $\mu = \mathbb{E}[f(X)]$, by the sample mean $\widehat{\mu}_n = n^{-1} \sum_{i=0}^{n-1} f(z)$ for
  - f in a Hilbert space, $\mathscr{H}$, with reproducing kernel K (don't need to know $\|\cdot\|_{\mathscr{H}}$ explicitly) or
  - $f \sim \mathscr{GP}(0, K)$

Observations about discrepancy (cont'd)

- The mean square discrepancy of an IID random sample is

$$\mathbb{E}\left[D^2\left(F_{\boldsymbol{X}}, \{\boldsymbol{z}_i \overset{\text{IID}}{\sim} F_{\boldsymbol{X}}\}_{i=0}^{n-1}; K\right) = \frac{1}{n}\left[\int_{\mathcal{X}} K(\boldsymbol{x},\boldsymbol{x})\,\mathrm{d}F_{\boldsymbol{X}}(\boldsymbol{x}) - \int_{\mathcal{X}\times\mathcal{X}} K(\boldsymbol{x},\boldsymbol{x}')\,\mathrm{d}F_{\boldsymbol{X}}(\boldsymbol{x})\mathrm{d}F_{\boldsymbol{X}}(\boldsymbol{x}')\right]\right.$$

- The deterministic cubature error bound cannot be used constructively to bound the error because the norm of f cannot be estimated

- If $f \sim \mathscr{GP}(0,K)$, and the parameters of K are estimated by empirical Bayes (maximum likelihood), then one can construct credible intervals for the cubature error

$$\mathbb{P}_{f\in\mathscr{GP}(0,K)}\left[\left|\int_{\mathscr{X}} f(\boldsymbol{x})\,\mathrm{d}F_{\boldsymbol{X}}(\boldsymbol{x}) - \frac{1}{n}\sum_{i=0}^{n-1} f(\boldsymbol{z}_i)\right| \le \frac{2.58D\left(F_{\boldsymbol{X}}, \{\boldsymbol{z}_i\}_{i=0}^{n-1}; K\right)}{\sqrt{n}}\right] \ge 99\,\%$$

Example of the Centered discrepancy

$$K(\boldsymbol{t},\boldsymbol{x}) = \prod_{\ell=1}^{d}\left[1+\frac{\gamma_\ell^2}{2}\big(|t_\ell - 1/2| + |x_\ell - 1/2| - |t_\ell - x_\ell|\big)\right], \quad \mathcal{X} = [0,1]^d$$

$$\|f\|_{\mathcal{H}}^2 = \|f(0.5,\ldots,0.5)\|_2^2$$

$$+ \left\| \frac{\partial f(x_1,0.5,\ldots,0.5)}{\gamma_1\,\partial x_1} \right\|_2^2 + \left\| \frac{\partial f(0.5,x_2,0.5,\ldots,0.5)}{\gamma_2\,\partial x_2} \right\|_2^2 + \ldots$$

$$+ \left\| \frac{\partial f(x_1,x_2,0.5,\ldots,0.5)}{\gamma_1\gamma_2\,\partial x_1\partial x_2} \right\|_2^2 + \left\| \frac{\partial f(x_1,0.5,x_3,0.5,\ldots,0.5)}{\gamma_1\gamma_3\,\partial x_1\partial x_3} \right\|_2^2 + \ldots$$

$$+\ldots + \left\| \frac{\partial f(x_1,\ldots,x_d)}{\gamma_1\cdots\gamma_d\,\partial x_1\cdots\partial x_d} \right\|_2^2$$

Gibbs Sampler

Overcoming challenges of MCMC