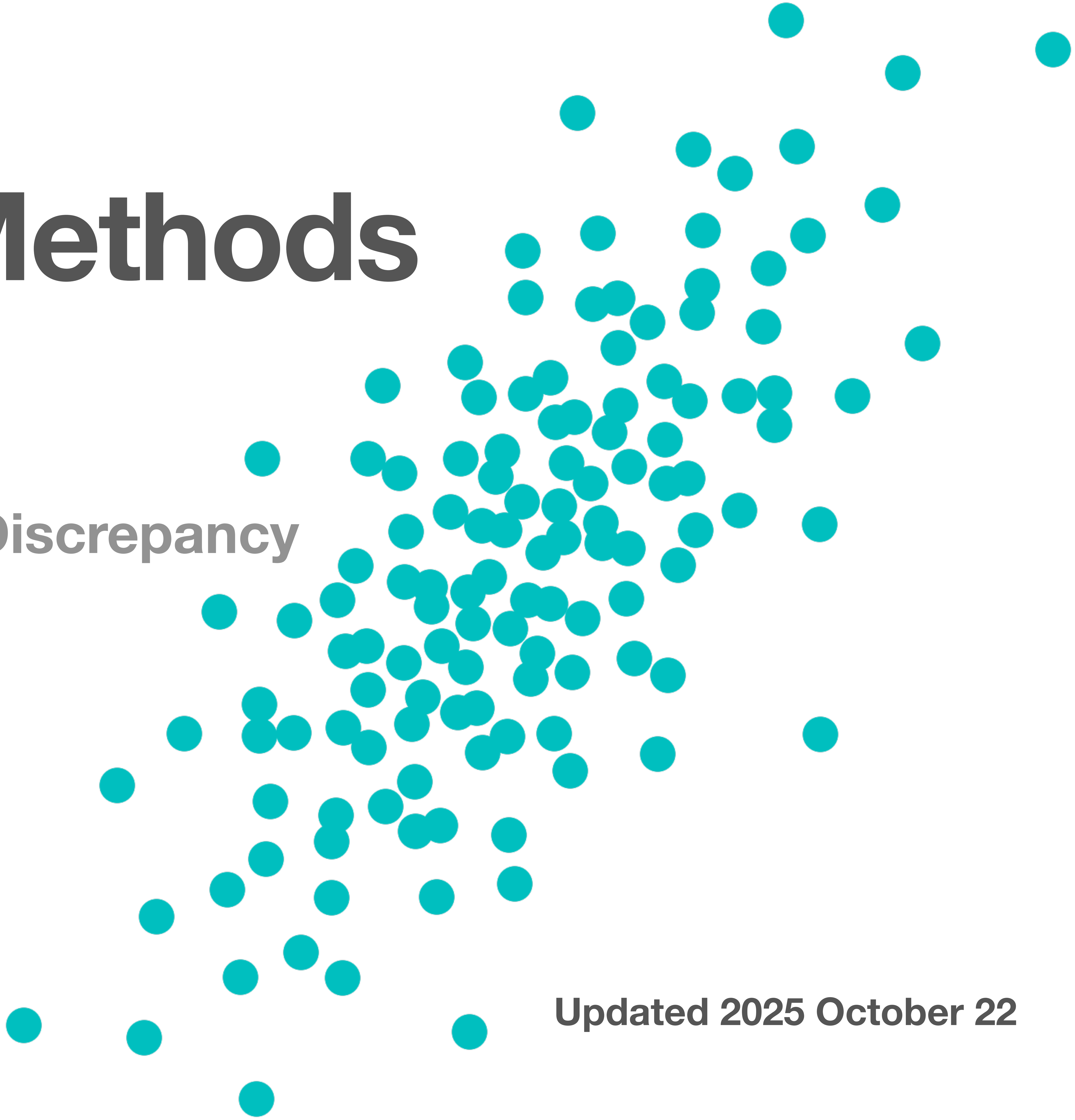
Monte Carlo Methods

Introduction
Generating Samples
Markov Chain Monte Carlo + Discrepancy
Improving Efficiency
Selected Topics

Git website and repository
Canvas



Fred Hickernell, Fall 2025

Updated 2025 October 22

Improving Efficiency

Owen, Chapters ??

