SLP1001: A Comprehensive Guide to Sleeping in Class **Spring 2026**

Lecture 1: Probability Basics

Instructor: Prof. Yitong Guo & Prof. Zhiyan Jin *Date: 19/Jan/2026*

1 Probability vs. Statistics: Two Sides of the Same Coin

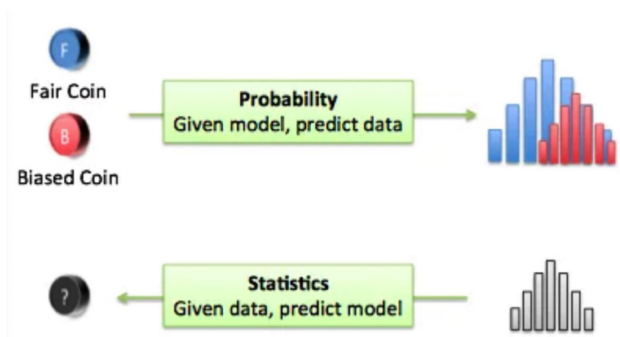
Before diving into the technical axioms, it is crucial to understand the conceptual difference between probability and statistics. While they are closely related and often used together, they represent inverse logical directions.

1.1 The Conceptual Difference

- **Probability** (Generative Perspective): We start with a known **model** (a population or a distribution) and use it to predict the characteristics of a **sample**. It answers the question: "Given that a coin is fair, what is the chance I get 10 heads in a row?"
- **Statistics** (Inference Perspective): We start with observed **data** (a sample) and use it to infer the underlying **model** (the population parameters). It answers the question: "Given that I just saw 10 heads in a row, is this coin actually fair?"

The Inverse Relationship

Probability: Population $\xrightarrow{\text{Predicts}}$ Sample
Statistics: Sample $\xrightarrow{\text{Infers}}$ Population



1.2 Illustrative Examples

***Example* 1 (The Candy Jar)** *Imagine a jar filled with 100 candies: 70 are blue and 30 are red.*

- ***The Probabilist's Task:*** *If I close my eyes and pull out 5 candies, what is the probability that at least 3 of them are red? (The "truth" of the jar is known; the outcome is uncertain).*
- ***The Statistician's Task:*** *I don't know the ratio of candies in the jar. I pull out 5 candies and see 4 red and 1 blue. What is my best estimate for the percentage of red candies in the entire jar? (The outcome is known; the "truth" of the jar is uncertain).*

***Example* 2 (Quality Control)** *A factory produces lightbulbs.*

- ***Probability:*** *If we know 1% of bulbs are defective, how many defective bulbs should we expect in a box of 1,000?*
- ***Statistics:*** *We test 1,000 bulbs and find 50 are defective. Does this mean the machine is broken, or was this just a "lucky" (or unlucky) sample?*

2 Formal Terminologies

To study probability rigorously, we need a mathematical framework to model and quantify the uncertainty in the real world. We define a **random experiment** as a repeatable procedure that can result in different outcomes, even when repeated in the same manner every time.

Definition 1 (Random Experiment) *A repeatable procedure that can result in different outcomes, even when repeated in the same manner every time.*

Definition 2 (Sample Space (S)) *The **sample space**, denoted by S, is the set of all possible outcomes of a random experiment.*

- ***Discrete:*** *Consists of a finite or countably infinite set of outcomes (e.g., rolling a die, $S = \{1, \dots, 6\}$).*
- ***Continuous:*** *Contains an interval of real numbers (e.g., the temperature outside).*

***Example* 3** *You bought 3 blind boxes which contain either Lord Voldemort or Harry Potter. Describe the set of possible outcomes.*

***Example* 4** *Which of the following are continuous?*

- *The sum of numbers on a pair of two dice.*
- *The possible sets of outcomes from flipping ten coins.*
- *The possible values of the temperature outside on any given day.*
- *The possible time that a person arrives at a restaurant.*

Definition 3 (Event) *An **event** is a subset of the sample space. We use set notation to describe events, such as $A = \{HH, HT, TH\}$ for getting at least one head in two coin tosses.*

Remark 1 *Since events are sets, we use standard set operations to find relationships between them:*

- ***Union** ($A \cup B$)**:*** *Outcomes in either A or B.*
- ***Intersection** ($A \cap B$)**:*** *Outcomes in both A and B.*
- ***Complement** (A' **or** A^c)**:*** *Outcomes in S that are not in A.*
- ***Useful Properties:*** $(E')' = E$, $A \cap B = B \cap A$, $A \cup B = B \cup A$, $(A \cup B)' = A' \cap B'$, $(A \cap B)' = A' \cup B'$

Proposition 1 (General Addition Rule) *For any two events A and B (whether joint or disjoint):*

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

REMINDER: disjoint (mutually exclusive) and independent.

