COMPSCI 240: Reasoning Under Uncertainty

Nic Herndon and Andrew Lan

University of Massachusetts at Amherst

Spring 2019

Lecture 19: Weak law of large numbers & Convergence in probability

Markov and Chebyshev Bounds

- **Markov Bound**
  - Informally: If a nonnegative RV has a small mean, then the probability that this RV takes a large value must also be small.
  - Formally: For a non-negative random variable X,

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

- **Chebyshev Bound**
  - Informally: If a RV has small variance, then the probability that it takes a value far from its mean is also small. Note that the Chebyshev inequality does not require the random variable to be nonnegative.
  - Formally: For a random variable X,

$$P(|X - E(X)| \geq c) \leq \frac{Var(X)}{c^2}$$

- The mean and the variance of a RV are only a rough summary of its properties, and we cannot expect the bounds to be close approximations of the exact probabilities.

Sample Mean

- Let $X_1, X_2, \cdots, X_n$ be a sequence of i.i.d. (either discrete or continuous) random variables with mean of μ and variance of σ^2.
- Its sample (empirical) mean can be computed as

$$\overline{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i$$

Note that $\overline{X}_n$ is also a random variable.

- We know that the expected value of the sample mean is

$$E\left[\overline{X}_n\right] = E\left[\frac{1}{n} \sum_{i=1}^{n} X_i\right] = \frac{1}{n} \sum_{i=1}^{n} E(X_i) = \frac{1}{n} n\mu$$
$$= \mu$$

- We also know that the variance and standard deviations of the sample mean are

$$\text{Var}(\overline{X}_n) = \text{Var}\left(\frac{\sum_{i=1}^{n} X_i}{n}\right) = \frac{1}{n^2} \cdot n \cdot \sigma^2$$
$$= \frac{\sigma^2}{n}$$
$$Std(\overline{X}_n) = \frac{\sigma}{\sqrt{n}}$$

The Weak Law of Large Numbers

- Let $X_1, X_2, \ldots$ be a sequence of i.i.d. (either discrete or continuous) random variables with mean μ and variance σ^2. For every $\epsilon > 0$, we have

$$P\left(|\overline{X}_n - \mu| \geq \epsilon\right) \to 0 \text{ as } n \to \infty.$$

- The weak law of large numbers states that if n is large, the bulk of the distribution of $\overline{X}_n$ will converge to (be concentrated around) μ.
- That is, if we consider a positive length interval $[\mu - \epsilon, \mu + \epsilon]$ around μ, then there is high probability that $\overline{X}_n$ will fall in that interval; as $n \to \infty$, this probability converges to 1. If ϵ is very small, we may have to wait longer (i.e., need a larger value of n) before this probability converges to 1.

The Weak Law of Large Numbers

- Let $X_1, X_2, \cdots$ be a sequence of i.i.d. (either discrete or continuous) random variable with mean μ and variance σ^2. For every $\epsilon > 0$, we have

$$P\left(|\overline{X}_n - \mu| \geq \epsilon\right) \rightarrow 0 \text{ as } n \rightarrow \infty.$$

- **Proof**:
- We know that the **Chebyshev bound** for a random variable X defines

$$P(|X - \mu| \geq \epsilon) \leq \frac{\text{Var}(X)}{\epsilon^2}$$

- Using this, we can write the weak law of large numbers as

$$P\left(|\overline{X}_n - \mu| \geq \epsilon\right) \leq \frac{\text{Var}(\overline{X}_n)}{\epsilon^2} = \frac{\sigma^2}{n\epsilon^2}$$

- Thus,

$$\begin{aligned}\lim_{n\rightarrow\infty} P\left(|\overline{X}_n - \mu| \geq \epsilon\right) &\leq \lim_{n\rightarrow\infty} \frac{\sigma^2}{n\epsilon^2} \\ &= 0\end{aligned}$$

Example 1

- Consider an event A with probability $p = P(A)$.
- We repeat the experiment n times.
- Let $\overline{X}_n$ be the fraction of time that event A occurs.
  This is the **empirical frequency** of A

  $$\overline{X}_n = \frac{X_1 + \cdots + X_n}{n},$$

  where $X_i = 1$ whenever A occurs, and 0 otherwise; thus $E[X_i] = p$.
- The weak law applies and shows that when n is large, the empirical frequency is most likely to be within ϵ of p.
- Loosely speaking, this allows us to conclude that empirical frequencies are faithful estimates of p.
- Alternatively, this is a step towards interpreting the probability p as the frequency of occurrence of A.

