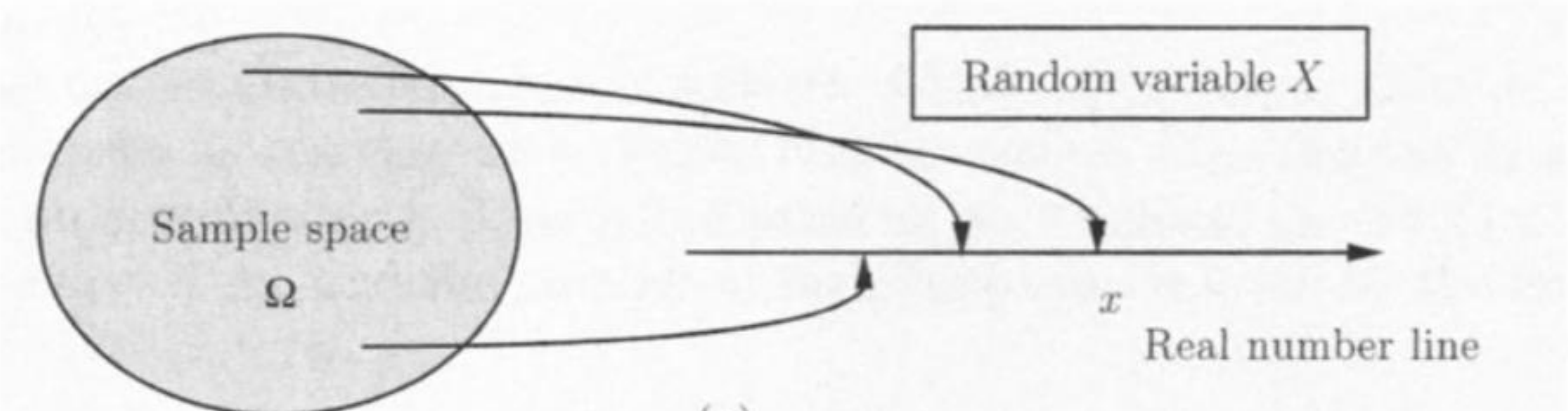
Lecture 6

Discrete random variables

Random variable

Given an experiment and the corresponding set of possible outcomes, a **random variable** associates a particular number with each outcome. We refer to this number as the numerical value or simply the value of the random variable. Mathematically, a random variable is a real-valued function of the experimental outcome.



(a)

How the Topic of "Random Variables" from Probability Theory Helps in Programming

The concept of **Random Variables** is not just abstract mathematics; it's a fundamental and powerful tool that bridges probability theory with practical programming. It provides a framework for handling uncertainty, complexity, and data-driven decision-making in software development.

Examples:

1. Modeling and Simulation (The Most Direct Application)

Often, it's too expensive, dangerous, or impractical to run a real-world experiment. Instead, we create a digital model where random variables are the core components.

- **Example: Queueing System Simulation**
  - **Problem:** Optimize the number of checkout counters in a supermarket to minimize customer wait times without having cashiers idle.
  - **Implementation:**
    - TimeBetweenCustomerArrivals is a random variable (e.g., with an exponential distribution).
    - ServiceTimePerCustomer is another random variable (e.g., with a normal distribution).
  - **In Code:** We generate values for these random variables over thousands of iterations to observe the system's behavior and find the optimal number of counters.

How the Topic of "Random Variables" from Probability Theory Helps in Programming

2. Content and Procedural Generation

- Random variables are the "fuel" for creating unpredictable, dynamic, and diverse content.
- **Example: Game Terrain Generation**
  - Terrain height can be defined by a random variable with a specific distribution (like Perlin noise, which correlates values) to create believable mountains and valleys, rather than just random noise.
- **Example: AI Opponent Behavior**
  - An NPC's decision on "where to shoot" or "which path to take" can be a probabilistic choice based on a discrete random variable. For instance, P(shot_to_head) = 0.7 and P(shot_to_body) = 0.3. This creates more complex and engaging behavior than purely random or deterministic actions.

How the Topic of "Random Variables" from Probability Theory Helps in Programming

3. Machine Learning and Data Science

- This field is fundamentally built upon the concept of random variables.
- **Data is a realization of random variables.** Every feature in a dataset (a column like height, salary, stock_price) is treated as a random variable.
- **Model Predictions are random variables.** When a model predicts the probability of a customer churning, it is essentially estimating the parameter of a Bernoulli random variable for the outcome ("will churn" / "will not churn").
- **Algorithms:** Many core algorithms (e.g., Linear Regression) have probabilistic interpretations where the model's errors (residuals) are assumed to be random variables, typically with a normal distribution.