Definition 4 (Probability Function, $P(\omega)$) *The probability function $P(\omega)$ gives the probability for each outcome $\omega \in S$. For an event A consisting of outcomes $\{\omega_1, \omega_2, \dots\}$, the probability of the event is the sum of the probabilities of its outcomes:*

$$P(A) = P(\omega_1) + P(\omega_2) + P(\omega_3) + \dots$$

Remark 2 *For discrete sample space, we have* ***probability mass function****, where $P(\omega)$ gives the probability for each outcome $\omega \in S$*

2.1 Axioms of Probability

Based on the frequentist view, as the number of experiments N goes to infinity, the relative frequency of an outcome (n/N) approaches its probability $P(\omega)$. All probability functions must satisfy:

1. $0 \leq P(A) \leq 1$ for any event A.
2. $P(S) = 1$, where S is the entire sample space.
3. For any two events A and B, if they are **mutually exclusive** (*in lecture note: they do not contain any common outcome*) $(A \cap B = \emptyset)$, then $P(A \cup B) = P(A) + P(B)$.

3 Independence and Conditional Concepts

Definition 5 (Independence) *Two events A and B are* ***independent*** *if and only if:*

$$P(A \cap B) = P(A) \cdot P(B)$$

Conceptually, this means knowing something about A does not tell us whether B happens.

***Example* 5 (Dependency)** *Roll a fair die twice. Let A be "the first number is 1" ($P(A) = 1/6$) and B be "the sum is > 2" ($P(B) = 35/36$). Here, $P(A \cap B) = 5/36 \neq P(A)P(B)$, so A and B are dependent.*

4 Discrete Random Variables

Definition 6 (Random Variable) *A* ***Random Variable (R.V.)*** *X is a map from the sample space S to the set of real numbers $\mathbb{R}$.*

To understand this concept, consider the following examples of how experimental outcomes are mapped to numerical values:

***Example* 6 (Counting Outcomes: Coin Tosses)** *Toss a coin three times and let the random variable X count the number of heads.*

- ***Sample Space (S):*** $\{(t,t,t), (t,t,h), (t,h,t), (h,t,t), (t,h,h), (h,t,h), (h,h,t), (h,h,h)\}$.
- ***The Mapping:*** *If the outcome $s = (t,t,h)$ occurs, then $X(s) = 1$.*

- ***Range of X:*** $\{0, 1, 2, 3\}$.

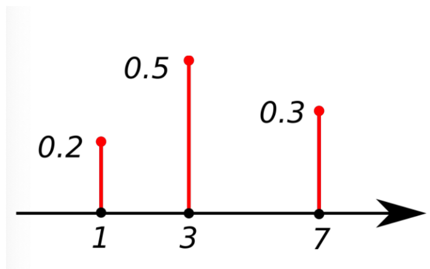
***Example 7* (Aggregating Results: Rolling Dice)** *Roll a fair die twice and let the random variable X denote the summation of the two numbers.*

- ***The Mapping:*** *An outcome like $(4, 6)$ maps to $X = 10$.*
- ***Range of X:*** $\{2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12\}$.
- ***Event Notation:*** *The event that the sum is larger than 10 can be written as $\{\omega : X(\omega) > 10\}$ or simply $\{X > 10\}$.*

4.1 Probability Mass Function (PMF)

For a discrete R.V. X, the distribution is a list of probabilities associated with its possible values. A function $f(x)$ is a PMF if:

1. $f(x_i) \geq 0$ for all x_i.
2. $\sum_{i=1}^{n} f(x_i) = 1$.
3. $f(x_i) = P(X = x_i)$.



4.2 Deriving a PMF: Step-by-Step

To obtain the PMF for a random variable (e.g., X = number of heads in three coin tosses), follow these steps:

- **Step 1:** Identify the range of X (e.g., $\{0, 1, 2, 3\}$).
- **Step 2:** Identify the specific outcomes in the sample space S that correspond to each event $\{X = x\}$.
- **Step 3:** Calculate the probability for each event, $P(X = x)$.
- **Step 4:** Construct the function $f(x)$ or a table representing the PMF.