Test 2 Oct 29

MATH 565 Monte Carlo Methods, Fred Hickernell, Fall 2025

Variance Reduction

$$\mu = \mathbb{E}[f(\boldsymbol{X})] = \mathbb{E}[g(\boldsymbol{X})], \quad \boldsymbol{X} \sim \mathcal{U}[0,1]^d$$

Is it better to use

$$\hat{\mu}_{f,n} = \frac{1}{n}\sum_{i=0}^{n-1} f(\boldsymbol{X}_i) \quad \text{or} \quad \hat{\mu}_{g,n} = \frac{1}{n}\sum_{i=0}^{n-1} g(\boldsymbol{X}_i), \qquad \boldsymbol{X}_0, \boldsymbol{X}_1, \ldots \overset{\text{IID}}{\sim} \mathcal{U}[0,1]^d$$

to approximate μ? Since both sample means are unbiased, we can compare their variances (mean squared errors):

$$\operatorname{var}(\hat{\mu}_{f,n}) = \frac{\operatorname{var}(f(\boldsymbol{X}))}{n} \qquad \operatorname{var}(\hat{\mu}_{g,n}) = \frac{\operatorname{var}(g(\boldsymbol{X}))}{n}$$

so it depends on the variances of the two functions

Keister example

$$\mu = \int_{\mathbb{R}^d} \cos(\|\boldsymbol{t}\|)\exp(-\|\boldsymbol{t}\|^2)\,\mathrm{d}\boldsymbol{t}$$

$$\boldsymbol{x} = \big(\Phi(at_1), \ldots, \Phi(at_d)\big),$$ Φ is the CDF of the standard normal

$$\mathrm{d}\boldsymbol{x} = a^d\phi(at_1)\cdots\phi(at_d)\,\mathrm{d}\boldsymbol{t} = \frac{a^d\exp(-a^2\|\boldsymbol{t}\|^2/2)}{(2\pi)^{d/2}}\,\mathrm{d}\boldsymbol{t}$$

$$\boldsymbol{t} = \big(\Phi^{-1}(x_1), \ldots, \Phi^{-1}(x_d)\big)/a$$

$$\mu = \int_{[0,1]^d} \frac{(2\pi)^{d/2}}{a^d} \cos\left(\left\|\big(\Phi^{-1}(x_1), \ldots, \Phi^{-1}(x_d)\big)\right\|/a\right)$$

$$\times \exp\left(-\frac{2-a^2}{2a^2}\left\|\big(\Phi^{-1}(x_1), \ldots, \Phi^{-1}(x_d)\big)\right\|^2\right)\mathrm{d}\boldsymbol{x}, \qquad (a^2 \le 2)$$

Whole family $f(\boldsymbol{x}; a)$ for which $\mu = \mathbb{E}[f(\boldsymbol{X}; a)]$ for $\boldsymbol{X} \sim \mathcal{U}[0,1]^d$

Variable transformations

Suppose that you want to approximate

$$\mu = \int_{\mathcal{T}} g(\boldsymbol{t})\,\mathrm{d}\boldsymbol{t}$$

by Monte Carlo. You need to write it as an expectation. Let $\boldsymbol{t} = \boldsymbol{\psi}(\boldsymbol{x})$, for some invertible transformation, $\boldsymbol{\psi} : \mathcal{X} \to \mathcal{T}$. Define the Jacobian determinant as

$$J(\boldsymbol{x};\boldsymbol{\psi}) := \det\left(\left(\frac{\partial \psi_i}{\partial x_j}\right)_{i,j=1}^{d}\right), \quad J(\boldsymbol{t};\boldsymbol{\psi}^{-1}) := \det\left(\left(\frac{\partial \psi_i^{-1}}{\partial t_j}\right)_{i,j=1}^{d}\right), \; J(\boldsymbol{\psi}(\boldsymbol{x});\boldsymbol{\psi}^{-1}) = \frac{1}{J(\boldsymbol{x};\boldsymbol{\psi})}$$

