Tool: Law of Large Numbers

$E[X]$

Terminology:

- *Sample*: A random number x_i whose values follow a distribution $p(x)$.

In words:

- As the number of samples $N \to \infty$, the average of samples converges to the expected value with probability 1.

In symbols:

$$P\left[\lim_{N\to\infty} \frac{1}{N}\left(\sum_{n=1}^{N} x_i\right) = E[X]\right] = 1.$$

Or:

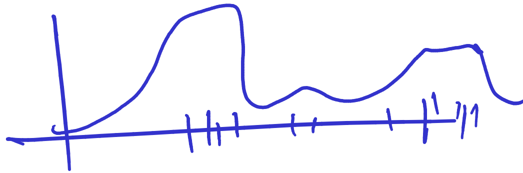
$$E[X] \approx \frac{1}{N}\left(\sum_{n=1}^{N} x_i\right)$$

Sampling: Approximating Expected Values

Integrals and sums in expected values are often challenging to evaluate.

- How can we approximate an expected value?

  Idea: Draw random samples. Make sure they are distributed according to $p(x)$.
- What is a Monte Carlo method?



$$\mathbb{E}[X] = \frac{1}{N}\sum_{i=1}^{N} x_i$$

outcome is not deterministic

→ Las Vegas methods: If they give a result, they give the right one

Expected Values with Hard-to-Sample Distributions

- Computing the sample mean requires samples from the distribution $p(x)$ of the random variable X. What if such samples aren't available?

has a dist. that's hard to sample from

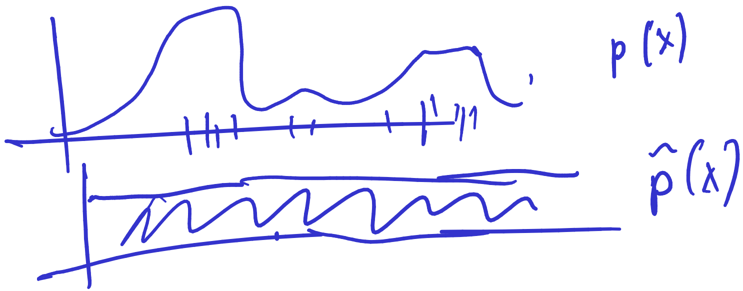
$$E[X] = \int_{\mathbb{R}} x \cdot p(x)\, dx$$

Assume $\tilde{p}(x) \neq 0$ $\rightarrow$

$$= \int_{\mathbb{R}} x \cdot \frac{p(x)}{\tilde{p}(x)}\, \hat{p}(x)\, dx$$

Assume $\hat{p}$ is dist. function of $\tilde{X}$. $\rightarrow$

$$= E\left[\tilde{X} \cdot \frac{p(\tilde{X})}{\tilde{p}(\hat{X})} \right]$$



Switching Distributions for Sampling

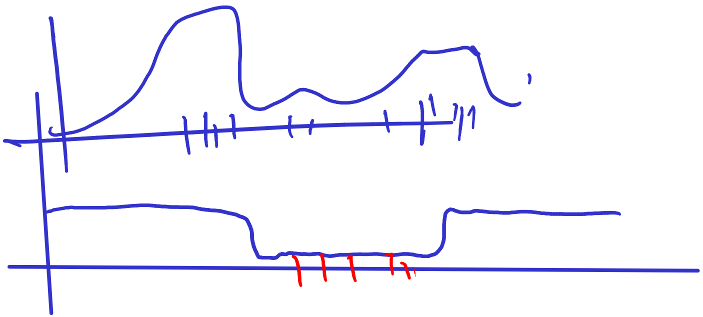
Found:

$$E[X] = E\left[\tilde{X} \cdot \frac{p(\tilde{X})}{\tilde{p}(\tilde{X})}\right]$$

- How do we apply this for sampling?

$$E(X) = E\left(\tilde{X} \cdot \frac{p(\tilde{X})}{\tilde{p}(\tilde{X})}\right) = \frac{1}{N} \cdot \sum_{i=1}^{N} \tilde{X}_i \cdot \frac{p(\tilde{X}_i)}{\tilde{p}(\tilde{X}_i)}$$

- When is this a good way to sample?



Want: samples contribute as equally as possible

$$\rightsquigarrow \frac{p(x)}{\tilde{p}(x)} \approx 1$$

is the best possible scenario.

Dealing with Unknown Scaling

- What if a distribution function is only known up to a constant factor, e.g.

$$p(x) = C \cdot \underbrace{\begin{cases} 1 & \text{point } x \text{ is in IL,} \\ 0 & \text{it isn't.} \end{cases}}_{\tilde{p}(x) \quad \hat{p}}$$

Typically $\int_{\mathbb{R}} \hat{p} \neq 1$. We need to find C so that $\int p = 1$, i.e.

$$C = \frac{1}{\int_{\mathbb{R}} \hat{p}(x) \mathrm{d}x}.$$

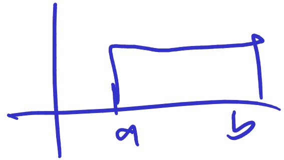
$\tilde{X}$ dist. according to $\tilde{p}$.