How the Topic of "Random Variables" from Probability Theory Helps in Programming

4. Algorithms and Data Structures

- **Randomized Algorithms:** These are algorithms that use a source of randomness to achieve good performance *on average*. The running time of such an algorithm (e.g., Randomized Quicksort) is itself a random variable, and we analyze its *expected* complexity.
- **Hash Functions:** A good hash function distributes keys uniformly across buckets, mimicking a discrete uniform random variable. This understanding allows us to analyze the *average-case* time complexity of operations in hash tables (O(1)).
- **Sampling Methods:** For large datasets, it's often impossible to process everything. We take a random sample. Understanding that the sample is a collection of realizations of random variables allows us to make correct inferences about the entire population (e.g., calculating confidence intervals).

How the Topic of "Random Variables" from Probability Theory Helps in Programming

5. Testing and Debugging

- **Fuzz Testing (Fuzzing):** Instead of using predefined test cases, the program is fed random (often malformed) input data. The goal is to find edge cases and vulnerabilities. Generating this data effectively relies on using different probability distributions for the random inputs.
- **Load Testing:** Simulating user load on a server often uses distributions (e.g., Poisson) to model the time between user requests, making the test more realistic than using a constant arrival rate.

6. Computer Graphics and Image Processing

- **Global Illumination (Path Tracing):** Modern rendering algorithms that create photorealistic images are based on tracing thousands of random rays from the camera. The integral calculating the light arriving at a point is solved using Monte Carlo integration by sampling random directions over a hemisphere.

How the Topic of "Random Variables" from Probability Theory Helps in Programming

Summary for a Programmer

Understanding random variables enables you to:

- **Generate random numbers with *specific properties*, not just "any" random number.** Different problems require different distributions (uniform, normal, exponential, etc.).
- **Analyze system behavior under uncertainty.** You move beyond seeing randomness as "magic" and start to *quantitatively* assess its impact.
- **Build more complex, intelligent, and realistic systems**—from game AI to financial models.
- **Work with data correctly** in Machine Learning, understanding its probabilistic nature.

Random variable

Example. *Let X be the number of points that comes up when a dice is rolled*. The result of this experiment is the loss of one and only one face, which one is impossible to predict. Random variable X can take one of the following values $x_1 = 1, x_2 = 2, x_3 = 3, x_4 = 4, x_5 = 5, x_6 = 6$.

Example. In an experiment involving two rolls of a die the following are examples of random variables:

-The sum of the two rolls.

-The number of sixes in the two rolls.

Example. In an experiment involving a sequence of 5 tosses of a coin the number of heads in the sequence is a random variable.

We will use upper case characters to denote random variables, and lower case characters to denote real numbers such as the numerical values of a random variable.

Discrete random variable

A random variable is called **discrete** if its range (the set of values that it can take) is either finite or countably infinite.

Concepts Related to Discrete Random Variables

Starting with a probabilistic model of an experiment:

- A **discrete random variable** is a real-valued function of the outcome of the experiment that can take a finite or countably infinite number of values.
- A discrete random variable has an associated **probability mass function (PMF)**, which gives the probability of each numerical value that the random variable can take.
- A **function of a discrete random variable** defines another discrete random variable, whose PMF can be obtained from the PMF of the original random variable.

Probability mass function

For a discrete random variable X, **probability mass function (PMF)** $p_X(x)$ characterize a random variable is through the probabilities of the values that it can take:

$$\boldsymbol{p_X(x) = \mathbf{P}(\{X = x\})}$$

Example. Let the experiment consist of two independent tosses of a fair coin, and let X be the number of heads obtained. Then the PMF of X is

$$p_X(x) = \begin{cases} \frac{1}{4}, if\ x = 0\ or\ x = 2 \\ \frac{1}{2}, if\ x = 1 \\ 0, otherwise \end{cases}$$

Note that $\sum_x p_X(x) = 1$

Find the distribution of X.

Probability mass function

Calculation of the PMF of a Random Variable X

For each possible value x of X:

1. Collect all the possible outcomes that give rise to the event $\{X = x\}$.
2. Add their probabilities to obtain $p_X(x)$.

Example. A four-sided dice is tossed 2 times. Let a, b – number of points on the first and second tosses. Find distribution for random variables $\mathrm{X} = \max\{a, b\}, Y = a + b$.

The Bernoulli random variable

Consider the toss of a coin, which comes up a head with probability p, and a tail with probability 1 - p. The Bernoulli random variable takes the two values 1 and 0, depending on whether the outcome is a head or a tail:

$$X = \begin{cases} 1, if\ a\ head \\ 0, if\ a\ tail \end{cases}$$

$$p_X(k) = \begin{cases} p, if\ k = 1 \\ 1 - p, if\ k = 0 \end{cases}$$

The binomial random variable

A coin is tossed n times. At each toss, the coin comes up a head with probability p, and a tail with probability 1 - p, independent of prior tosses. Let X be the number of heads in the n-toss sequence. We refer to X as a binomial random variable with parameters n and p. The PMF of X consists of the binomial probabilities:

$$p_X(k) = P(X = k) = C_n^k p^k q^{n-k}$$

Example. There are 10% non-standard details in the lot. Four details were selected at random. Write a binomial distribution for a discrete random variable X—the number of non-standard details.