Then

$$\mu = \int_{\mathcal{X}} g(\boldsymbol{\psi}(\boldsymbol{x}))\,\left|J(\boldsymbol{x};\boldsymbol{\psi})\right|\,\mathrm{d}\boldsymbol{x} = \int_{\mathcal{X}} \underbrace{\frac{g(\boldsymbol{\psi}(\boldsymbol{x}))}{\varrho_{\boldsymbol{X}}(\boldsymbol{x})}\,\left|J(\boldsymbol{x};\boldsymbol{\psi})\right|}_{f(\boldsymbol{x})}\,\varrho_{\boldsymbol{X}}(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} = \mathbb{E}_{\boldsymbol{X}\sim\varrho_{\boldsymbol{X}}}[f(\boldsymbol{X})]$$

Variable transformations

$$\mu = \int_{\mathcal{T}} g(\boldsymbol{t})\,\mathrm{d}\boldsymbol{t} = \,?$$

Let $\boldsymbol{t} = \boldsymbol{\psi}(\boldsymbol{x})$, for some invertible transformation, $\boldsymbol{\psi} : \mathcal{X} \to \mathcal{T}$. Then

$$\mu = \int_{\mathcal{X}} g(\boldsymbol{\psi}(\boldsymbol{x}))\,\left|J(\boldsymbol{x};\boldsymbol{\psi})\right|\,\mathrm{d}\boldsymbol{x} = \int_{\mathcal{X}} \underbrace{\frac{g(\boldsymbol{\psi}(\boldsymbol{x}))}{\varrho(\boldsymbol{x})}\,\left|J(\boldsymbol{x};\boldsymbol{\psi})\right|}_{f(\boldsymbol{x})}\;\varrho(\boldsymbol{x})\;\mathrm{d}\boldsymbol{x} = \mathbb{E}_{\boldsymbol{X}\sim\varrho}[f(\boldsymbol{X})]$$

If $\boldsymbol{X} \sim \mathcal{U}[0,1]^d$ (because we know how generate uniform samples), then

$$\mu = \int_{[0,1]^d} \underbrace{g(\boldsymbol{\psi}(\boldsymbol{x}))\,\left|J(\boldsymbol{x};\boldsymbol{\psi})\right|}_{f(\boldsymbol{x})}\;\mathrm{d}\boldsymbol{x} = \mathbb{E}_{\boldsymbol{X}\sim\mathcal{U}[0,1]^d}[f(\boldsymbol{X})]$$

Random variable transformations

Let

- $\boldsymbol{T}$ be a random variable with sample space $\mathscr{T}$ and density $\varrho_{\boldsymbol{T}}$
- $\boldsymbol{X}$ be a random variable with sample space $\mathscr{X}$ and density $\varrho_{\boldsymbol{X}}$
- $\boldsymbol{T} = \boldsymbol{\psi}(\boldsymbol{X})$, for some invertible transformation, $\boldsymbol{\psi} : \mathscr{X} \to \mathscr{T}$. Then

$$\varrho_{\boldsymbol{T}}(\boldsymbol{t}) = \varrho_{\boldsymbol{X}}(\boldsymbol{\psi}^{-1}(\boldsymbol{t}))|J(\boldsymbol{t};\boldsymbol{\psi}^{-1})|, \qquad \varrho_{\boldsymbol{T}}(\boldsymbol{\psi}(\boldsymbol{x})) = \varrho_{\boldsymbol{X}}(\boldsymbol{x})|J(\boldsymbol{\psi}(\boldsymbol{x}));\boldsymbol{\psi}^{-1})| = \frac{\varrho_{\boldsymbol{X}}(\boldsymbol{x})}{|J(\boldsymbol{x};\boldsymbol{\psi})|}$$

