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?

$E[X] = \int x \ p(x) \ dx$

$= \sum_{i=1}^{N} x_i \ p(x_i)$

$\approx \sum_{i=1}^{N} \frac{x_i}{N}$

$E[f(X)] = \int f(x) \ p(x) \ dx$
\[ E[X] = \int x \, p(x) \, dx \]

Sample from \( \tilde{p}(x) \)

\[ E[\tilde{X}] = \int x \, \frac{p(x)}{\tilde{p}(x)} \, \frac{\tilde{p}(x)}{\tilde{p}(x)} \, dx \]

\[ \approx E \left[ \tilde{X} \cdot \frac{p(\tilde{X})}{\tilde{p}(\tilde{X})} \right] \]
\[ E[X \frac{p(X)}{\bar{p}(X)}] = \frac{1}{N} \sum_{i=1}^{N} \frac{\tilde{s}_i}{\bar{p}(s_i)} \]
Switching Distributions for Sampling

Found:

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

Why is this useful for sampling?

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

\[ \iiint_{\text{IL}} \text{Snow} (x, y) \cdot p(x, y) \, dx \, dy \]

\[ \int_0^{40} \int_0^{40} \text{Snow} (r) \, dr \]

\[ p(x, y) = \begin{cases} 1 & \text{if } (x, y) \text{ in IL} \\ 0 & \text{otherwise} \end{cases} \]
Dealing with Unknown Scaling

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

\[ p(x) = C \cdot \begin{cases} 
1 & \text{point } x \text{ is in IL,} \\
0 & \text{it isn’t.} 
\end{cases} 
\]

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

\[
C = \frac{1}{\int_{\mathbb{R}} q(x) dx}.
\]

Idea: Use sampling.
\[ \int \frac{q(x)}{\tilde{p}(x)} \tilde{p}(x) \, dx \]

\[ E \left[ E \left[ \frac{q(x)}{\tilde{p}(x)} \mid \tilde{p} \right] \right] = E \left[ \frac{q(x)}{\tilde{p}(x)} \right] \]

\[ \tilde{p}(x) = \frac{1}{n} \Rightarrow n \cdot \int q(x) \tilde{p}(x) \, dx \sim \frac{N}{N} \leq q(x) \]
Demo: Computing $\pi$ using Sampling
Demo: Errors in Sampling
The Central Limit Theorem states that with

$$S_N := X_1 + X_2 + \cdots + X_n$$

for the \((X_i)\) independent and identically distributed according to random variable \(X\) with variance \(\sigma^2\), we have that

$$\frac{S_N - NE[X]}{\sqrt{\sigma^2 N}} \to N(0, 1),$$

i.e. that term approaches the normal distribution. As we increase \(N\), \(\sigma^2\) stays fixed, so the asymptotic behavior of the error is

$$\left| \frac{1}{N} S_N - E[X] \right| = O \left( \frac{1}{\sqrt{N}} \right).$$
The Central Limit Theorem uses the fact that given \( N \) identically distribution samples of random variable \( X \) with variance \( \sigma^2[X] \), the average of the samples will have variance \( \sigma^2[X]/N \). Since \( \sigma^2[X] = E[(E[X] - X)^2] \) is the expected square of the deviation, it tells us how far away the average of the samples is expected to be from the real mean. Why is this the case?
\[ \sigma^2 [X] = E[(X - E[X])^2] \]

\[ = |E[X^2] - E[X]^2| \]

\[ \sigma^2 [S_n] = E[S_n^2] = \frac{1}{N} \sum_{i=1}^{N} x_i \]

\[ \sigma^2 [C_{S_n}] = E[(\frac{1}{N} \sum_{i=1}^{N} x_i)^2] \]

\[ = \frac{1}{N} \sum_{i=1}^{N} \sum_{j=1}^{N} E[X_i X_j] \]
\[\sigma^2 (s_n) = \frac{1}{N^2} \sum_{i=1}^{N} \sum_{i < j} E[x_i, x_j] = \frac{N \cdot \sigma^2 (x)}{N^2} + \sum_{i=1}^{N} \sum_{j \neq i} E[x_i, x_j]\]

\[E[x_i, x_j] = \int \int_{x_i, x_j} p(x_i) p(x_j) dx_i dx_j\]

\[= \left( \int x_i p(x_i) dx_i \right) \left( \int x_j p(x_j) dx_j \right)\]

\[= E[x_i] E[x_j]\]
Monte Carlo Methods: The Good and the Bad

What are some advantages of MC methods?

What are some disadvantages of MC methods?
Computers and Random Numbers

```c
int getRandNumber()
{
    return 4; // chosen by fair dice roll.
    // guaranteed to be random.
}
```

[from xkcd]

How can a computer make random numbers?
Random Numbers: What do we want?

What properties can ‘random numbers’ have?

- Have a specific distribution (e.g. ‘uniform’—each value in given interval is equally likely)
- Real-valued/integer-valued
- Repeatable (i.e. you may ask to exactly reproduce a sequence)
- Unpredictable
  - V1: ‘I have no idea what it’s going to do next.’
  - V2: No amount of engineering effort can get me the next number.
- Uncorrelated with later parts of the sequence (Weaker: Doesn’t repeat after a short time)
- Usable on parallel computers