Overview
- Distribution, expected value
- LLC / Sampling
- Error in Sampling
- RNG
What types of problems can we solve with the help of random numbers?

We can compute (potentially) complicated averages.

- Where does ‘the average’ web surfer end up? (PageRank)
- How much is my stock portfolio/option going to be worth?
- How will my robot behave if there is measurement error?
Random Variables

What is a **random variable**?

A **random variable** $X$ is a function that depends on ‘the (random) state of the world’.

**Example:** $X$ could be
- ‘how much rain tomorrow?’, or
- ‘will my buttered bread land face-down?’

**Idea:** Since I don’t know the entire state of the world (i.e. all the influencing factors), I can’t know the value of $X$.

→ Next best thing: Say something about the *average* case.

To do that, I need to know how likely each individual value of $X$ is. I need a **probability distribution**.
What kinds of probability distributions are there?

Continuous (dist/r.v.)

- Infinitely many possible values
- Each individual value has prob. 0, but ranges have finite probability;
  e.g. \( \Pr(7 \leq X \leq 7.5) = 0.1 \)

Discrete (dist/r.v.)

- Finite / discrete number of values
- E.g. die: 1, \ldots, 6
  - \( p = \frac{1}{6}, \ldots, \frac{1}{6} \)

**Demo:** Plotting Distributions with Histograms

In both cases, probabilities must be non-negative.
\[
\rho \geq 0 \\
\int p(x) \, dx = 1 \\
\sum p_i = 1
\]

\[
\begin{align*}
\Pr(0 \leq x \leq 0.1) & + \Pr(0.1 < x < 0.2) \\
& \quad + \Pr(0.2 \leq x \leq 0.4) \\
& \quad + \Pr(0.4 \leq x \leq 1) = 1
\end{align*}
\]
Expected Values/Averages: What?

Define 'expected value' of a random variable.

\[
\text{discrete: } \quad \mathbb{E}(X) = \sum_{i=1}^{N} x_i \cdot p(x_i)
\]

Define variance of a random variable.

\[
\sigma^2(X) = \mathbb{E}((X - \mathbb{E}(X))^2) = \mathbb{E}(X^2) - \mathbb{E}(X)^2
\]

\(\sigma^2(X)\) is the distance from avg. squared error average that
Expected Value: Example I

What is the expected snowfall in Champaign?

\[ E(X) = \sum_{x} x \cdot p(x) \text{ d}x \]

\( X = \) amount of snow

\( p(x) \)

Why approximate \( E(X) \)?
**Tool: Law of Large Numbers**

Terminology:

- **Sample**: A sample \( s_1, \ldots, s_N \) of a discrete random variable \( X \) (with potential values \( x_1, \ldots, x_n \)) selects each \( s_i \) such that \( s_i = x_j \) with probability \( p(x_j) \).

In words:

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

What can samples tell us about the distribution?

\[
P \left[ \lim_{N \to \infty} \frac{1}{N} \left( \sum_{i=1}^{N} s_i \right) = \mathbb{E}[X] \right] = 1
\]
Two difficulties in applying CLN:

1 - knowing the distribution

2 - being able to make samples following that distribution

\[ \int_0^1 p(x) \, dx = 0.79 \]
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** (MC) method?
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?

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

\[
= \int x \cdot \frac{p(x)}{\hat{p}(x)} \cdot \hat{p}(x) \, dx
\]

\[
= \mathbb{E}[X \cdot \frac{p(x)}{\hat{p}(x)}]
\]

Suppose $\hat{p}(x)$ is a distribution function that is easy to sample from.

Let $\hat{X}$ be distributed according to $\frac{p(x)}{\hat{p}(x)}$. 
Switching Distributions for Sampling

Found:

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

Why is this useful for sampling?

**In-class activity:** Monte-Carlo Methods
What is the expected snowfall in Illinois?