TAM 598 Lecture 10 :

QUANTIFYING UNCERTAINTY IN MONTE CARLO ESTIMATES

Announcements:

- HW 2 covers lectures 4-8 ; due on Feb 26

I. VISUALIZING MONTE CARLO UNCERTAINTY

expectation we want:

$$I = \mathbb{E}[g(x)] = \int g(x)\, p(x)\, dx$$

$I = \bar{I}_N + \Diamond$

where $X \sim p(x)$

Monte Carlo Approach:

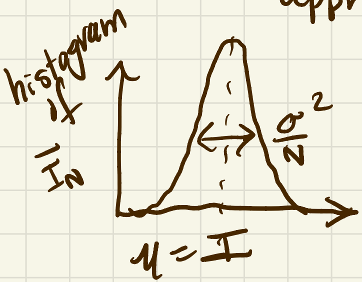
(1) obtain $x_1, x_2, \ldots, x_N$ iid samples of X

(2) $\bar{I}_N = \dfrac{g(x_1) + g(x_2) + \ldots g(x_N)}{N} \rightarrow I$ as $N \to \infty$

* Using small N leads to different estimates.
This is a type of epistemic uncertainty.

II. THINKING ABOUT EPISTEMIC UNCERTAINTY USING THE CENTRAL LIMIT THEOREM

CLT : sum of a large # of independent RVs is approximately normal distributed.



say $X_1, X_2, \ldots$ iid random variables with mean μ and variance σ^2

$$S_N = \frac{X_1 + X_2 + \ldots X_N}{N}$$

$$\sim N\left(\mu = I, \frac{\sigma^2}{N}\right)$$

where σ^2 is the variance of $\mathbb{V}[g(x)]$

III. QUANTIFYING EPISTEMIC UNCERTAINTY IN MONTE CARLO ESTIMATES

what can we say about I, given our estimate $\bar{I}_N$?

we have:

$$I = \mathbb{E}[g(x)] = \int g(x)\,p(x)\,dx$$

$$\boxed{\bar{I}_N = \frac{g(x_1) + \ldots g(x_N)}{N}} \longrightarrow I \quad \text{as } N \to \infty$$

Note that the RV's $Y_i = g(x_i)$ are iid with expectation

$$\mathbb{E}[Y_i] = \mathbb{E}[g(x_i)] = I$$

and say we have finite variance

$$\mathbb{V}[Y_i] = \sigma^2 < \infty$$

this means that our Monte Carlo estimates are normal distributed

$$\boxed{\tilde{I}_N \sim N\left(I, \frac{\sigma^2}{N}\right)}$$

so: $\bar{I}_N = I \pm \frac{\sigma}{\sqrt{N}} Z$ where $Z \sim N(0,1)$

$$\Longrightarrow \quad I = \bar{I}_N \pm \frac{\sigma}{\sqrt{N}} Z \quad \Longrightarrow \quad \boxed{I \sim N\left(\bar{I}_N, \frac{\sigma^2}{N}\right)}$$

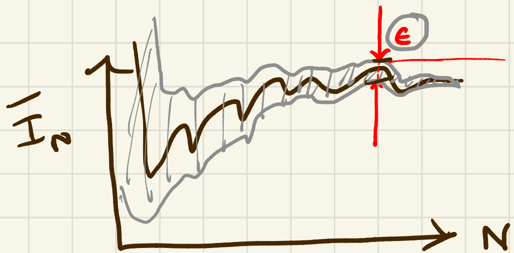
We can approximate σ^2 using our MC estimator

$$\tilde{\sigma}_N^2 = \frac{1}{N}\sum_{j=1}^{N} g^2(x_j) - \bar{I}_N^2$$

Then $I \sim N\left(\bar{I}_N, \frac{\tilde{\sigma}_N^2}{N}\right)$ and $\underbrace{I \approx \bar{I}_N \pm \frac{2}{\sqrt{N}}\tilde{\sigma}_N}$

95% predictive interval

For example, say we want to reduce the 95% confidence interval to a range $\underline{\epsilon}$. What is a good choice of N?



choose N so that

$$\epsilon \leq \frac{4\sigma}{\sqrt{N}}$$

$$N \geq 16\sigma^2 / \epsilon^2$$

and approximate σ^2 by $\bar{\sigma}_N^2$

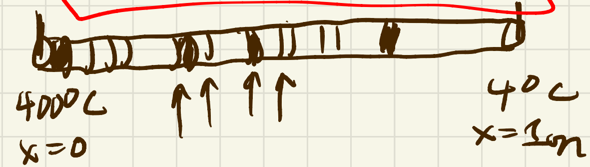
IV. Uncertainty Propagation Through a Boundary Value Problem

steady-state heat equation, heterogeneous 1D rod, no sources

$$\frac{d}{dx}\left(c(x) \frac{d}{dx} T(x) \right) = 0 \qquad x \in [0, 1] \text{ m}$$

$$\boxed{\begin{aligned} T(x=0) &= 400^\circ C \\ T(x=1m) &= 4^\circ C \end{aligned}}$$

we are uncertain about conductivity $c(x)$



What we know:

- rod is exactly 1m
- rod is from laminated segments of two materials
  - (0) fiberglass, $c_0 = 0.3$ W/mK — 30%
  - (1) steel $c_1 = 0.8$ W/mK — 70%
- rod is made of D total segments, $D \approx 100$

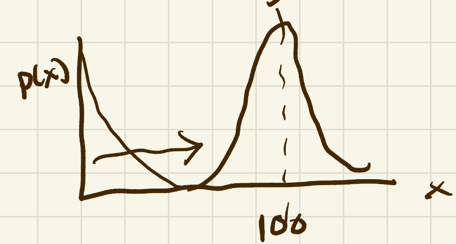
We are uncertain about $c(x)$ so we'll describe it as a random variable. We'll use what we know to inform the choice of R.V.

(1) # of segments D is a discrete RV, $\mathbb{E}[D] = 100$

$$D \sim \text{Poisson}(100)$$

$$p(x) = \frac{\lambda^x e^{-\lambda}}{x!}, \quad \mathbb{E}[x] = \lambda$$

Poisson (λ)



(2) coordinates of the segments? $x_0 = 0$, $x_D = 1$

assume the intermediate coords are obtained by uniformly sampling $[0,1]$ and sorting the numbers

$$(x_1, x_2, \ldots, x_{D-1}) = \text{sort}(u_1, \ldots, u_{D-1})$$

where $u_i \sim \mathcal{U}[0,1]$

(3) how do we define the material type of each segment?

segment d is made of material $M_d = \begin{cases} 0 & \text{fiberglass} \\ 1 & \text{steel} \end{cases}$

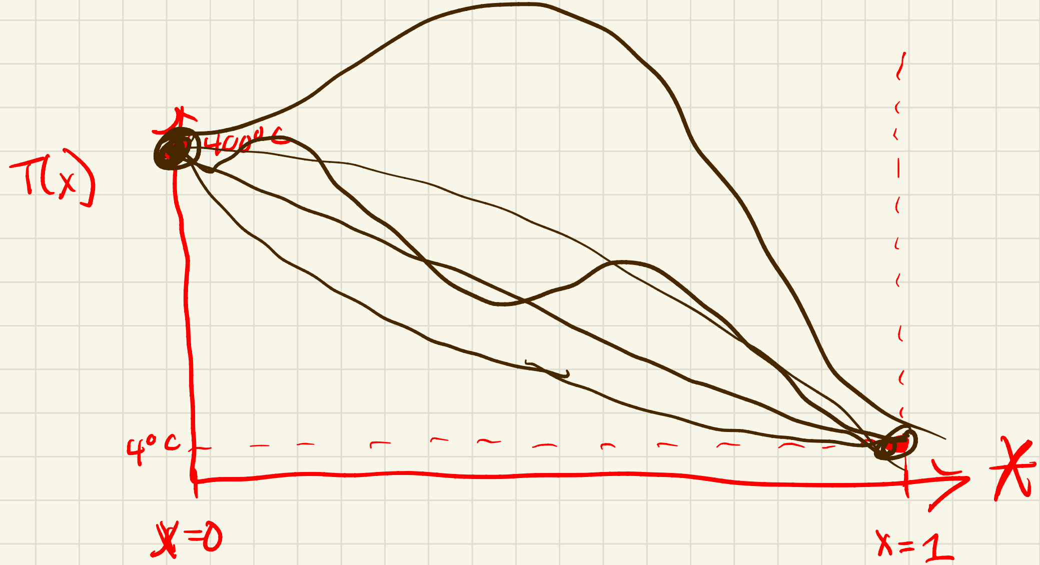
$$M_d \sim \text{categorical}\,(0.3,\ 0.7) \quad \text{for} \quad d = 1, \dots, D$$

Putting all of these together, we have our model for thermal conductivity

$$c(x) = \sum_{d=1}^{D} c_{M_d} \mathbb{1}_{[x_{d-1},\, x_d]}(x)$$

where if segment d is $\begin{cases} M_d = 0 & \text{then } c_0 = 0.3 \\ M_d = 1 & \text{then } c_1 = 0.8 \end{cases}$

$c(x)$ is a random function, depends on D, x_d, M_d


T(x)
400°C
4°C
x=0
x=1
x

Implementation in python: two classes

(1) Rod - a specific rod. Has the following methods:

Rod.get_conductivity(x)

Rod.__repr__()

Rod.plot()

(2) RandomRod - a random rod

RandomRod.rvs() samples a random rod

and Fipy for the 1D heat equation solver.