***Example 8* (Equally Likely Outcomes)** *Suppose $S = \{a, b, c, d, e, f\}$ with equally likely outcomes ($P(\omega) = 1/6$). Let X be defined as follows:*

Outcome	*a*	*b*	*c*	*d*	*e*	*f*
X	*0*	*0*	*1.5*	*1.5*	*2*	*3*

The resulting PMF $f(x)$ is:

- $f(0) = P(\{a, b\}) = 2/6 = 1/3$.
- $f(1.5) = P(\{c, d\}) = 2/6 = 1/3$.

- $f(2) = P(\{e\}) = 1/6$.
- $f(3) = P(\{f\}) = 1/6$.

From this, we can calculate $P(0 \leq X < 2) = f(0) + f(1.5) = 2/3$.

***Example* 9 (Functional PMF)** *Verify that* $f(x) = \frac{2x+1}{25}$ *for* $x = 0, 1, 2, 3, 4$ *is a valid PMF.*

- ***Verification:*** $\sum_{x=0}^{4} \frac{2x+1}{25} = \frac{1+3+5+7+9}{25} = \frac{25}{25} = 1$.
- ***Calculation:*** $P(X \leq 1) = f(0) + f(1) = \frac{1+3}{25} = \frac{4}{25}$.

4.3 Cumulative Distribution Function (CDF)

Definition 7 (CDF) *The CDF* $F(x)$ *is defined as:*

$$F(x) = P(X \leq x) = \sum_{x_i \leq x} f(x_i)$$

Remark 3 (Properties of the CDF) *Any valid CDF must satisfy the following mathematical properties:*

1. $0 \leq F(x) \leq 1$ *for all* x.
2. $F(x)$ *is a non-decreasing function: if* $x \leq y$*, then* $F(x) \leq F(y)$.
3. $F(x)$ *is defined for all real numbers, even if the random variable only takes discrete integer values.*

4.4 Relationship Between PMF and CDF

The CDF uniquely determines the PMF of a discrete random variable. In a CDF plot, the probability $P(X = x)$ corresponds to the magnitude of the "jump" or vertical step at point x.

***Proposition* 2 (Deriving PMF from CDF)** *The probability mass function* $f(x)$ *can be derived from the CDF using the formula:*

$$f(x) = F(x) - \lim_{y \to x^-} F(y)$$

***Example* 10 (Step-Function Analysis)** *Consider a random variable* X *with the following CDF:*

$$F(x) = \begin{cases} 0, & x < -2 \\ 0.2, & -2 \leq x < 0 \\ 0.7, & 0 \leq x < 2 \\ 1, & 2 \leq x \end{cases}$$

To find the PMF, we calculate the jumps at the points of discontinuity:

- *At* $x = -2$*:* $f(-2) = F(-2) - 0 = 0.2$.
- *At* $x = 0$*:* $f(0) = F(0) - F(-0.1) = 0.7 - 0.2 = 0.5$.
- *At* $x = 2$*:* $f(2) = F(2) - F(1.9) = 1 - 0.7 = 0.3$.

4.5 Expectation and Variance

In financial applications, such as the **Markowitz Model**, we quantify "Return" as the Mean and "Risk" as the Variance.

Definition 8 (Mean/Expectation) *The "long-run average" value:*

$$E[X] = \sum_x xP(X = x) = \sum_x xf(x)$$

Definition 9 (Variance) *Quantifies how far a random variable is from its mean, on average:*

$$Var(X) = E[(X - E[X])^2] = \sum_x (x - E[X])^2 f(x)$$

A more useful computational formula is (reminder: derive):

$$Var(X) = E[X^2] - (E[X])^2$$

4.6 Useful Formulas

1. **Linearity of expectation:** $E[\sum_i c_i X_i] = \sum_i c_i E[X_i]$, where c_i's are constant. This requires **no assumption** on the dependence of X_i.
2. **Linearity of variance:** $Var(\sum_i c_i X_i) = \sum_i c_i^2 Var(X_i)$, where c_i's are constant. X_i's need to be independent with each other.
3. **Expectation of a function:** $E[g(X)] = \sum_x g(x)P(X = x) = \sum_x g(x)f(x)$.
4. **Note:** In general, $E[g(X)] \neq g(E[X])$.