Idea: Use sampling.

$$\int \hat{p}(x)\,dx = \int \frac{\hat{p}(x)}{\tilde{p}(x)} \cdot \tilde{p}(x)\,dx = E\left[\frac{\hat{p}(\tilde{X})}{\tilde{p}(\tilde{X})}\right] \approx \frac{1}{N} \sum_{i=1}^{N} \frac{\hat{p}(\tilde{X}_i)}{\tilde{p}(\tilde{X}_i)}$$

$\tilde{p}$: actually a dist. function, sample-able

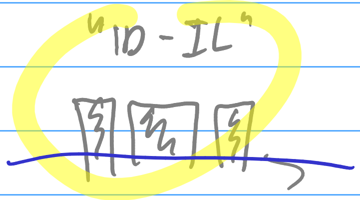
If we choose $\tilde{p}$ to be the uniform distribution



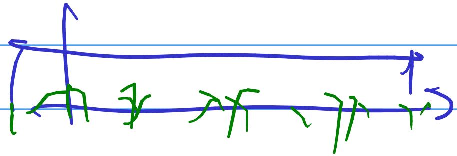
on $[a,b]$, then

$$\int \hat{p}(x)\,dx \simeq \frac{1}{N}\sum_{i=1}^{N}\frac{\hat{p}(\tilde{x}_i)}{\tilde{p}(\tilde{x}_i)} = \frac{b-a}{N}\sum_{i=1}^{N}\hat{p}(\tilde{x}_i)$$

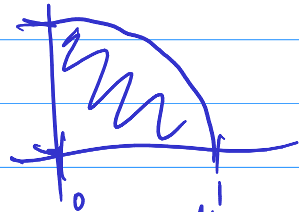
$\tilde{p}$ $\frac{1}{b-a}$



$\hat{p}(x)$



$\tilde{p}(x)$



$$\frac{(1-0)(1-0)}{N}\sum_{i=1}^{N}\hat{p}(\hat{x}_i)$$

Demo: Computing π using Sampling
Demo: Errors in Sampling

"on average"

$$\left| E[X] - \frac{1}{N}\sum_{i=1}^{N} X_i \right| = O\left(\frac{1}{\sqrt{n}}\right)$$

Sampling: Error

The Central Limit Theorem states that with

$$S_n := x_1 + x_2 + \cdots + x_n$$

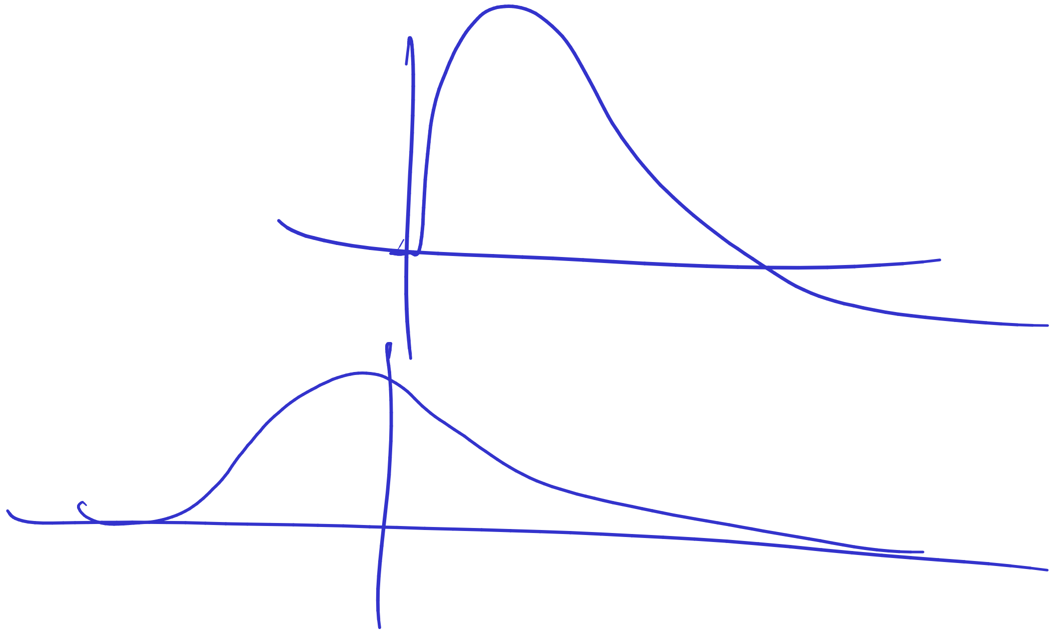
for the (x_i) independent and identically distributed we have that

$$\frac{S_n - n\,E[x_i]}{\sqrt{\sigma^2[x_i]n}} \to \mathcal{N}(0,1),$$

i.e. that term approaches the normal distribution. Or, short and imprecise:

$$\left|\frac{1}{n}S_n - E[x_i]\right| = O\left(\frac{1}{\sqrt{n}}\right).$$

In-class activity: Monte-Carlo Methods



Monte Carlo Methods: The Good and the Bad

- What are some *advantages of MC methods?*

- What are some *disadvantages* of MC methods?