Recap

• (Ch 4) Random variables
  • Expected value, variance and covariance
  • Towards the weak law of large numbers

Today

• (Ch 4) Random variables
  • Weak law of large numbers
  • Simulation examples
  • Continuous random variables
Towards the weak law of large numbers

• The weak law says that if we repeat an experiment many times, the average of the observations will “converge” to the expected value.

• The weak law justifies using simulations (instead of calculations) to estimate the expected values of random variables.
Markov’s inequality

• For any random variable $X$ and constant $a > 0$

$$P(|X| \geq a) \leq \frac{E[|X|]}{a}$$

• In words, a random variable is unlikely to have an absolute value much larger than the mean of its absolute value

• For example, if $a = 10E[|X|]$

$$P(|X| \geq 10E[|X|]) \leq 0.1$$
Chebyshev’s inequality

• For any random variable $X$ and constant $a > 0$

\[
P(|X - E[X]| \geq a) \leq \frac{\text{var}[X]}{a^2}
\]

• To rephrase, let $a = k\sigma$ where $\sigma = \text{std}[X]$

\[
P(|X - E[X]| \geq k\sigma) \leq \frac{1}{k^2}
\]

• In words, the probability that $X$ is greater than $k$ standard deviations from the mean is small
Proof of Chebyshev’s inequality

• Apply Markov’s inequality to $U = (X - E[X])^2$, so $E[U] = \text{var}[X]$

\[
P(|U| \geq w) \leq \frac{\mathbb{E}[|U|]}{w} = \frac{\mathbb{E}[U]}{w} = \frac{\text{var}[X]}{w}
\]

• Substitute $U = (X - E[X])^2$ and $w = a^2$

\[
P\left( (X - E[X])^2 \geq a^2 \right) \leq \frac{\text{var}[X]}{a^2}
\]

\[
P\left( |X - E[X]| \geq a \right)
\]
IID samples and the sample mean

• Say you have a random variable $X$ with probability distribution $P(X)$

  \[ \text{randomly} \]

• Say you generate independent samples $\{x_i\}$ so that their histogram resembles $P(X)$ more closely as the number of samples increases

• We call $\{x_i\}$ independent identically distributed (IID) samples of $P(X)$

• The sample mean of $\{x_i\}$ is a random variable:

  \[ X_N = \frac{1}{N} \sum_{i=1}^{N} x_i \]
Expected value of sample mean

• By linearity of expected value

\[ E[X_N] = E\left[\frac{1}{N}\sum_{i=1}^{N} x_i\right] = \frac{1}{N} \sum_{i=1}^{N} E[x_i] \]

• Since each \( x_i \) is a sample drawn from \( P(X) \), we have \( E[x_i] = E[X] \)

\[ E[X_N] = \frac{1}{N} \sum_{i=1}^{N} E[X] = E[X] \]
Variance of sample mean

- By the scaling property of variance and independence of samples $x_i$

$$\text{var}[X_N] = \text{var}\left[\frac{1}{N} \sum_{i=1}^{N} x_i\right] = \frac{1}{N^2} \text{var}\left[\sum_{i=1}^{N} x_i\right] = \frac{1}{N^2} \sum_{i=1}^{N} \text{var}[x_i]$$

- Since each $x_i$ is drawn from $P(X)$, we have $\text{var}[x_i] = \text{var}[X]$

$$\text{var}[X_N] = \frac{1}{N^2} \sum_{i=1}^{N} \text{var}[X] = \frac{\text{var}[X]}{N}$$
Weak law of large numbers (WLLN)

• Given a random variable $X$ with finite variance, probability distribution $P(X)$ and sample mean $X_N$

• For any positive number $\epsilon$

$$\lim_{N \to \infty} P(|X_N - E[X]| \geq \epsilon) = 0$$

• In words: for a large enough set of IID samples, the sample mean $X_N$ will be very close to the expected value $E[X]$ with high probability
Proof of WLLN

• Apply Chebyshev’s inequality to the sample mean $X_N$

$$ P \left( \left| X_N - \mathbb{E} [X_N] \right| \geq \varepsilon \right) \leq \frac{\text{var} \left[ X_N \right]}{\varepsilon^2} $$

