

Lecture 9: Monte Carlo estimates of various statistics

Professor Ilias Bilonis

Estimating the probability density function via histograms

Estimating the probability density function via histograms

don't need to know $p(x)$ specifically, just need to be able to sample from it (either synthetically or experimentally)

- Let X be a random variable and $Y = g(X)$.
- We want to approximate the PDF $p(y)$ of Y .

Estimating the probability density function via histograms

- Take M small bins $[b_0, b_1]$, ..., $[b_{M-1}, b_M]$ in the y space.
- We will approximate $p(y)$ with a constant inside each bin:

$$\hat{p}_M(y) = \sum_{j=1}^M \hat{c}_j 1_{[b_{j-1}, b_j]}(y),$$

piecewise
approximation

- Where the c_j 's are coefficients to be determined.

Estimating the probability density function via histograms

- We will approximate $p(y)$ with a constant inside each bin:

$$\hat{p}_M(y) = \sum_{j=1}^M c_j 1_{[b_{j-1}, b_j]}(y),$$

- The constants are just:

$$c_j = p(b_{j-1} \leq Y \leq b_j) = F(b_j) - F(b_{j-1})$$

$\approx \bar{F}_N(b_j) - \bar{F}_N(b_{j-1})$ — know how to find $\bar{F}_N(y)$

Estimating the probability density function via histograms

- So, we can approximate the constants with the empirical CDF:

$$\bar{c}_{j,N} := \bar{F}_N(b_j) - \bar{F}_N(b_{j-1}) \rightarrow c_j \text{ a.s.}$$

- The intuitive meaning of this is:

$$\bar{c}_{j,N} = \frac{\text{number of samples that fall in bin } [b_{j-1}, b_j]}{N}$$

Estimating the probability density function via histograms

- The intuitive meaning of this is:

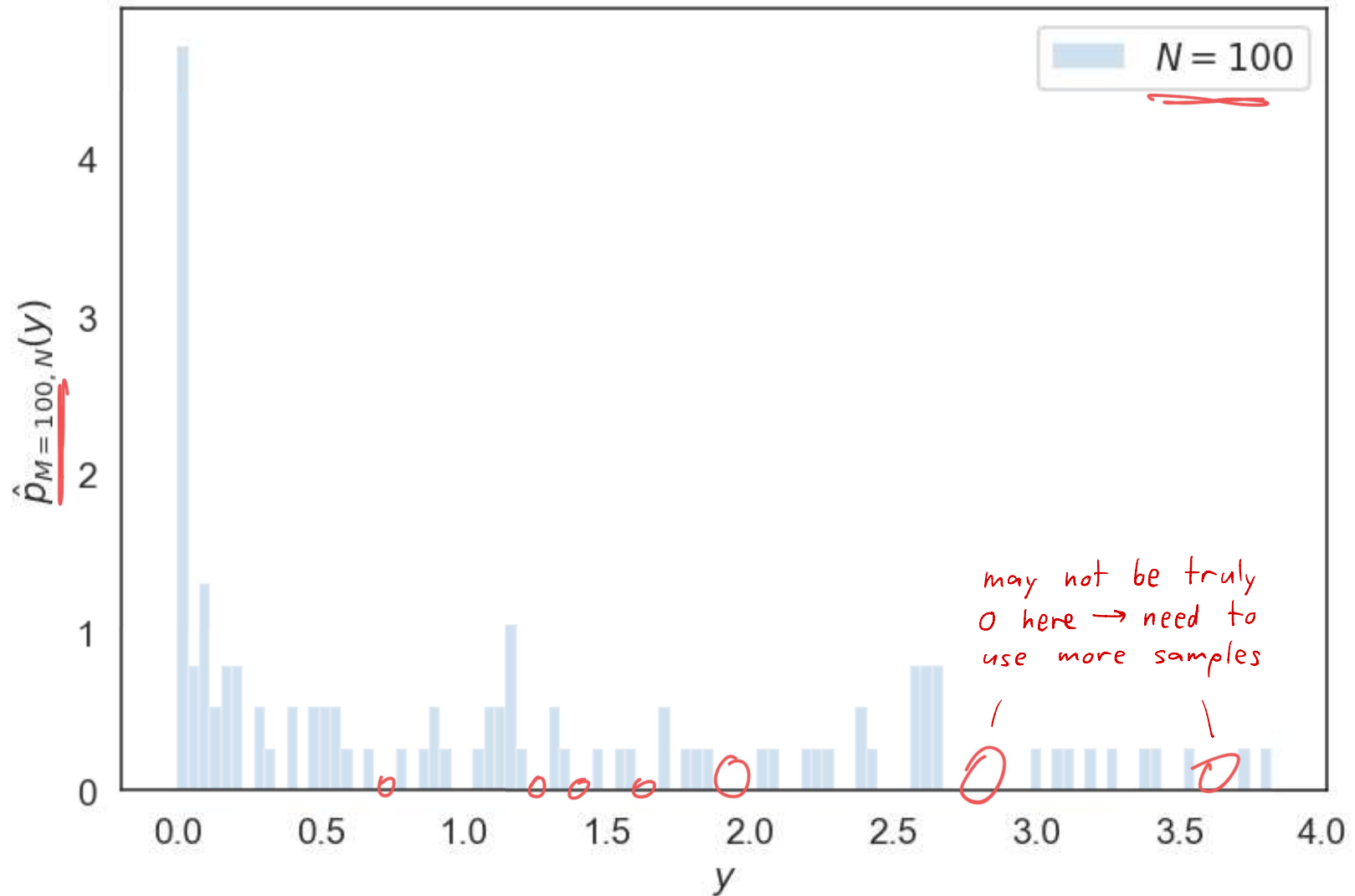
$$\bar{c}_{j,N} = \frac{\text{number of samples that fall in bin } [b_{j-1}, b_j]}{N}$$