***Example* 11 (The Hat Check Problem)** *n people leave hats; they are returned randomly. Let X be the number of people getting their own hat back. Define indicators $X_i = 1$ if the i^{th} person gets their hat back, else 0. Since $P(X_i = 1) = 1/n$, by linearity:*

$$E[X] = E\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} E[X_i] = n \cdot \frac{1}{n} = 1$$

Remarkably, the average number of matches is always 1, regardless of n.

4.7 Common Discrete Random Variables

Discrete random variables are used to model scenarios where we count occurrences.

4.7.1 Bernoulli Distribution: $X \sim$ Bern(p)

The simplest discrete R.V., modeling a single trial with two outcomes: Success (1) or Failure (0).

- **PMF:** $f(0) = 1 - p, \quad f(1) = p$.
- **Mean:** $E[X] = p$.
- **Variance:** $\mathrm{Var}(X) = p(1 - p)$.

4.7.2 Binomial Distribution: $X \sim \textbf{Bin}(n, p)$

Models the total number of successes in n independent Bernoulli trials.

- **PMF:** $f(x) = \binom{n}{x} p^x (1-p)^{n-x}$ for $x = 0, 1, \ldots, n$.
- **Mean:** $E[X] = np$.
- **Variance:** $\mathrm{Var}(X) = np(1-p)$.

4.7.3 Geometric Distribution: $X \sim \textbf{Geom}(p)$

Models the number of trials until the *first* success occurs.

- **PMF:** $f(x) = (1-p)^{x-1} p$ for $x = 1, 2, \ldots$.
- **Mean:** $E[X] = 1/p$.
- **Variance:** $\mathrm{Var}(X) = (1-p)/p^2$.

4.7.4 Poisson Distribution: $X \sim \textbf{Pois}(\lambda)$

Counts the number of rare events occurring in a fixed interval of time or space.

- **PMF:** $f(x) = \frac{e^{-\lambda}\lambda^x}{x!}$ for $x = 0, 1, \ldots$.
- **Property:** $E[X] = \mathrm{Var}(X) = \lambda$.

Remark 4 *The Poisson distribution can be viewed as the limiting case of the Binomial distribution. This relationship is known as the **Law of Rare Events**, which applies when the number of trials n is very large and the probability of success p is very small.*

Consider a random variable $X \sim \mathrm{Bin}(n, p)$. We are interested in the behavior of the PMF as $n \to \infty$ and $p \to 0$, while the average number of successes $\lambda = np$ remains constant.

***Proposition* 3 (Derivation of the Poisson PMF)** *Starting with the Binomial PMF:*

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$$

Substitute $p = \frac{\lambda}{n}$:

$$P(X = k) = \frac{n!}{k!\,(n-k)!} \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-k}$$

Rearranging the terms to isolate the n-dependent parts:

$$P(X = k) = \underbrace{\frac{n(n-1)\ldots(n-k+1)}{n^k}}_{(1)} \cdot \frac{\lambda^k}{k!} \cdot \underbrace{\left(1 - \frac{\lambda}{n}\right)^n}_{(2)} \cdot \underbrace{\left(1 - \frac{\lambda}{n}\right)^{-k}}_{(3)}$$

Taking the limit as $n \to \infty$:

1. $\lim_{n\to\infty} \frac{n}{n} \cdot \frac{n-1}{n} \ldots \frac{n-k+1}{n} = 1$.

2. $\lim_{n\to\infty}\left(1-\frac{\lambda}{n}\right)^n = e^{-\lambda}$ *(by the definition of the exponential constant).*
3. $\lim_{n\to\infty}\left(1-\frac{\lambda}{n}\right)^{-k} = 1^{-k} = 1.$

Combining these results, we obtain the Poisson PMF:

$$P(X=k) \approx \frac{e^{-\lambda}\lambda^k}{k!}$$

***Example* 12** *This approximation is preferred in Data Science when n is large (typically $n \geq 20$) and p is small ($p \leq 0.05$).*

- *Modeling the number of website visitors who click an ad (millions of impressions, but very low click-through rate).*
- *Computing the Poisson PMF is computationally cheaper than calculating high-order factorials and combinations required for the Binomial distribution.*

