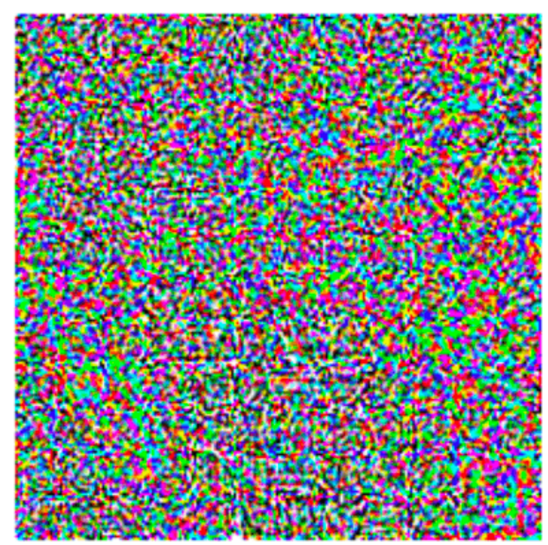
ENM 5310: Data-driven Modeling and Probabilistic Scientific Computing

Lecture #9: Sampling methods

Paris Perdikaris
February 14, 2023

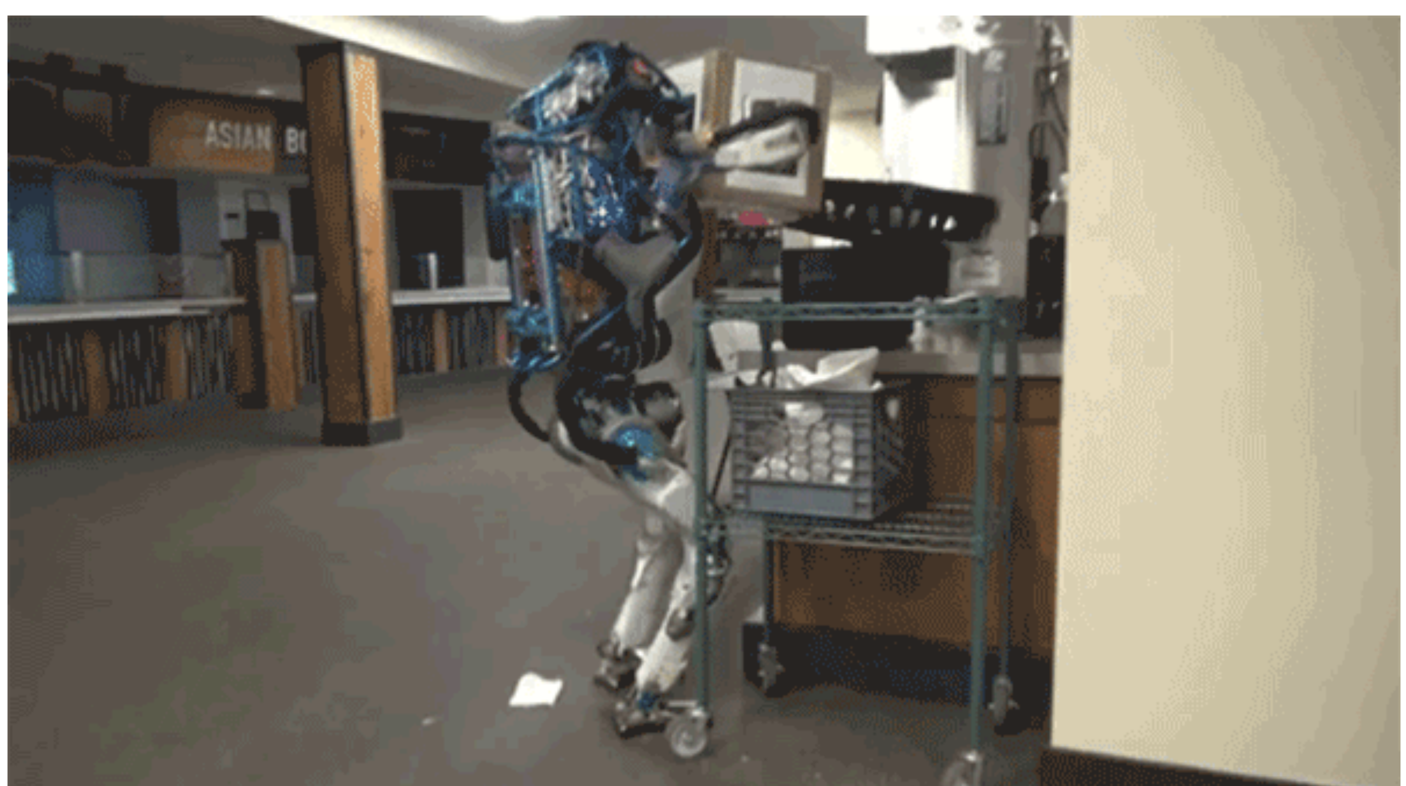
Penn
UNIVERSITY *of* PENNSYLVANIA

AI bloopers

 $+ .007 \times$  $=$ 

“panda”	“nematode”	“gibbon”
57.7% confidence	8.2% confidence	99.3 % confidence

AI bloopers



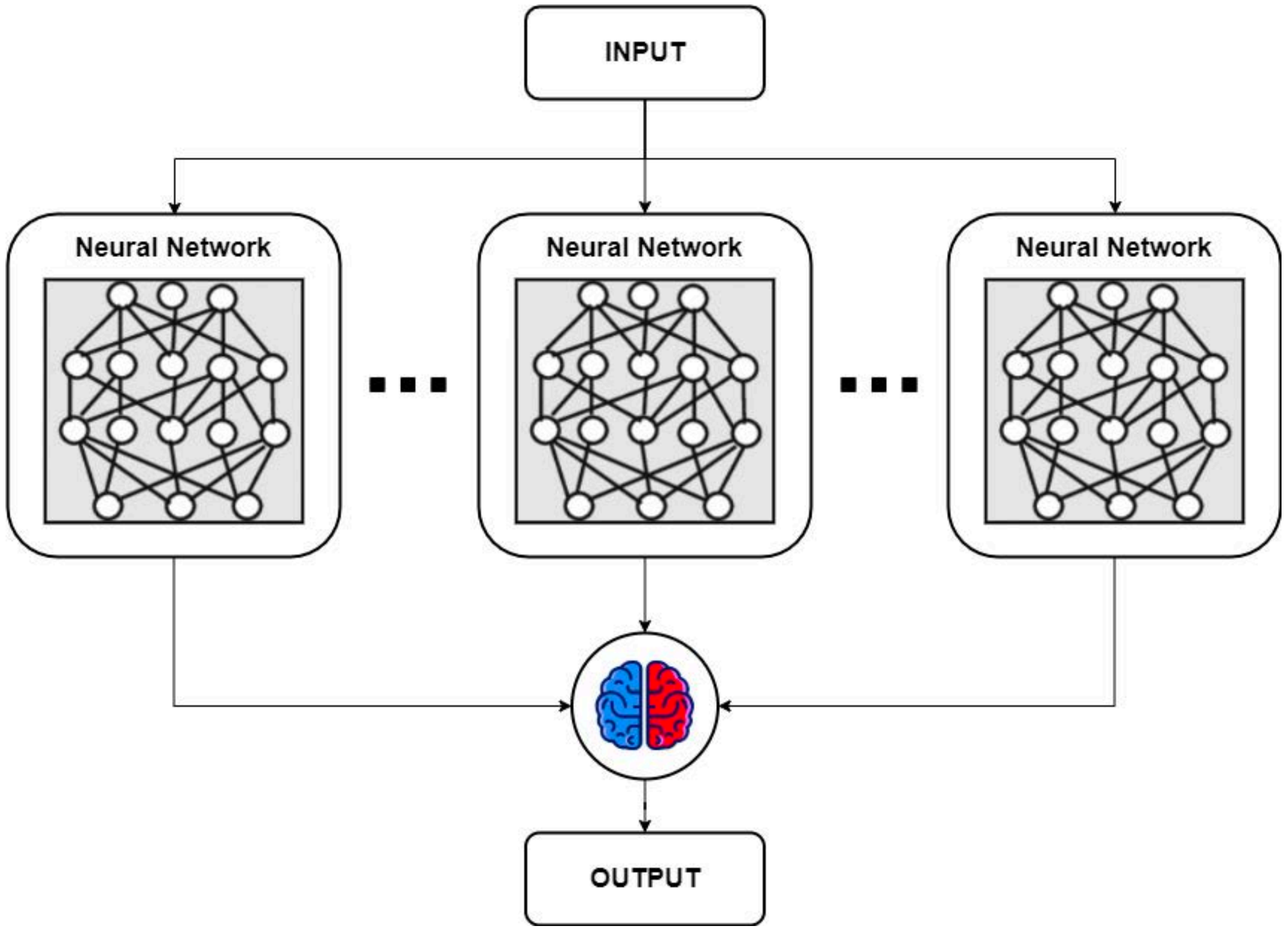
A need for robustness and uncertainty quantification

https://losslandscape.com/

A need for robustness and uncertainty quantification

Becomes particularly important when:

- We are working with small data-sets (over-fitting regime).
- We need to make high-consequence decisions.
- We require performance/accuracy guarantees.
- We work under a limited budget.



$$p(\theta|\mathcal{D}) = \frac{p(\mathcal{D}|\theta)p(\theta)}{p(\mathcal{D})} = \frac{p(\mathcal{D}|\theta)p(\theta)}{\int p(\mathcal{D}|\theta)p(\theta)d\theta}$$