• Substitute $E[X_N] = E[X]$ and substitute $\text{var}[X_N] = \text{var}[X]/N$

$$ P \left( \left| X_N - \mathbb{E} [X] \right| \geq \varepsilon \right) \leq \frac{\text{var} \left[ X \right]}{N\varepsilon^2} \xrightarrow{N \to \infty} 0 $$
Simulation: airline overbooking example

An airline has a flight with 6 seats. They always sell 12 tickets for this flight and ticket holders show up independently with probability $p$. Plot the following quantities as a function of $p$.

- Expected number of ticket holders that show up
- Probability that the flight is overbooked
- Expected number of ticket holders who show up but don’t fly given that the flight is overbooked
Approximating expected value

Expected number of ticket holders that show up

```python
results = np.zeros((10, 2))
numTrials = 100000
numTickets = 12
numSeats = 6
for i,p in enumerate(np.linspace(0.1, 1.0, num=10)):
    arrivals = np.random.random((numTickets, numTrials)) < p
    numArrivals = arrivals.sum(axis=0)
    results[i] = [p, numArrivals.mean()]
results
```
Approximating probability

Probability that the flight is overbooked

```python
results = np.zeros((10, 2))
numTrials = 100000
numTickets = 12
numSeats = 6
for i, p in enumerate(np.linspace(0.1, 1.0, num=10)):
    arrivals = np.random.random((numTickets, numTrials)) < p
    numArrivals = arrivals.sum(axis=0)
    indicatorOverbooked = numArrivals > numSeats
    results[i] = [p, indicatorOverbooked.mean()]
results
```
Approximating conditional expected value

Expected number of ticket holders who show up but don’t fly given that the flight is overbooked

```python
results = np.zeros((10, 2))
numTrials = 100000
numTickets = 12
numSeats = 6
for i, p in enumerate(np.linspace(0.1, 1.0, num=10)):
    arrivals = np.random.random((numTickets, numTrials)) < p
    numArrivals = arrivals.sum(axis=0)
    indicatorOverbooked = numArrivals > numSeats
    numDontFly = numArrivals[indicatorOverbooked] - numSeats
    results[i] = [p, numDontFly.mean()]
results
```
Continuous random variables

• So far we have been talking about discrete random variables

• Some random variables can take on a continuous set of values
  • Temperature
  • Height
  • Sample mean $X_N$ (whoops!)

• Defining samples spaces, outcomes and events for continuous random variables is beyond the scope of CS 361
Probability density function (pdf)

- For a continuous random variable $X$, the probability that $X = x$ is essentially zero for all (or most) $x$, so we can’t define $P(X = x)$.

- Instead, we define the **probability density function (pdf)** over an infinitesimally small interval $dx$:

$$p(x)dx = P(X \in [x, x + dx])$$

- For $a < b$

$$\int_{a}^{b} p(x) \, dx = P(X \in [a, b])$$
Properties of the probability density function

- $p(x)$ is a bit like a discrete random variable’s probability distribution
  - $p(x) \geq 0$ for all $x$
  - The probability of $X$ taking some value is 1

$$\int_{-\infty}^{\infty} p(x) \, dx = 1$$

- $p(x)$ is **not** like a discrete random variable’s probability distribution
  - $p(x)$ is not the probability that $X = x$
  - $p(x)$ can exceed 1
Probability density function: height example

- Suppose we heard that Napoleon was 62.5 inches tall, rounded up to the nearest half inch. What is the pdf of his height $H$?

- Assume that $H$ is equally likely to be any value in $(62, 62.5]$ inches

$$p(h) = \begin{cases} 
\frac{c}{2} & \text{if } h \in (62, 62.5] \\
0 & \text{if } h \notin (62, 62.5] 
\end{cases}$$

where $c$ is a constant

- Then

$$1 = \int_{-\infty}^{\infty} p(h) \, dh = \int_{62}^{62.5} c \, dh = \frac{c}{2} \implies c = 2$$