$$\mu = \int_{\mathscr{T}} \underbrace{h(\boldsymbol{t})\,\varrho_{\boldsymbol{T}}(\boldsymbol{t})}_{g(\boldsymbol{t})}\,\mathrm{d}\boldsymbol{t} = \mathbb{E}_{\boldsymbol{T}\sim\varrho_{\boldsymbol{T}}}[h(\boldsymbol{T})] = \mathbb{E}_{\boldsymbol{X}\sim\varrho_{\boldsymbol{X}}}[h(\boldsymbol{\psi}(\boldsymbol{X}))] \qquad = \int_{\mathscr{X}} h(\boldsymbol{\psi}(\boldsymbol{x}))\,\varrho_{\boldsymbol{X}}(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}$$

$$\text{or } = \int_{\mathscr{X}} \frac{h(\boldsymbol{\psi}(\boldsymbol{x}))\,\varrho_{\boldsymbol{T}}(\boldsymbol{\psi}(\boldsymbol{x}))}{\varrho_{\boldsymbol{X}}(\boldsymbol{x})}\,|J(\boldsymbol{x};\boldsymbol{\psi})|\,\varrho_{\boldsymbol{X}}(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} \qquad = \int_{\mathscr{X}} h(\boldsymbol{\psi}(\boldsymbol{x}))\,\varrho_{\boldsymbol{X}}(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}$$

Importance sampling

Suppose that you want to approximate

$$\mu = \mathbb{E}_{\boldsymbol{T} \sim \varrho_{\boldsymbol{T}}}[h(\boldsymbol{T})] = \int_{\mathcal{T}} h(\boldsymbol{t})\, \varrho_{\boldsymbol{T}}(\boldsymbol{t})\,\mathrm{d}\boldsymbol{t} = \int_{\mathcal{T}} \underbrace{h(\boldsymbol{t}) \overbrace{\frac{\varrho_{\boldsymbol{T}}(\boldsymbol{t})}{\varrho_{\boldsymbol{X}}(\boldsymbol{t})}}^{\text{likelihood ratio}}}_{f(\boldsymbol{t})} \varrho_{\boldsymbol{X}}(\boldsymbol{t})\,\mathrm{d}\boldsymbol{t} = \mathbb{E}_{\boldsymbol{X} \sim \varrho_{\boldsymbol{X}}}[f(\boldsymbol{X})]$$

We try to choose $\varrho_{\boldsymbol{X}}$ to make $\mathrm{var}[f(\boldsymbol{X})]$ small when doing IID sampling. If $\boldsymbol{T} = \boldsymbol{\psi}(\boldsymbol{X})$ for $\boldsymbol{X} \sim \mathcal{U}[0,1]^d$—since often find generate non-uniform samples from uniform ones—then

$$\mu = \int_{\mathcal{T}} h(\boldsymbol{t})\, \varrho_{\boldsymbol{T}}(\boldsymbol{t})\,\mathrm{d}\boldsymbol{t} = \mathbb{E}_{\boldsymbol{T} \sim \varrho_{\boldsymbol{T}}}[h(\boldsymbol{T})] = \mathbb{E}_{\boldsymbol{X} \sim \mathcal{U}[0,1]^d}[h(\boldsymbol{\psi}(\boldsymbol{X}))] = \int_{[0,1]^d} h(\boldsymbol{\psi}(\boldsymbol{x}))\,\mathrm{d}\boldsymbol{x}$$

Ex. Brownian motion with drift

In computational finance certain outcomes, such as positive option payoffs, may be rare

- Adding a drift to the Brownian motion increases the probability that asset paths yield positive payoffs and
- Reduces the variance of our estimator of the option price (population mean)
- Paths with positive payoffs receive smaller sample weights to compensate and keep the mean the same

$$\text{price} = \int_{\mathbb{R}^d} \text{payoff}(\boldsymbol{x})\, \varrho(\boldsymbol{x})\, \mathrm{d}\boldsymbol{x}, \quad \varrho(\boldsymbol{x}) = \frac{\exp(-\boldsymbol{x}^\top \Sigma^{-1} \boldsymbol{x}/2)}{\sqrt{(2\pi)^d |\Sigma|}}, \ \ \Sigma = \big(\min(t_i, t_j)\big)_{i,j=1}^d$$

