COMPSCI 240: Reasoning Under Uncertainty

Andrew Lan and Nic Herndon

University of Massachusetts at Amherst

Spring 2019
Lecture 18: Limit Theorems
Overview

- Let $X_1, X_2, \cdots, X_n$ be a sequence of i.i.d. (either discrete or continuous) random variables with mean of $\mu$ and variance of $\sigma^2$.

- Limit theorems are mostly concerned with the sum of these random variables (which forms another random variable): 

$$S_n = X_1 + X_2 + \cdots + X_n$$

especially when $n$ is very large.

- Then, the mean and variance of $S_n$ can be computed as

$$E[S_n] = E[X_1] + E[X_2] + \cdots + E[X_n] = n\mu$$

$$\text{var}(S_n) = \text{var}(X_1) + \text{var}(X_2) + \cdots + \text{var}(X_n) = n\sigma^2$$
Overview

• Let us introduce a new RV:

$$Z_n = \frac{S_n - n\mu}{\sigma \sqrt{n}}$$

• The mean and variance of $Z_n$ is

$$E[Z_n] = 0$$

$$\text{var}(Z_n) = 1$$

• The **central limit theorem** states that the distribution of $Z_n$ becomes the **standard normal variable** as $n$ increases.
Markov and Chebyshev Bounds

We will learn these two bounds to prove the Central Limit Theorem. More specifically, Markov Inequality $\rightarrow$ Chebyshev Inequality $\rightarrow$ Central Limit Theorem.

- **Markov Bound:** For a non-negative random variable $X$,

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

- **Chebyshev Bound:** For a random variable $X$,

  $$P(|X - E[X]| \geq c) \leq \frac{\text{var}(X)}{c^2}$$
The Markov Bound

- **Markov Bound**: For any non-negative random variable, \( P(X \geq a) \leq \frac{E[X]}{a} \)

- **Proof**: Introduce a new RV \( Y_a \) where

\[
Y_a = \begin{cases} 
0 & \text{if } X < a \\
 a & \text{if } X \geq a
\end{cases}
\]

- Then, we know that \( Y_a \leq X \), which yields

\[
E[Y_a] \leq E[X]
\]

- On the other hand,

\[
E[Y_a] = a \cdot P(Y_a = a) = a \cdot P(X \geq a)
\]

- Thus,

\[
P(X \geq a) \leq \frac{E[X]}{a}
\]
The Markov Bound

- Let $X$ be a continuous random variable with uniform density over $[0,4]$.
- Its mean can be computed as

$$E[X] = \int_{0}^{4} \frac{1}{4} dx = 2$$

- Then, the Markov inequality asserts that

$$P(X \geq 0) \leq 1 \text{ whereas } P(X \geq 0) = 1.$$  

$$P(X \geq 1) \leq 1 \text{ whereas } P(X \geq 1) = \frac{3}{4}.$$  

$$P(X \geq 2) \leq 1 \text{ whereas } P(X \geq 2) = 0.5.$$  

These are uninformative...

$$P(X \geq 3) \leq \frac{2}{3} \text{ whereas } P(X \geq 3) = \frac{1}{4}.$$  

$$P(X \geq 4) \leq \frac{2}{4} \text{ whereas } P(X \geq 4) = 0.$$  

- The Markov inequality can be quite loose.
The Chebyshev Bound

- **Chebyshev Bound:**
  \[ P(|X - E[X]| \geq c) \leq \frac{\sigma^2}{c^2}, \text{ for all } c > 0 \]

- **Proof:** Let us introduce a non-negative RV \((X - \mu)^2\).
  - We can apply the Markov bound on this RV:
    \[ P((X - \mu)^2 \geq a) \leq \frac{E[(X - \mu)^2]}{a} = \frac{\sigma^2}{a} \]
  - We will only consider \(a > 0\) where \(a\) can be defined as \(a = c^2\). Then, we have
    \[ P((X - \mu)^2 \geq c^2) \leq \frac{\sigma^2}{c^2} \]
  - Since the event \((X - \mu)^2 \geq c^2\) is identical to the event \(|X - \mu| \geq c\),
    \[ P((X - \mu)^2 \geq c^2) = P(|X - \mu| \geq c) \leq \frac{\sigma^2}{c^2} \]
  - The Chebyshev bound is often more powerful than the Markov bound because it also uses information on the variance of X.
An Alternate Form of The Chebyshev Bound

- Let $c = k\sigma$, then

$$P(|X - \mu| \geq k\sigma) = P \left( \left| \frac{X - \mu}{\sigma} \right| \geq k \right) = \frac{1}{k^2}$$

- The probability that a RV takes a value more than $k$ standard deviations from its mean is at most $1/k^2$. 
The Chebyshev Bound

- Let $X$ be a continuous random variable with uniform density over $[0,4]$.
- Its mean is
  \[ E[X] = \int_{0}^{4} \frac{1}{4} dx = 2. \]
- Its variance is
  \[ \text{var}(X) = E[X^2] - E[X]^2 = \frac{4}{3} \]
- Then, the Chebyshev inequality asserts that
  \[ P(|X - 2| \geq 0) \leq 1 \] so is not informative
  \[ P(|X - 2| \geq 1) \leq 1 \] so is not informative
  \[ P(|X - 2| \geq 2) \leq \frac{1}{3} \] or equivalently \( P(X \geq 4 \cup X \leq 0) \leq \frac{1}{3} \)
- We can also derive that
  \[ P(0 < X < 4) \geq \frac{2}{3} \]
Example

Suppose we know that the number of items produced in a factory during a week is a random variable with mean of 50 and variance of 25.

- What can be said about the probability that this week’s production will exceed 75?

By Markov’s inequality

\[ P(X \geq 75) \leq \frac{50}{75} = \frac{2}{3} \]
Example

Suppose we know that the number of items produced in a factory during a week is a random variable with mean of 50 and variance of 25.

- What can be said about the probability that this week’s production will be between 40 and 60?
  By Chebyshev’s inequality

\[
P(|X - 50| \geq 10) \leq \frac{25}{100} = \frac{1}{4}
\]

Thus,

\[
P(40 \leq X \leq 60) = 1 - P(|X - 50| \geq 10) \geq \frac{3}{4}
\]