

ENM 531: Data-driven Modeling and Probabilistic Scientific Computing

Lecture #11: Sampling methods

Paris Perdikaris
February 25, 2021



AI bloopers



“panda”
57.7% confidence

+ .007 ×



“nematode”
8.2% confidence

=

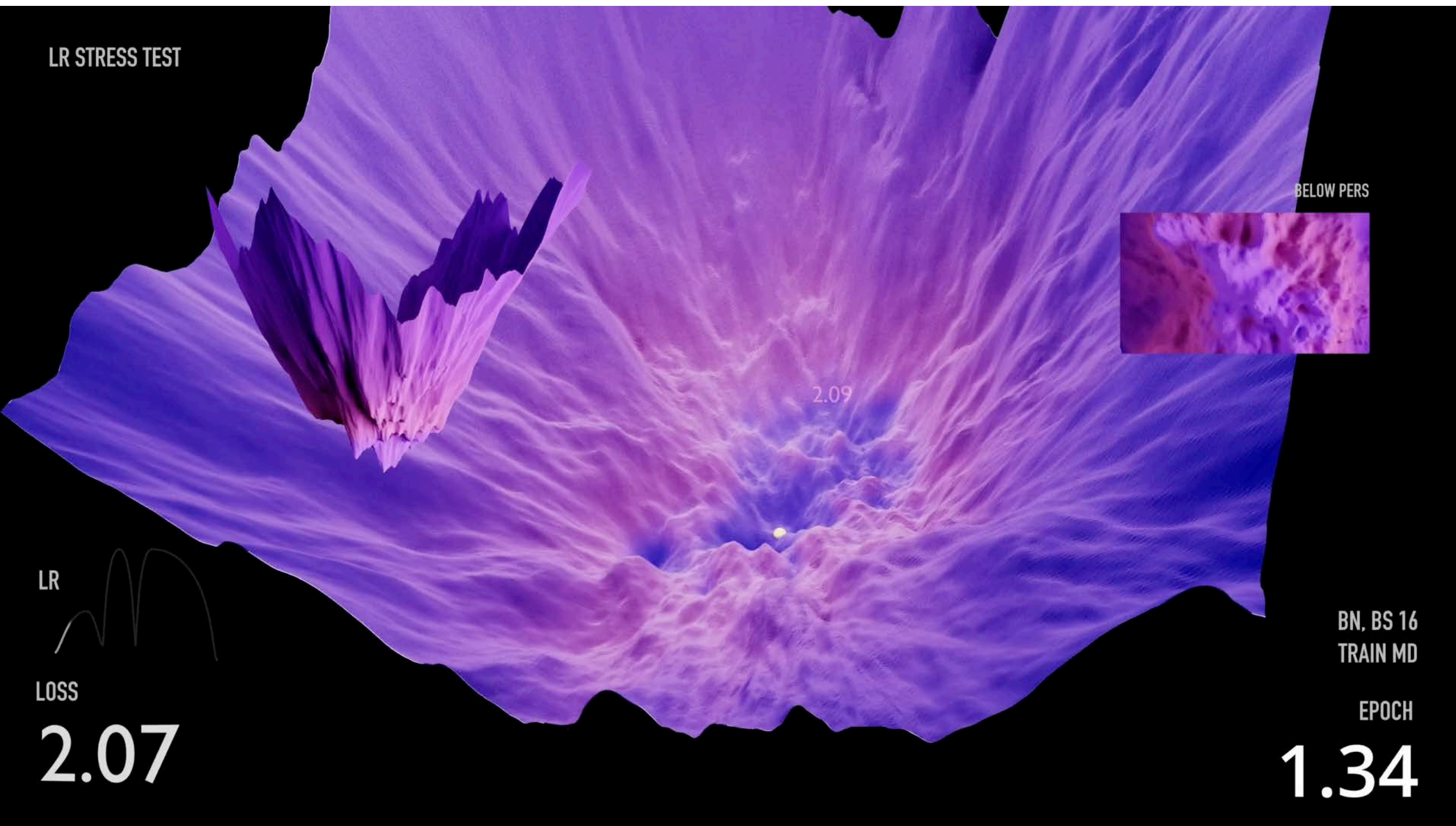


“gibbon”
99.3 % confidence

AI bloopers



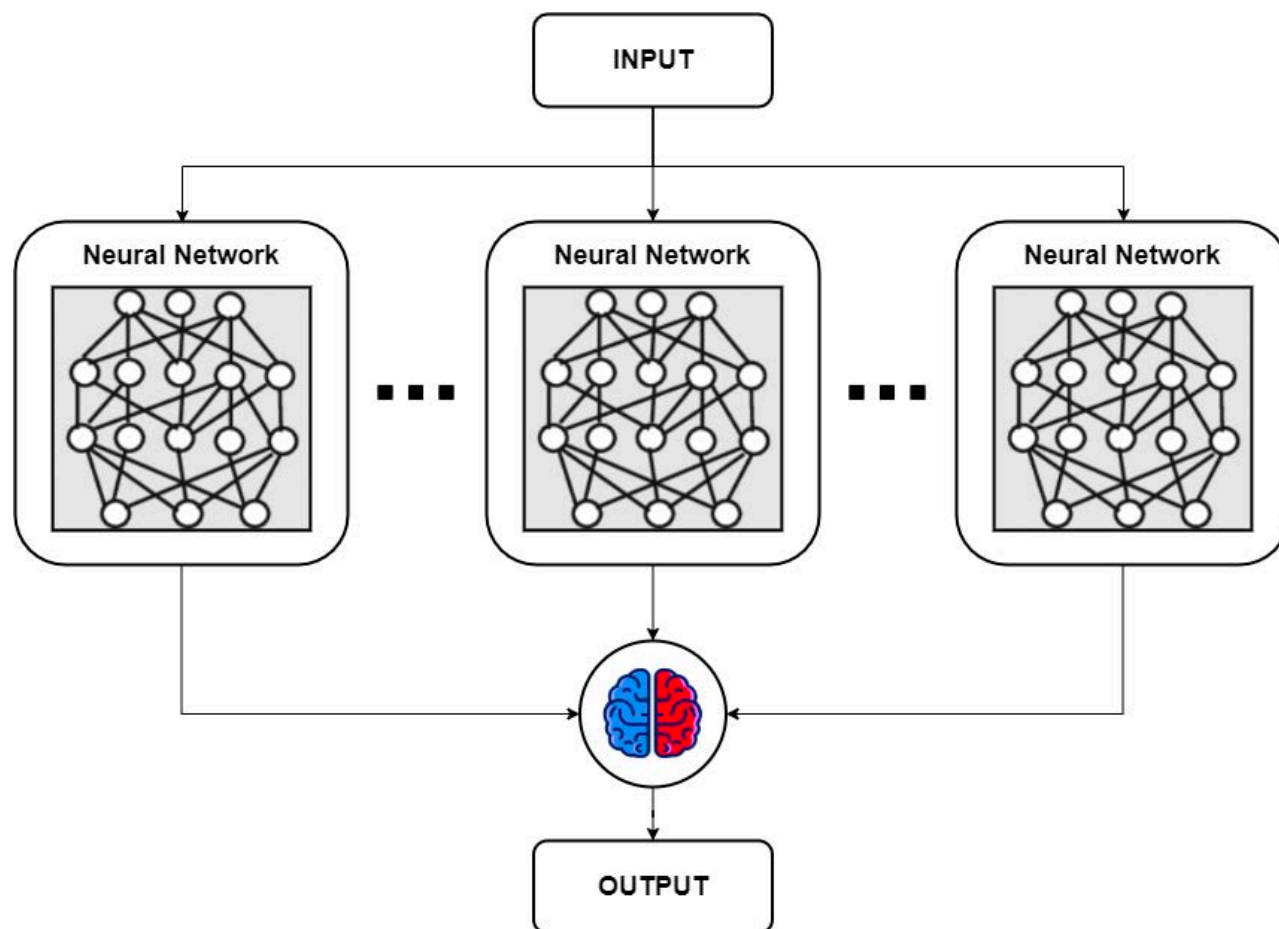
A need for robustness and uncertainty quantification



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.



