Lecture 9: Monte Carlo estimates of various statistics

Professor Ilias Bilionis

Estimating the cumulative distribution function

PREDICTIVE
SCIENCE LABORATORY

Estimating the variance

- Take a random variable $X \sim p(x)$ and some function $g(x)$.
- Consider the derived random variable $Y = g(X)$.
- We would like to estimate the cumulative distribution function of Y:

$$F(y) = p(Y \leq y) = p(g(X) \leq y)$$

PREDICTIVE
SCIENCE LABORATORY

Estimating the variance

- We would like to estimate the cumulative distribution function of Y:

$$F(y) = p(Y \leq y) = p(g(X) \leq y)$$

- Consider the indicator function of a set A:

$$1_A(y) = \begin{cases} 1, & y \text{ in } A \\ 0, & \text{otherwise} \end{cases}$$

- Using it, we can rewrite $F(y)$ as an expectation:

$$F(y) = p(Y \leq y) = p(g(X) \leq y) = \mathbb{E}\left[1_{(-\infty, y]}(g(X))\right]$$

PREDICTIVE
SCIENCE LABORATORY

Estimating the CDF

$$F(y) = \mathbb{E}\left[1_{(-\infty,y]}(g(X))\right]$$

- Take $X_1, X_2, \ldots$ independent identical copies of X.

- Estimate the CDF using a sample average:

$$\bar{F}_N(y) = \frac{1}{N}\sum_{i=1}^{N} 1_{[-\infty,y]}(g(X_i)) = \frac{\text{number of } g(X_i) \leq y}{N}$$

- This estimate is called the empirical CDF.

Estimating the CDF

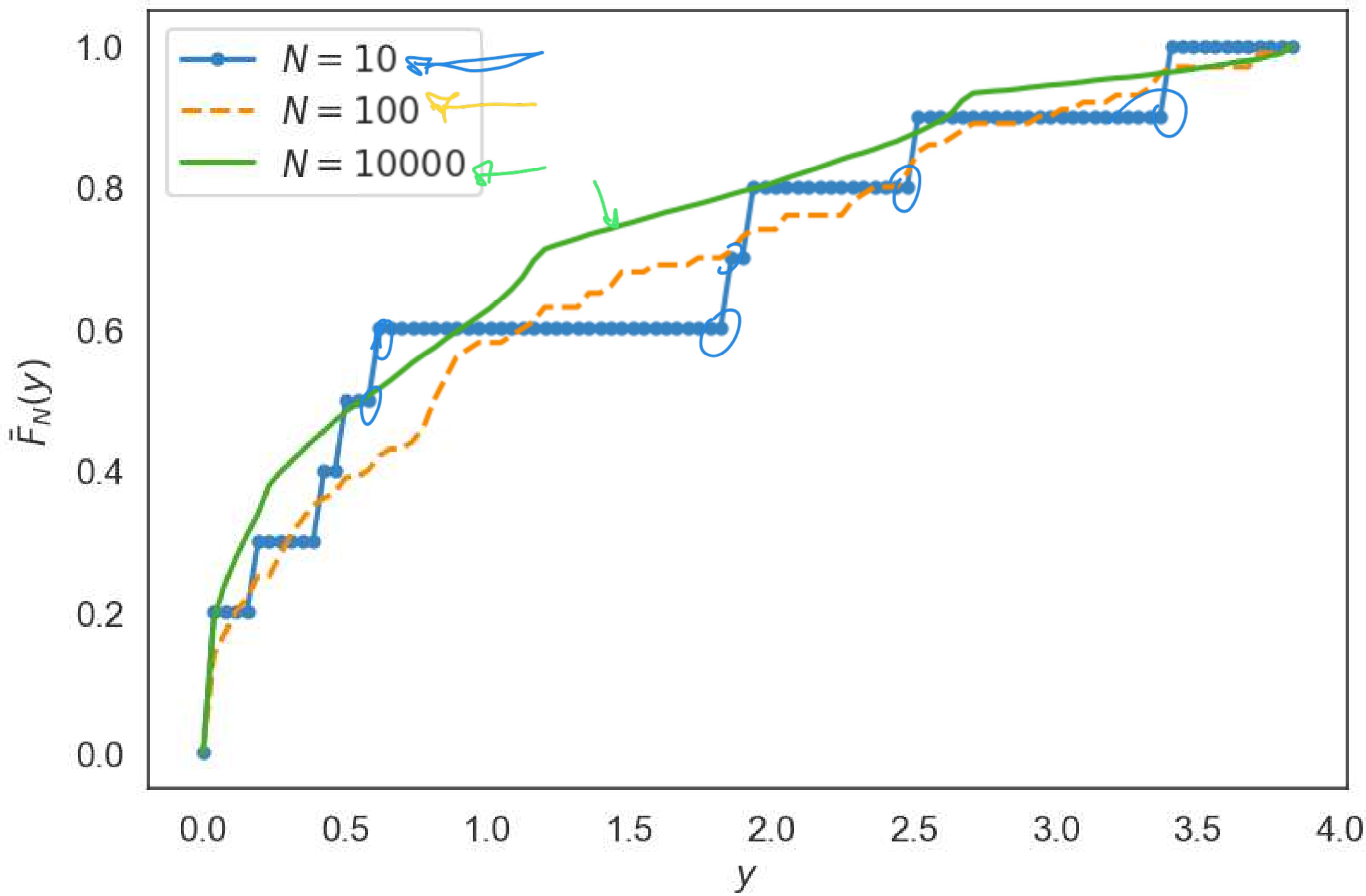
- Take $X_1, X_2, \ldots$ independent identical copies of X.

- Estimate the CDF using a sample average:

$$\bar{F}_N(y) = \frac{1}{N}\sum_{i=1}^{N} 1_{[-\infty, y]}(g(X_i)) = \frac{\text{number of } g(X_i) \le y}{N}$$

- This estimate is called the empirical CDF.

Example: 1D CDF



PREDIC
SCIENCE LABORATORY