Distribution	**Breakdown Context**	**Notation**	**PMF** $f(x)$	$E[X]$	$\mathrm{Var}(X)$
Bernoulli	Whether the machine breaks down on a specific day.	$\mathrm{Bern}(p)$	$p^x(1-p)^{1-x}$	p	$p(1-p)$
Binomial	The number of breakdowns during the first n days.	$\mathrm{Bin}(n,p)$	$\binom{n}{x}p^x(1-p)^{n-x}$	np	$np(1-p)$
Geometric	The first day the machine breaks down.	$\mathrm{Geom}(p)$	$(1-p)^{x-1}p$	$1/p$	$\frac{1-p}{p^2}$
Poisson	The number of breakdowns over a long period at rate λ.	$\mathrm{Pois}(\lambda)$	$\frac{e^{-\lambda}\lambda^x}{x!}$	λ	λ

Table 1: Comparison of Discrete Distributions in a Reliability Context

5 Continuous Random Variables

A random variable is continuous if it can take any value in an interval of real numbers.

5.1 Probability Density Function (PDF)

Unlike discrete variables, $P(X=x)=0$ for any specific value x in the continuous case. We use the PDF $f(x)$ to describe probabilities over intervals:

$$P(a \leq X \leq b) = \int_a^b f(x)dx$$

Remark 5 (Properties of PDF) *PDF for a continuous distribution should satisfy the following properties:*

1. $f(x) \geq 0$ *for all* x.
2. $\int_{-\infty}^{\infty} f(x)dx = 1.$

Definition 10 (CDF of a Continuous R.V.) *The CDF, denoted by $F(x)$, represents the area under the PDF curve to the left of x:*

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(t)dt$$

Remark 6 (Fundamental Theorem of Calculus) *If the PDF $f(x)$ is continuous, then the PDF is the derivative of the CDF:*

$$f(x) = \frac{d}{dx}F(x)$$

This mirrors the discrete case where the PMF is found by the "jump" in the CDF.

***Example* 13 (Bus Stop Interarrival Time)** *Suppose X is the interarrival time (in hours) that you have to wait if you show up at a bus stop at an arbitrary time. The PDF is given by:*

$$f(x) = \begin{cases} 5e^{-5x}, & x \geq 0 \\ 0, & x < 0 \end{cases}$$

*To find the **CDF**, $F(x)$, we integrate the PDF:*

1. *For $x < 0$: $F(x) = 0$.*
2. *For $x \geq 0$:*

$$F(x) = \int_{0}^{x} 5e^{-5t}dt = \left[-e^{-5t}\right]_0^x = -e^{-5x} - (-e^0) = 1 - e^{-5x}$$

Thus, the complete CDF is:

$$F(x) = \begin{cases} 1 - e^{-5x}, & x \geq 0 \\ 0, & x < 0 \end{cases}$$

5.2 Continuous Mean and Variance

The formulas follow the discrete logic but replace summation with integration:

- **Mean:** $E[X] = \int_{-\infty}^{\infty} xf(x)dx$.
- **Variance:** $\mathrm{Var}(X) = \int_{-\infty}^{\infty}(x - E[X])^2 f(x)dx = E[X^2] - (E[X])^2$.

5.3 Common Continuous Distributions

5.3.1 Continuous Uniform Distribution: $X \sim U(a, b)$

The R.V. is equally likely to take any value within the interval $[a, b]$.

- **PDF:** $f(x) = \frac{1}{b-a}$ for $a \leq x \leq b$.
- **Mean:** $E[X] = \frac{a+b}{2}$.
- **Variance:** $\mathrm{Var}(X) = \frac{(b-a)^2}{12}$.

5.3.2 Normal (Gaussian) Distribution: $X \sim N(\mu, \sigma^2)$

The most important distribution in statistics, characterized by its bell-shaped curve.

- **PDF:** $f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$.
- **Mean:** $E[X] = \mu$; **Variance:** $\text{Var}(X) = \sigma^2$.
- **Standard Normal (Z):** A normal distribution with $\mu = 0$ and $\sigma = 1$. Any $X \sim N(\mu, \sigma^2)$ can be standardized using $Z = \frac{X-\mu}{\sigma}$.

5.3.3 Exponential Distribution: $X \sim \textbf{Exp}(\lambda)$

The Exponential distribution is often used to model the time between events in a Poisson process where events occur at a constant average rate λ.

- **PDF:** $f(x) = \lambda e^{-\lambda x}$ for $x \geq 0$.
- **Mean:** $E[X] = 1/\lambda$.
- **Variance:** $\text{Var}(X) = 1/\lambda^2$.

