Lecture Summary: Approximating CDFs with Continuous Models

Source: Lecture 4.3.docx

Key Points

- **Transitioning from Discrete to Continuous CDFs:**
  - As the alphabet size of a discrete random variable increases, the stepwise nature of its cumulative distribution function (CDF) begins to resemble a continuous line.
  - Continuous CDFs provide a simpler description, avoiding the complexity of stepwise structures in large datasets.
- **Advantages of Continuous CDFs:**
  - Simplifies probability calculations, especially for large datasets.
  - Shifts focus from exact values to intervals, aligning with real-world scenarios.
  - Reduces computational effort compared to summing probabilities for discrete steps.
- **Examples:**
  - **Binomial Distribution:**
    * For $n = 10$, steps in the CDF are visible, with jumps at each integer.
    * For $n = 100$, the steps shrink, and the CDF appears smoother.
    * For large n, the CDF becomes indistinguishable from a continuous curve, allowing for approximation.
  - **Meteorite Weights:**
    * Raw data spans from 0.01 grams to 60 tons.
    * Histogram-derived CDFs show small steps that approximate a smooth curve.
    * Continuous approximations capture essential probabilities without requiring precise stepwise values.
- **Continuous Approximation of Uniform Distributions:**
  - For $X \sim \text{Uniform}\{1, 2, \ldots, 100\}$:
    * Discrete CDF: Stepwise increments of 1/100.
    * Continuous approximation: Linear function $F_X(x) = \frac{x}{100}$ for $x \in [1, 100]$.
  - Approximation simplifies interval probability calculations:

$$P(3.2 \leq X \leq 10.6) \approx F_X(10.6) - F_X(3.2) = \frac{10.6}{100} - \frac{3.2}{100} = 0.074.$$

  - Exact CDF yields 0.07, showing a small difference due to approximation.
- **Continuous Approximations for Binomial Distributions:**
  - Discrete binomial CDFs are cumbersome for large n.

- Simple continuous approximations, like logistic functions, provide smooth alternatives:

$$F_X(x) \approx \frac{1}{1 + e^{-k(x-\mu)}},$$

  where μ is the mean, and k is a scaling parameter.
- These approximations are accurate for calculating probabilities over intervals, with slight deviations.

- **Theory of Continuous CDFs:**
  - A function $F(x)$ is a valid CDF if:
    * $F(x)$ is non-decreasing.
    * $\lim_{x\to-\infty} F(x) = 0$ and $\lim_{x\to\infty} F(x) = 1$.
    * $F(x)$ is right-continuous.
  - Continuous CDFs provide a framework for transitioning from discrete to continuous random variables.

Simplified Explanation

Why Approximate CDFs? As datasets grow, discrete CDFs with many steps become impractical. Continuous CDFs offer simpler and computationally efficient alternatives.

Examples: - Uniform Distribution: Replace stepwise increments with a linear approximation. - **Binomial Distribution:** Use smooth functions to approximate large-step CDFs.

Key Concept: Continuous CDFs are non-decreasing functions starting at 0 and ending at 1. They simplify probability calculations over intervals.

Conclusion

In this lecture, we:

- Transitioned from discrete to continuous CDFs for large datasets.
- Discussed the benefits of approximating discrete CDFs with continuous functions.
- Highlighted theoretical properties of continuous CDFs.

Continuous approximations enhance efficiency and clarity in statistical modeling and probability calculations.