Lecture Summary: Sum of Independent Random Variables and the Law of Large Numbers

Source: lec7.4.pdf

Key Points

- **Sum of Random Variables:**
  - For n random variables $X_1, X_2, \ldots, X_n$:
  $$S = \sum_{i=1}^{n} X_i.$$
  - The expected value of S is:
  $$\mathbb{E}[S] = \sum_{i=1}^{n} \mathbb{E}[X_i].$$
  - If X_i are pairwise uncorrelated, the variance of S is:
  $$\text{Var}(S) = \sum_{i=1}^{n} \text{Var}(X_i).$$
  - Independence implies pairwise uncorrelation, but the converse is not true.
- **Linear Combinations of Random Variables:**
  - For a linear combination of random variables:
  $$L = \sum_{i=1}^{n} a_i X_i,$$
  where a_i are constants:
  $$\mathbb{E}[L] = \sum_{i=1}^{n} a_i \mathbb{E}[X_i],$$
  $$\text{Var}(L) = \sum_{i=1}^{n} a_i^2 \text{Var}(X_i), \quad \text{if pairwise uncorrelated.}$$
- **Sample Mean:**
  - For $X_1, X_2, \ldots, X_n$ iid with mean μ and variance σ^2, the sample mean is:
  $$\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i.$$

- Expected value:
$$\mathbb{E}[\bar{X}] = \mu.$$
- Variance:
$$\mathrm{Var}(\bar{X}) = \frac{\sigma^2}{n}.$$

- **Weak Law of Large Numbers (WLLN):**
  - States that as $n \to \infty$, the sample mean $\bar{X}$ converges to the true mean μ in probability:
  $$P(|\bar{X} - \mu| > \delta) \leq \frac{\sigma^2}{n\delta^2}.$$
  - Derived using Chebyshev's inequality.
  - Highlights the shrinking variance of $\bar{X}$ as n increases.

- **Examples and Applications:**
  - **Bernoulli(p) Trials:**
    * Variance: $\sigma^2 = p(1-p)$.
    * Probability of sample mean lying within $[p-\delta, p+\delta]$ increases with n.
  - **Uniform and Normal Distributions:**
    * Variances depend on distribution parameters and scale with $1/n$ for the sample mean.
  - **Real Data Examples:**
    * Iris dataset (sepal length): Sample variance $\sigma^2 = 0.1242$ for 50 samples.
    * Taj Mahal air quality (PM2.5): Sample variance $\sigma^2 = 15.92$ for 11 samples, highlighting issues with small datasets.
    * IPL scores: Large sample size (1598 observations) enables high confidence in the sample mean estimate.

- **Insights:**
  - Larger sample sizes lead to smaller confidence intervals for the sample mean.
  - Small datasets may result in wide confidence intervals, limiting precision.
  - The WLLN provides probabilistic guarantees for the convergence of the sample mean to the true mean.

Simplified Explanation

Sum of Random Variables: - Mean and variance of the sum depend on whether the variables are uncorrelated or independent.

Sample Mean: - Converges to the true mean as the sample size increases (WLLN).

Applications: - Analyze datasets like Iris, Taj Mahal air quality, and IPL scores to see how sample size affects confidence in results.

Key Takeaway: Larger datasets improve the reliability of statistical estimates, with smaller variances in sample means.

Conclusion

In this lecture, we:

- Discussed the properties of sums and means of random variables.
- Introduced the Weak Law of Large Numbers and its implications.
- Examined the effects of sample size on confidence intervals using real-world examples.

The WLLN and related concepts are foundational for understanding how sample statistics approximate population parameters.