Understanding the CDF The CDF, $F(x) = 1 - e^{-\lambda x}$ for $x \geq 0$, represents the probability that an event occurs *at or before* time x. Its intuition is rooted in the Poisson process:

- **Connection to Poisson:** Let K be the number of events in interval $[0, x]$. The probability that the first event occurs *after* time x is equivalent to zero events occurring in that interval: $P(X > x) = P(K = 0) = \frac{e^{-\lambda x}(\lambda x)^0}{0!} = e^{-\lambda x}$.
- **Accumulation of Success:** Since $F(x) = P(X \leq x) = 1 - P(X > x)$, we obtain $F(x) = 1 - e^{-\lambda x}$.
- **Behavior at Boundaries:** At $x = 0$, $F(0) = 0$, meaning an event cannot happen in zero time. As $x \to \infty$, $F(x) \to 1$, indicating the event is guaranteed to eventually occur.

A Codes Used to Demonstrate Ideas

```python
import numpy as np
import matplotlib.pyplot as plt
from scipy.stats import binom, poisson

# Set the average rate (lambda)
lam = 5

# Define different values of n to show convergence
# We will calculate p = lam / n for each case
n_values = [10, 50, 500]
x = np.arange(0, 15) # Range of successes to plot

# Create plots
fig, axes = plt.subplots(1, 3, figsize=(18, 6), sharey=True)

for i, n in enumerate(n_values):
    p = lam / n

    # Calculate Probability Mass Functions (PMF)
```

```python
    binom_pmf = binom.pmf(x, n, p)
    poisson_pmf = poisson.pmf(x, lam)

    # Plotting side-by-side bars
    axes[i].bar(x - 0.2, binom_pmf, width=0.4, label=f'Binomial (n={n}, p={p:.2f})', alpha=0.7)
    axes[i].bar(x + 0.2, poisson_pmf, width=0.4, label=f'Poisson ($\lambda$={lam})', alpha=0.7, color='orange')

    axes[i].set_title(f'Trials (n) = {n}')
    axes[i].set_xlabel('Number of Successes')
    axes[i].set_xticks(x)
    axes[i].legend()

axes[0].set_ylabel('Probability')
plt.suptitle(f'Convergence of Binomial to Poisson Distribution ($\lambda = {lam}$)', fontsize=16)
plt.tight_layout(rect=[0, 0.03, 1, 0.95])
plt.savefig('binom_poisson_convergence.png')
```

