Probability Distributions: Definitions, Formulas, Creation, and Use Cases

1. **Uniform Distribution**

* **Definition**: Every outcome in a range has an equal probability of occurring.
* **Formula**: For a continuous uniform distribution from (a) to (b), the probability density function (PDF) is: [ f(x) =

{1b−a0for a≤x≤botherwise

]

* **Creation**: Use random number generators within the desired range.
* **Why Create**: To model scenarios where each outcome is equally likely.
* **When Needed**: Random sampling, simulations where each event is equally probable.
* **Use Case**: Assigning people to teams randomly.
* **Example**: Rolling a fair die.

2. **Normal Distribution**

* **Definition**: A continuous distribution shaped like a bell curve; most values cluster around the mean.
* **Formula**: The PDF is: [ f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x - \mu)^2}{2\sigma^2}} ] where (\mu) is the mean and (\sigma) is the standard deviation.
* **Creation**: Use statistical functions in software (e.g., Excel, Python's numpy library).
* **Why Create**: To model natural phenomena and measurement errors.
* **When Needed**: Analyzing data distributions, statistical testing.
* **Use Case**: Heights of people, IQ scores.
* **Example**: Analyzing test scores in a class.
* **Dataset Example**: Heights of individuals in a population.

3. **Binomial Distribution**

* **Definition**: Discrete distribution of the number of successes in a fixed number of trials, each with the same probability of success.
* **Formula**: [ P(X = k) = \binom{n}{k} p^k (1 - p)^{n - k} ] where (n) is the number of trials, (k) is the number of successes, and (p) is the probability of success.
* **Creation**: Use software functions or binomial experiments.
* **Why Create**: To model binary outcomes (success/failure).
* **When Needed**: Quality control, medical trials.
* **Use Case**: Flipping a coin multiple times.
* **Example**: Number of heads in 10 coin flips.
* **Dataset Example**: Results of a survey where responses are "yes" or "no".

4. **Poisson Distribution**

* **Definition**: Models the number of events occurring within a fixed interval of time or space.
* **Formula**: [ P(X = k) = \frac{\lambda^k e^{-\lambda}}{k!} ] where (\lambda) is the average rate of occurrence.
* **Creation**: Use software functions or observe event counts over time/space.
* **Why Create**: To model rare events.
* **When Needed**: Analyzing occurrences over time/space.
* **Use Case**: Counting the number of emails received per hour.
* **Example**: Number of cars passing through a toll booth per hour.
* **Dataset Example**: Incidents of a rare disease in a population.

5. **Exponential Distribution**

* **Definition**: Models the time between events in a Poisson process.
* **Formula**: [ f(x|\lambda) = \lambda e^{-\lambda x} ] for (x \geq 0), where (\lambda) is the rate parameter.
* **Creation**: Use software functions or observe time intervals.
* **Why Create**: To model waiting times.
* **When Needed**: Analyzing time intervals between events.
* **Use Case**: Time until the next customer arrives at a store.
* **Example**: Time between arrivals of buses at a bus stop.
* **Dataset Example**: Time intervals between successive earthquakes.

6. **Log-Normal Distribution**

* **Definition**: If a variable is log-normally distributed, its natural logarithm is normally distributed.
* **Formula**: The PDF is: [ f(x) = \frac{1}{x\sigma\sqrt{2\pi}} e^{-\frac{(\ln x - \mu)^2}{2\sigma^2}} ] for (x > 0).
* **Creation**: Transform normally distributed data using the exponential function.
* **Why Create**: To model skewed distributions.
* **When Needed**: Modeling financial data, survival times.
* **Use Case**: Analyzing stock prices.
* **Example**: Distribution of income levels.
* **Dataset Example**: Salaries of employees in a company.

Real-World Scenario Questions with Datasets

1. **Normal Distribution Example**:
   * **Scenario**: Analyze the test scores of students in a large class to determine if they follow a normal distribution.
   * **Dataset**: Test scores of 200 students.
   * **Question**: Are the test scores normally distributed? What are the mean and standard deviation?
2. **Binomial Distribution Example**:
   * **Scenario**: Conduct a survey where each respondent can answer "yes" or "no" to a single question.
   * **Dataset**: Responses from 150 participants.
   * **Question**: What is the probability of getting exactly 100 "yes" responses if the probability of a "yes" response is 0.6?
3. **Poisson Distribution Example**:
   * **Scenario**: Monitor the number of customer service calls received per hour at a call center.
   * **Dataset**: Number of calls received each hour for a month.
   * **Question**: What is the average rate of calls per hour? What is the probability of receiving exactly 5 calls in an hour?
4. **Exponential Distribution Example**:
   * **Scenario**: Measure the time between arrivals of customers at a store.
   * **Dataset**: Time intervals (in minutes) between successive customer arrivals.
   * **Question**: What is the average time between customer arrivals? What is the probability that the next customer will arrive within 3 minutes?

Creating and Analyzing Distributions in Data Science

1. **Normal Distribution**:
   * **Python Example**:

import numpy as np  
import matplotlib.pyplot as plt  
  
data = np.random.normal(loc=50, scale=10, size=1000)  
plt.hist(data, bins=30, density=True)  
plt.title('Normal Distribution')  
plt.xlabel('Value')  
plt.ylabel('Frequency')  
plt.show()

1. **Binomial Distribution**:
   * **Python Example**:

from scipy.stats import binom  
import matplotlib.pyplot as plt  
  
n = 10  # number of trials  
p = 0.5  # probability of success  
x = np.arange(0, n+1)  
pmf = binom.pmf(x, n, p)  
  
plt.bar(x, pmf)  
plt.title('Binomial Distribution')  
plt.xlabel('Number of Successes')  
plt.ylabel('Probability')  
plt.show()

1. **Poisson Distribution**:
   * **Python Example**:

from scipy.stats import poisson  
import matplotlib.pyplot as plt  
  
mu = 3  # average rate of occurrence  
x = np.arange(0, 15)  
pmf = poisson.pmf(x, mu)  
  
plt.bar(x, pmf)  
plt.title('Poisson Distribution')  
plt.xlabel('Number of Events')  
plt.ylabel('Probability')  
plt.show()

These examples and explanations cover the basics of different types of probability distributions, their uses, how to create them, and their applications in data science and analysis.