Example 2

- Let p be the fraction of voters who support a particular candidate for office.
- We interview n "randomly selected" voters and record $\overline{X}_n$, the fraction of them that support the candidate.
- We view $\overline{X}_n$ as our estimate of p and would like to investigate its properties (the true value of p is assumed to be unknown).
- The response of each person interviewed can be viewed as an independent Bernoulli random variable X, with success probability p and variance $\sigma^2 = p(1-p)$.
- The Chebyshev inequality yields

$$P(|\overline{X}_n - p| \geq \epsilon) \leq \frac{p(1-p)}{n\epsilon^2}$$

- Since $p(1-p) \leq 1/4$ (Example 5.3 in the textbook), we have

$$P(|\overline{X}_n - p| \geq \epsilon) \leq \frac{1}{4n\epsilon^2}$$

Example 2 (cont.)

$$P(|\overline{X}_n - p| \geq \epsilon) \leq \frac{1}{4n\epsilon^2}$$

- Let $\epsilon = 0.1$ and $n = 100$:

  $$P(|\overline{X}_{100} - p| \geq 0.1) \leq \frac{1}{4 \cdot 100 \cdot (0.1)^2} = 0.25$$

  That is, with a sample size of $n = 100$, the probability that our estimate is incorrect by more than 0.1 is no larger than 0.25.
- Let's say we'd like to have high confidence (probability at least 95%) that our estimate is within 0.01 of p accurate. How many voters should be sampled?

  $$P(|\overline{X}_n - p| \geq 0.1) \leq \frac{1}{4n(0.1)^2} \leq 1 - 0.95$$

  $$n \geq 50,000$$

Convergence in probability

- Let $Y_1, Y2, \ldots$ be a sequence of random variables (not necessarily independent), and let a be a real number.
- We say that the sequence Y_n **converges to a in probability**, if for every $\epsilon > 0$, we have

$$\lim_{n \to \infty} P(|Y_n - a| \geq \epsilon) = 0$$

- Given this definition, the weak law of large numbers simply states that the sample mean converges in probability to the true mean μ.

Example

- In order to estimate f, the true fraction of smokers in a large population, Alvin selects n people at random. His estimator $\overline{X}_n$ is obtained by dividing X_n, the number of smokers in the sample, by n, i.e., $\overline{X}_n = X_n/n$. Alvin choose the sample size n to be the smallest possible number for which the Chebyshev inequality yields a guarantee that

$$P(|\overline{X}_n - f| \geq \epsilon) \leq \delta$$

where ϵ and δ are some predefined tolerances. Determine how the value of n recommended by the Chebyshev inequality changes in the following cases.

  a) The value of ϵ is reduced to half its original value.
  b) The probability δ is reduced to half its original value.

Example (solution)

- The best guarantee that can be obtained from the Chebyshev inequality is

$$P(|\overline{X}_n - f| \geq \epsilon) \leq \frac{1}{4n\epsilon^2}$$

a) How should the value of n be updated if ϵ is reduced to half its original value?

$$\frac{1}{4n\epsilon^2} = \frac{1}{4n'\epsilon'^2} \Rightarrow n' = \frac{n\epsilon^2}{\epsilon'^2} = \frac{n\epsilon^2}{(\epsilon/2)^2} = 4n$$

The sample size should be four times larger.

b) How should the value of n be updated if the probability δ is reduced to half its original value?

$$\frac{1}{4n\epsilon^2} = \frac{2}{4n'\epsilon^2} \Rightarrow n' = 2n$$

The sample size should be doubled.

Usefulness of limit theorems

- Conceptually, they provide an interpretation of expectations (as well as probabilities) is terms of a long sequence of identical independent experiments.
- They allow for an *approximate* analysis of the properties of random variables such as X_n. This is to be contrasted with an *exact* analysis which would require a formula for the PMF or PDF of X_n, a complicated and tedious task when n is large.
- They play a major role in inference and statistics, in the presence of large data sets.