The frequentist approach:
Ensemble averaging

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

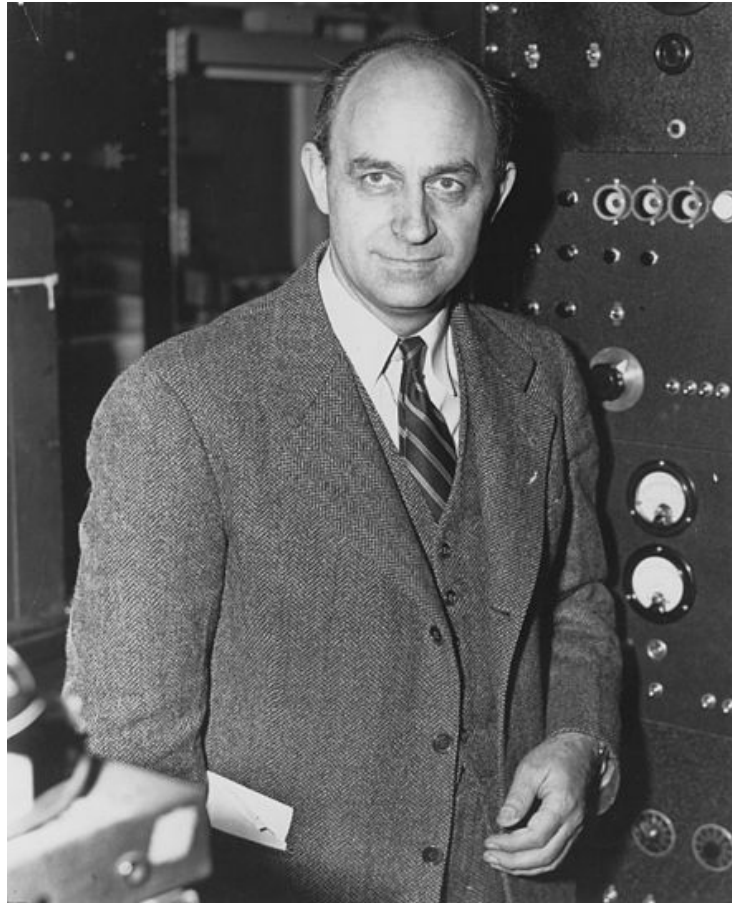
The Bayesian approach:
Probabilistic programming