- Putting everything together our approximation becomes:

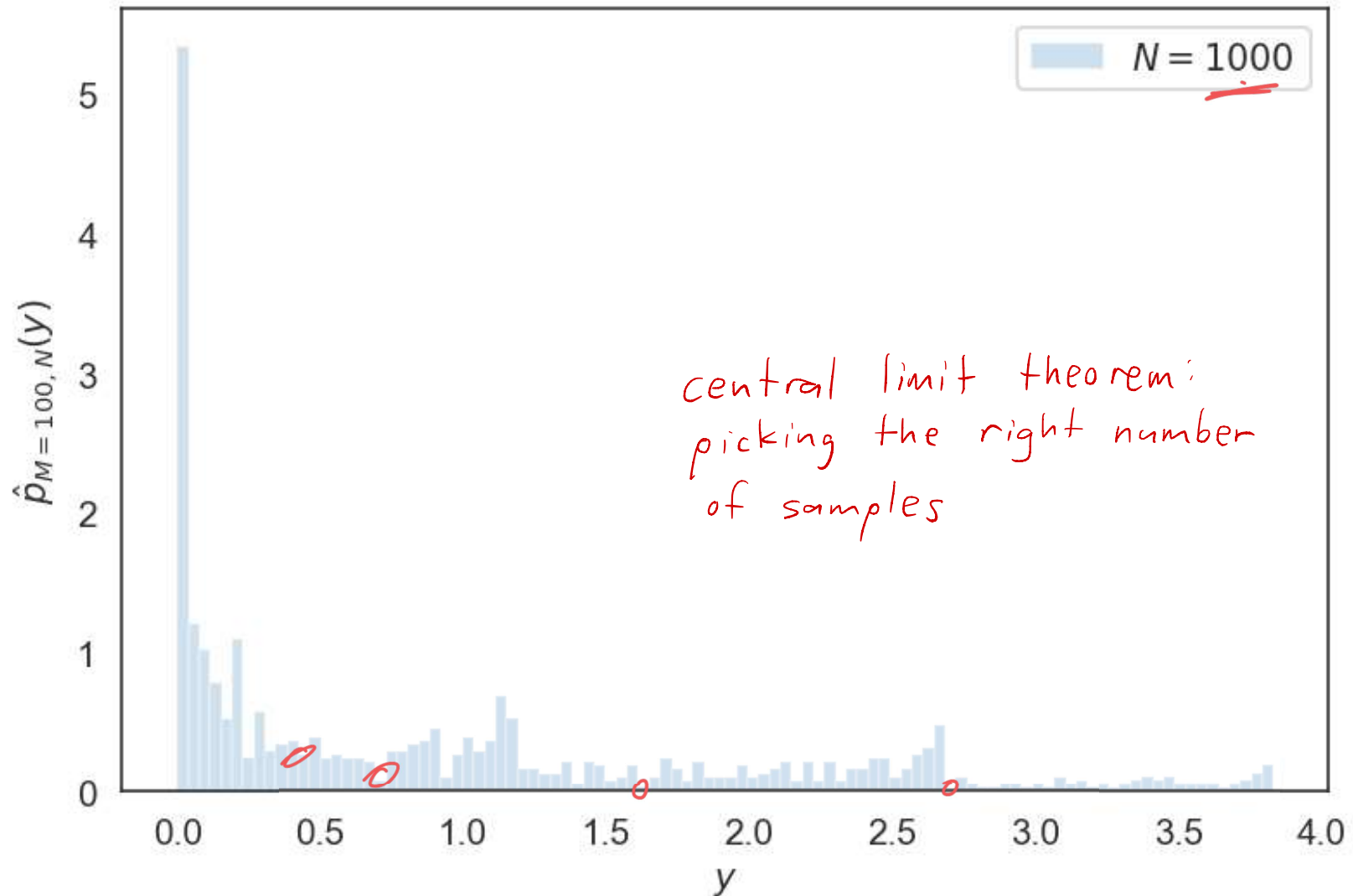
$$\hat{p}_{\bar{M},N}(y) = \sum_{j=1}^{\bar{M}} \bar{c}_{j,N} 1_{[b_{j-1}, b_j]}(y)$$

bins *# samples*

Example: 1D PDF



Example: 1D PDF



Example: 1D PDF