Here the density ϱ is a discretized Brownian motion

Ex. Brownian motion with drift cont'd

$$\text{price} = \int_{\mathbb{R}^d} \text{payoff}(\boldsymbol{x})\, \varrho(\boldsymbol{x})\, \mathrm{d}\boldsymbol{x}$$

$$\varrho(\boldsymbol{x}) = \frac{\exp(-\boldsymbol{x}^\top \Sigma^{-1} \boldsymbol{x}/2)}{\sqrt{(2\pi)^d |\Sigma|}}, \quad \Sigma = \bigl(\min(t_i, t_j)\bigr)_{i,j=1}^d$$

We importance sample:

$$\varrho_{\text{drift}}(\boldsymbol{x}) = \frac{\exp\bigl(-(\boldsymbol{x}-\boldsymbol{a})^\top \Sigma^{-1} (\boldsymbol{x}-\boldsymbol{a})/2\bigr)}{\sqrt{(2\pi)^d |\Sigma|}}, \qquad \boldsymbol{a} = \alpha(t_1, \ldots, t_d)^\top$$

$$\text{likelihood ratio} = \frac{\varrho(\boldsymbol{x})}{\varrho_{\text{drift}}(\boldsymbol{x})} = \exp(-\alpha x_d + \alpha^2 t_d/2)$$

$$\text{price} = \int_{\mathbb{R}^d} \text{payoff}(\boldsymbol{x})\, \exp(-\alpha x_d + \alpha^2 t_d/2)\, \varrho_{\text{drift}}(\boldsymbol{x})\, \mathrm{d}\boldsymbol{x}$$

Note that the weight gets smaller as αx_d increases or $\alpha^2 t_d$ decreases

Variable transformation / importance sampling observations

- Variable transformation and importance sampling are intertwined
- The choice of sampling density or variable transformation affects the variance of the sample mean
- You must choose the variable transformation so that
  - The Jacobian determinant, $J(\boldsymbol{x}; \boldsymbol{\psi})$, is finite
  - The probability density, $\varrho_{\boldsymbol{X}}(\boldsymbol{x})$, does not vanish unless integrand vanishes
- You might try pilot samples to help choose transformation

Control variates

$$\mu = \mathbb{E}[f(\boldsymbol{X})] \iff \mu = \mathbb{E}[f(\boldsymbol{X}) - \beta_1\eta_1(\boldsymbol{X}) - \cdots - \beta_m\eta_m(\boldsymbol{X})]$$

provided that $\mathbb{E}[\eta_1(\boldsymbol{X})] = \cdots \mathbb{E}[\eta_m(\boldsymbol{X})] = 0$. Want to arrange such that

$$\mathrm{var}[f(\boldsymbol{X}) - \beta_1\eta_1(\boldsymbol{X}) - \cdots - \beta_m\eta_m(\boldsymbol{X})] \le \mathrm{var}[f(\boldsymbol{X})]$$

Then

$\hat{\mu}_{\mathrm{CV},n} = \dfrac{1}{n}\displaystyle\sum_{i=0}^{n-1}[f(\boldsymbol{X}_i) - \beta_1\eta_1(\boldsymbol{X}_i) - \cdots - \beta_m\eta_m(\boldsymbol{X}_i)]$ converges faster than $\hat{\mu}_n = \dfrac{1}{n}\displaystyle\sum_{i=0}^{n-1} f(\boldsymbol{X}_i)$

How do we minimize $\mathrm{var}[f(\boldsymbol{X}) - \beta_1\eta_1(\boldsymbol{X}) - \cdots - \beta_m\eta_m(\boldsymbol{X})]$?

