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?

$$E[X] = \int \underline{x}\, p(x)\, dx$$

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

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

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

$$p(x)$$

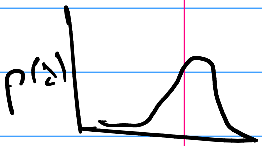
$$E[X] = \int x \, p(x) \, dx$$

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

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

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

$$E\left[\tilde{X}\,\frac{p(\tilde{X})}{\tilde{p}(\tilde{X})}\right] \approx \sum_{i=1}^{N} \frac{\tilde{s}_i}{N}\,\frac{p(s_i)}{\tilde{p}(s_i)}$$



$$\tilde{p}(x) = \frac{1}{n}$$

$$E\left[\tilde{X}\,\frac{p(\tilde{X})}{\tilde{p}(\tilde{X})}\right] = \frac{n}{N}\sum_{i=1}^{N}$$

Switching Distributions for Sampling

Found:

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

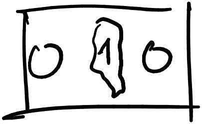
Why is this useful for sampling?

Expected Value: Example II

What is the expected snowfall in Illinois?

$$\iint \text{snow}(x,y)\cdot p(x,y)\,dx\,dy$$

$$\int_{IL} \text{snow}(r)\,dr$$



$$p(x,y) = \begin{cases} 1 : \text{if } (x,y) \text{ in } IL \\ 0 : \text{otherwise} \end{cases}$$

Dealing with Unknown Scaling

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

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

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.

$\int q(x) dx = \frac{1}{C}$ $\quad E[1]$

$$\int \frac{q(x)}{\tilde{p}(x)} \tilde{p}(x)\, dx$$

$$E_q[1] = E_{\tilde{p}}\left[\frac{q(x)}{\tilde{p}(x)}\right]$$

$$\tilde{p}(x) = \frac{1}{n} \Longrightarrow n \cdot \int q(x)\, \tilde{p}(x)\, dx$$

$$\approx \frac{n}{N} \sum_{i=1}^{N} q(s_i)$$

Demo: Computing π using Sampling

Demo: Errors in Sampling

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 according to random variable X with variance σ^2, we have that

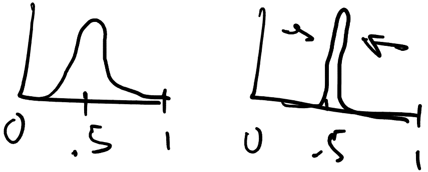
$$\frac{S_N - NE[X]}{\sqrt{\sigma^2 N}} \to \mathcal{N}(0, 1),$$

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

$$\left| \frac{1}{N} S_N - E[X] \right| = O\left(\frac{1}{\sqrt{N}} \right).$$

Proof Intuition for Central Limit Theorem

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?



$N \to$

$$\sigma^2[X] = E[(X - E[X])^2]$$

$$= \boxed{E[X^2] - E[X]^2}$$

$$\sigma^2[X] = E[X^2]$$

$$S_n = \frac{1}{N}(X_1 + \dots + X_n) = \frac{1}{N}\sum_{i=1}^{n} X_i$$

$$\sigma^2[S_n] = E\left[\left(\sum_{i=1}^{n} X_i\right)^2\right] \qquad \frac{1}{N}$$

$$= \sum_{i=1}^{n}\sum_{j=1}^{n} E[X_i X_j]$$

$$\sigma^2[S_n] = \frac{1}{N^2}\sum_{i=1}^{n} E[X_i^2] = \frac{N \cdot \sigma^2[X]}{N^2}$$

$$+ \sum_{i=1}^{n}\sum_{j\neq i} E[X_i X_j]$$

$$E[X_i X_j] = \iint x_i x_j\, p(x_i)\, p(x_j)\, dx_i\, dx_j$$

$$= \left(\int x_i\, p(x_i)\, dx_i\right)\left(\int x_j\, p(x_j)\, dx_j\right)$$

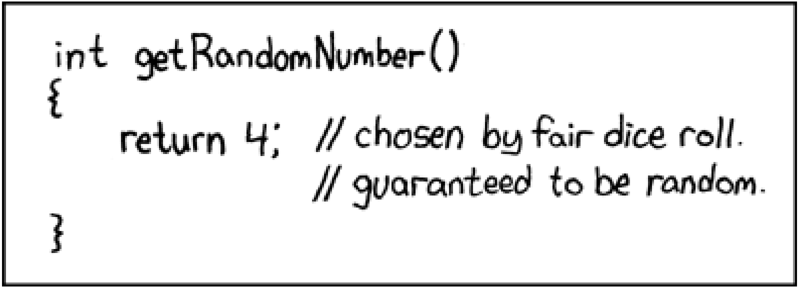
$$= E[X_i]\, E[X_j]$$

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



[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