

Lecture 9: Monte Carlo estimates of various statistics

Professor Ilias Bilonis

Estimating the cumulative distribution function

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 :

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

\ want to find this using
only samples of Y

Estimating the variance

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

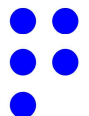
$$F(y) = p(Y \leq y) = p(g(X) \leq y)$$

- Consider the indicator function of a set A :

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

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

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



Estimating the CDF

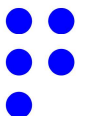
$$F(y) = \mathbb{E} \left[1_{(-\infty, y]}(g(X)) \right]$$

- Take X_1, X_2, \dots independent identical copies of X .
- Estimate the CDF using a sample average:

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

counts the number of times that $g(x) \leq y$

- This estimate is called the empirical CDF.



Example: 1D CDF