Control variates (cont'd)

$$\hat{\mu}_{\mathrm{CV},n} = \frac{1}{n}\sum_{i=0}^{n-1}[f(\boldsymbol{X}_i) - \beta_1\eta_1(\boldsymbol{X}_i) - \cdots - \beta_m\eta_m(\boldsymbol{X}_i)]$$

How do we minimize $\mathrm{var}[f(\boldsymbol{X}) - \beta_1\eta_1(\boldsymbol{X}) - \cdots - \beta_m\eta_m(\boldsymbol{X})]$? Use least squares regression with centered response and explanatory variables to choose $\hat{\boldsymbol{\beta}}$:

$$\hat{\boldsymbol{\beta}} = \underset{\boldsymbol{b}}{\mathrm{argmin}}\|\boldsymbol{y} - \mathsf{H}\boldsymbol{b}\|^2$$

where $\boldsymbol{y} = \begin{pmatrix} f(\boldsymbol{x}_0) - \widehat{\mu}_n \\ \vdots \\ f(\boldsymbol{x}_{n-1}) - \widehat{\mu}_n \end{pmatrix}, \quad \mathsf{H} = \begin{pmatrix} \eta_1(\boldsymbol{x}_0) - \widehat{\mu}_{\eta_1,n} & \cdots & \eta_m(\boldsymbol{x}_0) - \widehat{\mu}_{\eta_m,n} \\ \vdots & & \vdots \\ \eta_1(\boldsymbol{x}_{n-1}) - \widehat{\mu}_{\eta_1,n} & \cdots & \eta_m(\boldsymbol{x}_{n-1}) - \widehat{\mu}_{\eta_m,n} \end{pmatrix}$

$$\widehat{\mu}_n = \frac{1}{n}\sum_{i=0}^{n-1} f(\boldsymbol{x}_i), \qquad \widehat{\mu}_{\eta_1,n} = \frac{1}{n}\sum_{i=0}^{n-1}\eta_1(\boldsymbol{x}_i), \ \ldots, \ \widehat{\mu}_{\eta_m,n} = \frac{1}{n}\sum_{i=0}^{n-1}\eta_m(\boldsymbol{x}_i)$$

Least Squares

How do we solve

$$\hat{\beta} = \underset{\boldsymbol{b}}{\operatorname{argmin}} \left\| \boldsymbol{y} - \mathsf{H}\boldsymbol{b} \right\|^2 = \underset{\boldsymbol{b}}{\operatorname{argmin}} (\boldsymbol{y} - \mathsf{H}\boldsymbol{b})^T (\boldsymbol{y} - \mathsf{H}\boldsymbol{b})$$

Let

$$\boldsymbol{b} = (\mathsf{H}^\top \mathsf{H})^{-1} \mathsf{H}^\top \boldsymbol{y} + \boldsymbol{c}$$

Then

$$\begin{aligned}
\left\| \boldsymbol{y} - \mathsf{H}\boldsymbol{b} \right\|^2 &= \left\| \boldsymbol{y} - \mathsf{H}[(\mathsf{H}^\top \mathsf{H})^{-1} \mathsf{H}^\top \boldsymbol{y} + \boldsymbol{c}] \right\|^2 \\
&= \left\| [\mathsf{I} - \mathsf{H}(\mathsf{H}^\top \mathsf{H})^{-1} \mathsf{H}^\top] \boldsymbol{y} + \mathsf{H}\boldsymbol{c} \right\|^2 \\
&= \left([\mathsf{I} - \mathsf{H}(\mathsf{H}^T \mathsf{H})^{-1} \mathsf{H}^\top] \boldsymbol{y} + \mathsf{H}\boldsymbol{c} \right)^\top \left([\mathsf{I} - \mathsf{H}(\mathsf{H}^\top \mathsf{H})^{-1} \mathsf{H}^\top] \boldsymbol{y} + \mathsf{H}\boldsymbol{c} \right) \\
&= \boldsymbol{y}^\top [\mathsf{I} - \mathsf{H}(\mathsf{H}^\top \mathsf{H})^{-1} \mathsf{H}^\top] \boldsymbol{y} + \boldsymbol{c}^\top \mathsf{H}^\top \mathsf{H}\boldsymbol{c}
\end{aligned}$$

So the optimum comes by setting $\boldsymbol{c} = \boldsymbol{0}$

Control variates

- Should be clever in choosing the η_j
  - Good choices can decrease the sample size required
  - Bad choices only hurt in terms of increased computation time, but not increased sample size
- The β_j should be chosen by least squares for IID sampling
- For low discrepancy sequences, the β_j should be chosen to reduce the variation (not variance)