Listing 1: Convergence of Binomial Distribution to Poison Distribution

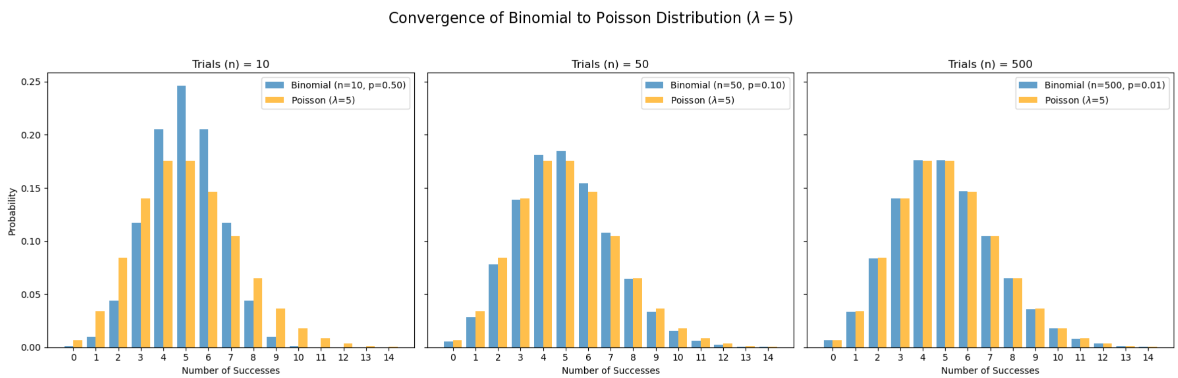


Figure 1: Key Observations: $n = 10$ (Small n): There is a visible gap between the Binomial and Poisson probabilities; $n = 50$ (Medium n): The approximation becomes quite accurate as the probability of success per trial ($p = 0.10$) decreases; $n = 500$ (Large n): The two distributions are virtually identical, demonstrating that the Poisson distribution is indeed the limiting form of the Binomial distribution when successes are rare ($p = 0.01$) but the number of opportunities (n) is vast.

```python
import numpy as np
import matplotlib.pyplot as plt

num_iterations = np.asarray([1,2,5,10,50,100]) #number of i.i.d RVs
experiment = 'dice' #valid values: 'dice', 'coins'
num_sample_space = {'dice':6,'coins':2} #max numbers represented on dice or coins
n_samples=100000 #number of samples to draw for each experiment

fig, fig_axes = plt.subplots(ncols=3, nrows=2, constrained_layout=True)

for i,N in enumerate(num_iterations):
    y = np.random.randint(low=1,high = num_sample_space[experiment]+1,size=(N,n_samples)).sum(axis=0)
    row = i//3;col=i%3
    bins=np.arange(start=min(y),stop=max(y)+2,step=1)
    fig_axes[row,col].hist(y,bins=bins,density=True, color='skyblue')
    fig_axes[row,col].set_title('N={} {}'.format(N,experiment))
```

Listing 2: Illutrative Example of Normal Distribution

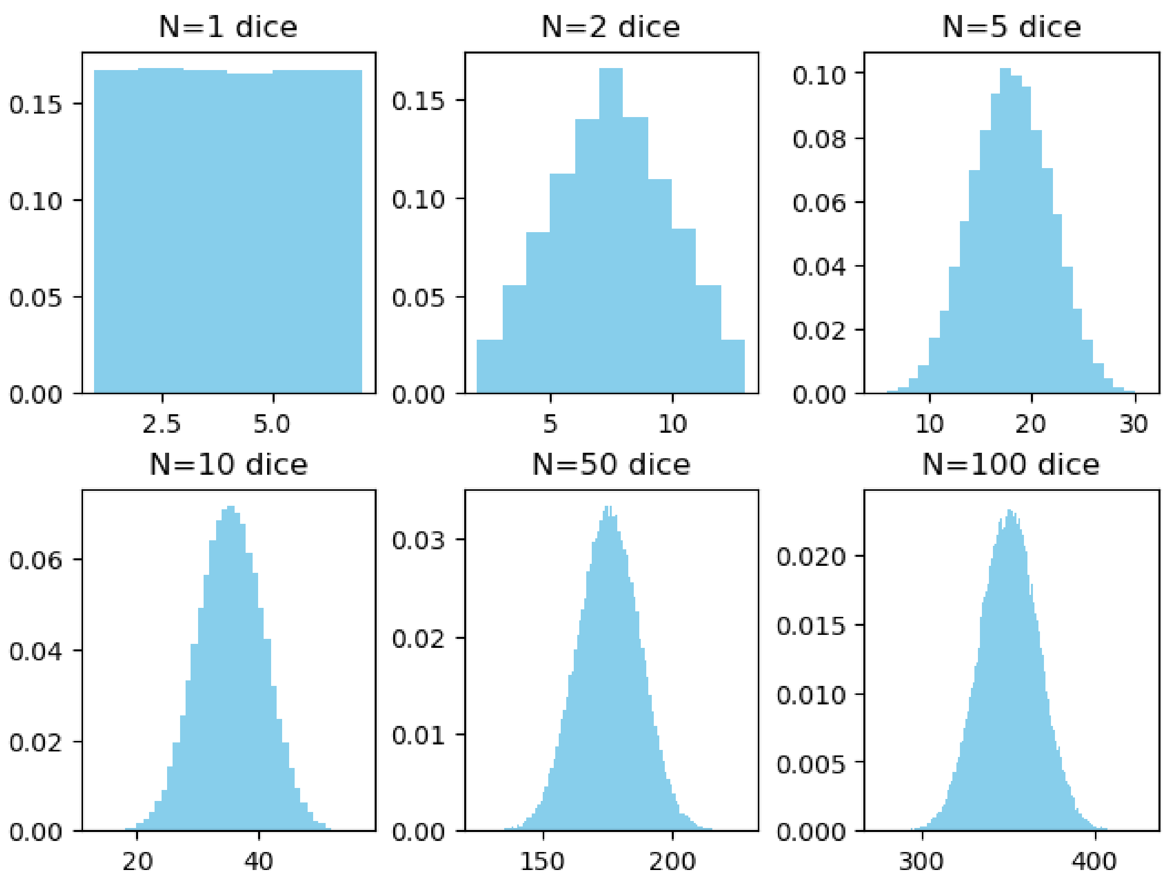


Figure 2: Convergence Example of Normal Distribution