The frequentist approach:
Ensemble averaging

The Bayesian approach:
Probabilistic programming

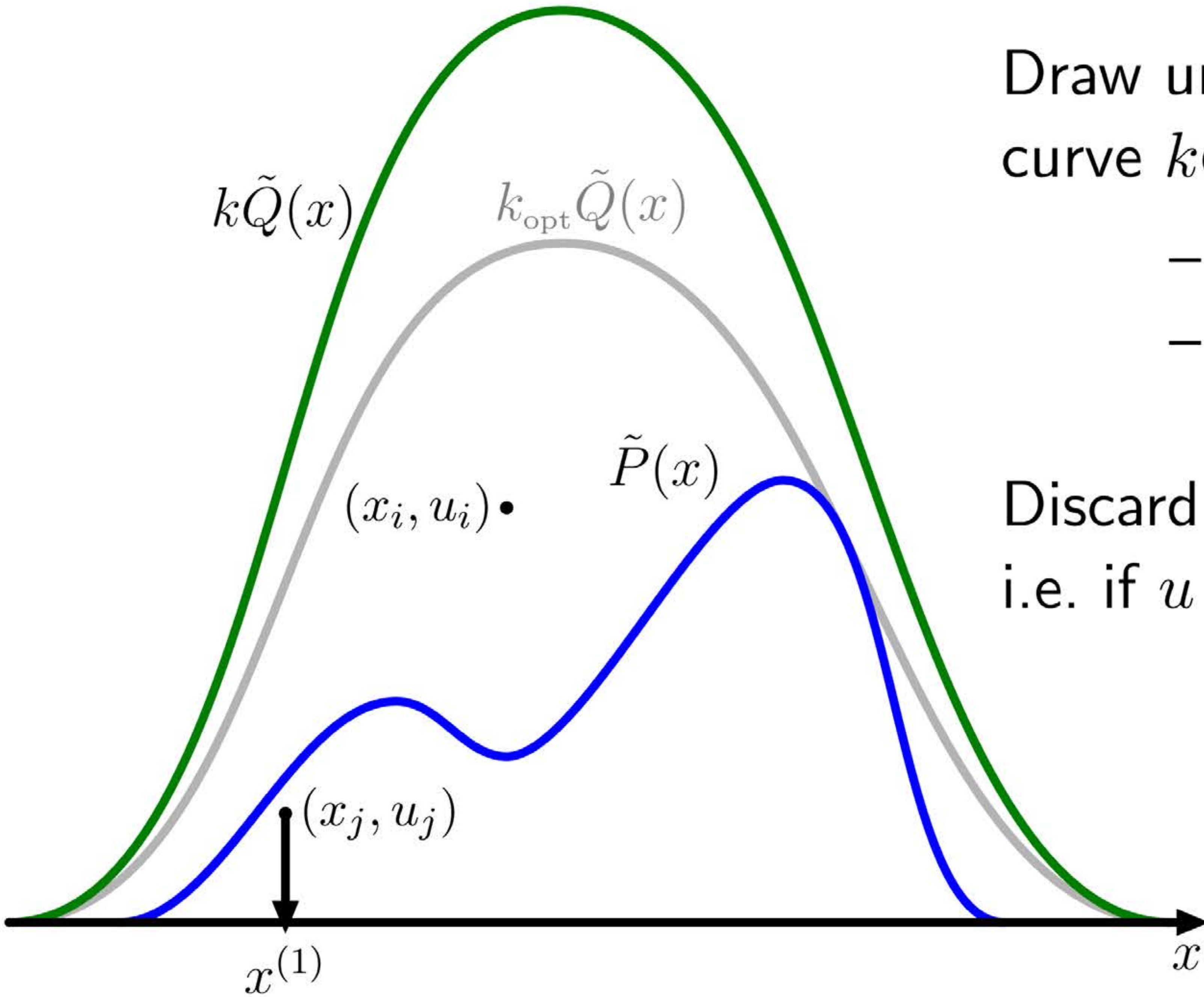
Monte Carlo approximation

$$\mathbb{E}_{x \sim p(x)}[f(x)] = \int f(x)p(x)dx \approx \frac{1}{n}\sum_{i=1}^{n} f(x_i),$$

where x_i are drawn iid from $p(x)$

Rejection sampling

Sampling underneath a $\tilde{P}(x) \propto P(x)$ curve is also valid



Draw underneath a simple curve $k\tilde{Q}(x) \geq \tilde{P}(x)$:

- Draw $x \sim Q(x)$
- height $u \sim \mathrm{Uniform}[0, k\tilde{Q}(x)]$

Discard the point if above $\tilde{P}$, i.e. if $u > \tilde{P}(x)$