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 $p(x)$

Monte Carlo approximation



Enrico Fermi



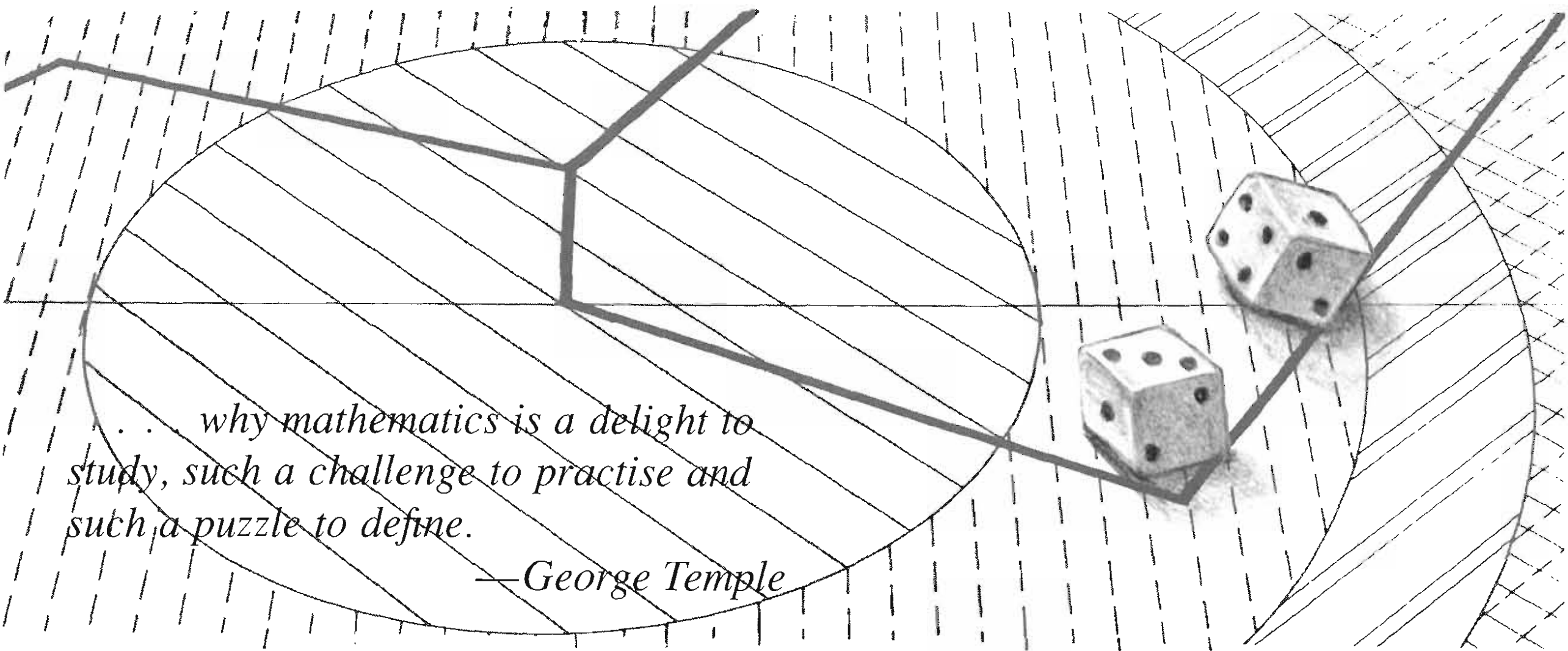
John von Neumann



Stan Ulam

THE BEGINNING *of the* MONTE CARLO METHOD

by N. Metropolis



*... why mathematics is a delight to
study, such a challenge to practise and
such a puzzle to define.*

—George Temple

Monte Carlo approximation



Monte Carlo approximation

Example: estimating π by Monte Carlo integration

MC approximation can be used for many applications, not just statistical ones. Suppose we want to estimate π . We know that the area of a circle with radius r is πr^2 , but it is also equal to the following definite integral:

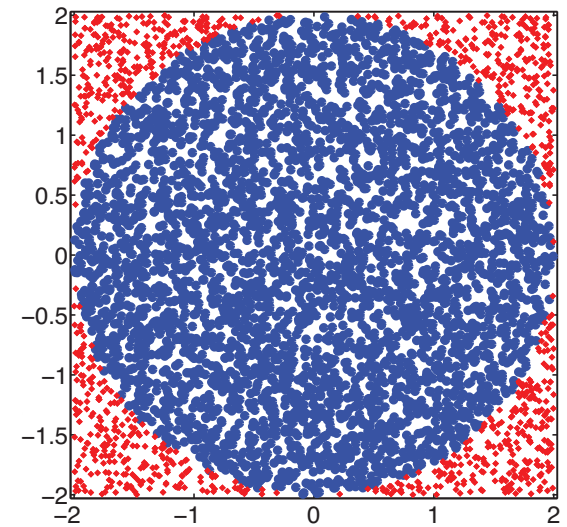
$$I = \int_{-r}^r \int_{-r}^r \mathbb{I}(x^2 + y^2 \leq r^2) dx dy \quad (2.99)$$

Hence $\pi = I/(r^2)$. Let us approximate this by Monte Carlo integration. Let $f(x, y) = \mathbb{I}(x^2 + y^2 \leq r^2)$ be an indicator function that is 1 for points inside the circle, and 0 outside, and let $p(x)$ and $p(y)$ be uniform distributions on $[-r, r]$, so $p(x) = p(y) = 1/(2r)$. Then

$$I = (2r)(2r) \int \int f(x, y) p(x) p(y) dx dy \quad (2.100)$$

$$= 4r^2 \int \int f(x, y) p(x) p(y) dx dy \quad (2.101)$$

$$\approx 4r^2 \frac{1}{S} \sum_{s=1}^S f(x_s, y_s) \quad (2.102)$$



We find $\hat{\pi} = 3.1416$ with standard error 0.09 (see Section 2.7.3 for a discussion of standard errors). We can plot the points that are accepted/ rejected as in Figure 2.19.