Tail Bounds and the Law of Large Numbers

Gregory M. Shinault

Big Idea

Tail bounds give us rough estimates for large deviations from the mean. This can be practical if a sample does not include any large deviations.

We also finally justify my assertion that the expected value is essentially the long run average.

This material corresponds to sections 9.1 and 9.2 of the textbook.

Markov Inequality

The Statement

Fact: Suppose $X \geq 0$ with probability 1. Then for any $c > 0$ we have
$$\mathbb{P}(X \geq c) \leq \frac{\mathbb{E}X}{c}.$$

Example

Suppose $X \geq 0$ and $\mathbb{E}X = 50$.

1. Give a bound on $\mathbb{P}(X \geq 60)$.
2. Assume further $\sigma^2 = 25$. Give a bound on $\mathbb{P}(X \geq 60)$.

Example | Lesson

The Markov inequality does not give us a method to use the information on the variance to refine our estimate of the tail probability.

Next we will develop a method to use the variance for more accurate estimates.

Chebyshev Inequality

The Statement

Suppose $\mathbb{E}X = \mu < \infty$ and $\mathrm{Var}(X) = \sigma^2 < \infty$. Then
$$\mathbb{P}(|X - \mu| \geq c) \leq \frac{\mathrm{Var}(X)}{c^2}.$$

Example

Suppose $X \geq 0$ and $\mathbb{E}X = 50$.

1. Give a bound on $\mathbb{P}(X \geq 60)$.
2. Assume further $\sigma^2 = 25$. Give a bound on $\mathbb{P}(X \geq 60)$.
3. Now assume further X is a binomial RV. Estimate $\mathbb{P}(X \geq 60)$.

Weak Law of Large Numbers

The Statement

Theorem: Suppose $X_1, X_2, \ldots$ is an IID sequence of RVs with mean $\mathbb{E}X_1 = \mu$ and variance $\mathrm{Var}(X_1) = \sigma^2$. Let $\bar{X}_n$ be the sample mean,

$$\bar{X}_n = \frac{X_1 + \cdots + X_n}{n}.$$

Then for any $\varepsilon > 0$ we have

$$\lim_{n\to\infty} \mathbb{P}(|\bar{X}_n - \mu| < \varepsilon) = 1.$$

Classic Example

Suppose we take an random sample of IID random variables: $X_1, X_2, \ldots, X_n$. We would like to use the sample mean

$$\hat{\mu} = \frac{1}{n}(X_1 + X_2 + \cdots + X_n)$$

to estimate the expected value of the random variables $\mathbb{E}X_j$. Suppose we know the variance of the RVs is no greater than 2.

Will this work? If so, how large must our sample be for the sample mean to be within 0.05 of the correct value of $\mathbb{E}X_j$ with probability at least 0.99?

Summary

Key Ideas

1. The Markov inequality provides a tail bound if you only know the mean.
2. The Chebyshev inequality provides a tail bound if you know the mean and the variance.
3. The WLLN is an important conceptual tool.