Lecture Summary: Marginal PMF for Multiple Random Variables

Source: Lec1.6.pdf

Key Points

- **Definition of Marginal PMF:**
  - The marginal PMF of a random variable X_1 in a set of n random variables $X_1, X_2, \ldots, X_n$ is given by summing the joint PMF over all other variables:

$$f_{X_1}(t) = \sum_{t_2, t_3, \ldots, t_n} f_{X_1, X_2, \ldots, X_n}(t, t_2, t_3, \ldots, t_n).$$

  - Marginalization simplifies complex joint distributions by focusing on individual variables.
- **Key Principle:**
  - To marginalize, keep the variable of interest and sum over the ranges of all other variables.
- **Examples:**
  - **Three Coin Tosses:**
    - Experiment: Toss a fair coin three times. Define X_1, X_2, X_3 as indicators for heads in each toss.
    - Joint PMF: $f_{X_1,X_2,X_3}(t_1, t_2, t_3) = \frac{1}{8}$ for all (t_1, t_2, t_3).
    - Marginal PMF of X_1:

$$f_{X_1}(0) = \sum_{t_2, t_3} f_{X_1,X_2,X_3}(0, t_2, t_3) = \frac{1}{2}, \quad f_{X_1}(1) = \frac{1}{2}.$$

  - **Three-Digit Lottery Numbers:**
    - Experiment: Generate a three-digit number.
    - Variables:
      - X: First digit (hundreds place).
      - Y: Modulo 2 (even/odd).
      - Z: Last digit (units place).
    - Marginal PMFs:

$$f_X(x) = \frac{1}{10},\ x \in \{0, 1, \ldots, 9\}, \quad f_Y(y) = \frac{1}{2},\ y \in \{0, 1\}.$$

  - **IPL Powerplay Overs:**
    - Experiment: Runs scored in six deliveries of the first over.
    - Random variables: $X_1, X_2, \ldots, X_6$.
    - Marginal PMF for X_1 (runs on the first ball):

$$f_{X_1}(0) = \frac{957}{1598},\ f_{X_1}(1) = \frac{429}{1598},\ f_{X_1}(4) = \frac{138}{1598}, \text{ and so on.}$$

- * Marginalization provides meaningful insights even when the joint PMF is too complex to compute.

- **Pairwise Marginalization:**
  - The joint PMF of two variables, X_1 and X_2, is computed by summing over all other variables:

$$f_{X_1,X_2}(t_1,t_2) = \sum_{t_3,\ldots,t_n} f_{X_1,X_2,\ldots,X_n}(t_1,t_2,t_3,\ldots,t_n).$$

- **General Formula for Marginal PMFs:**
  - For a subset of variables $X_{i_1}, X_{i_2}, \ldots, X_{i_k}$:

$$f_{X_{i_1},X_{i_2},\ldots,X_{i_k}}(t_{i_1},t_{i_2},\ldots,t_{i_k}) = \sum_{\text{all other variables}} f_{X_1,X_2,\ldots,X_n}(t_1,t_2,\ldots,t_n).$$

Simplified Explanation

Marginal PMFs: These focus on individual random variables or subsets by summing over the ranges of other variables.

Examples: - Coin tosses: $f_{X_1}(0) = f_{X_1}(1) = \frac{1}{2}$. - IPL runs: Marginal PMFs are derived directly from data proportions.

Why Use Marginals? They simplify analysis by reducing the complexity of joint distributions.

Conclusion

In this lecture, we:

- Defined marginal PMFs for multiple random variables.
- Explored examples from coin tosses, lottery numbers, and IPL cricket.
- Discussed practical use cases for marginalization in large datasets.

Marginal PMFs are essential tools for probabilistic analysis and simplifying complex distributions.