Data Science(Numerical Python, Discrete & Continuous Data And Their Mathematical Functions)

Charles Ndung'u, Python Instructor
www.kncmap.com.

December 2023

Introduction

calculus one basics

Power Rule:

If $f(x) = x^n$, where n is a constant, then $f'(x) = nx^{(n-1)}$.

Constant Rule:

If $f(x) = c$, where c is a constant, then $f'(x) = 0$.

Sum/Difference Rule:

If $f(x) = g(x) + h(x)$, then $f'(x) = g'(x) + h'(x)$. If $f(x) = g(x) - h(x)$, then $f'(x) = g'(x) - h'(x)$.

Product Rule:

If $f(x) = g(x) \cdot h(x)$, then $f'(x) = g'(x) \cdot h(x) + g(x) \cdot h'(x)$.

Quotient Rule:

If $f(x) = \frac{g(x)}{h(x)}$, then $f'(x) = \frac{g'(x) \cdot h(x) - g(x) \cdot h'(x)}{(h(x))^2}$.

Chain Rule:

If $f(x) = g(h(x))$, then $f'(x) = g'(h(x)) \cdot h'(x)$.

Trigonometric Functions:

Sine Function:

$$\frac{d}{dx}(\sin(x)) = \cos(x)$$

Cosine Function:

$$\frac{d}{dx}(\cos(x)) = -\sin(x)$$

Tangent Function:

$$\frac{d}{dx}(\tan(x)) = \sec^2(x)$$

Exponential Functions:

$\frac{d}{dx}(e^x) = e^x$

Logarithmic Functions:

$\frac{d}{dx}(\ln(x)) = \frac{1}{x}$

Implicit Differentiation:

When expressing a relationship implicitly, differentiate both sides with respect to x: $F(x, y) = 0$, $\frac{d}{dx}[F(x, y)] = 0$

1 Examples

Power Rule:

If $f(x) = x^n$, where n is a constant, then $f'(x) = nx^{(n-1)}$.
Example:

$$f(x) = x^3, \quad f'(x) = 3x^2$$

Constant Rule:

If $f(x) = c$, where c is a constant, then $f'(x) = 0$.
Example:

$$f(x) = 5, \quad f'(x) = 0$$

Sum/Difference Rule:

If $f(x) = g(x) + h(x)$, then $f'(x) = g'(x) + h'(x)$.
Example:

$$f(x) = 2x + 3x^2, \quad f'(x) = 2 + 6x$$

If $f(x) = g(x) - h(x)$, then $f'(x) = g'(x) - h'(x)$.
Example:

$$f(x) = 4x^2 - 7x, \quad f'(x) = 8x - 7$$

Product Rule:

If $f(x) = g(x) \cdot h(x)$, then $f'(x) = g'(x) \cdot h(x) + g(x) \cdot h'(x)$.
Example:

$$f(x) = x^2 \cdot \sin(x), \quad f'(x) = 2x\sin(x) + x^2\cos(x)$$

Quotient Rule:

If $f(x) = \frac{g(x)}{h(x)}$, then $f'(x) = \frac{g'(x) \cdot h(x) - g(x) \cdot h'(x)}{(h(x))^2}$.

Example:

$$f(x) = \frac{x^2}{\cos(x)}, \quad f'(x) = \frac{2x\cos(x) + x^2\sin(x)}{\cos^2(x)}$$

Chain Rule:

If $f(x) = g(h(x))$, then $f'(x) = g'(h(x)) \cdot h'(x)$.

Example:

$$f(x) = \sin(x^2), \quad f'(x) = 2x\cos(x^2)$$

Trigonometric Functions:

Sine Function:

$$\frac{d}{dx}(\sin(x)) = \cos(x)$$

Example:

$$f(x) = \sin(2x), \quad f'(x) = 2\cos(2x)$$

Cosine Function:

$$\frac{d}{dx}(\cos(x)) = -\sin(x)$$

Example:

$$f(x) = \cos(3x), \quad f'(x) = -3\sin(3x)$$

Tangent Function:

$$\frac{d}{dx}(\tan(x)) = \sec^2(x)$$

Example:

$$f(x) = \tan(x), \quad f'(x) = \sec^2(x)$$

Exponential Functions:

$\frac{d}{dx}(e^x) = e^x$

Example:

$$f(x) = e^{2x}, \quad f'(x) = 2e^{2x}$$

Logarithmic Functions:

$\frac{d}{dx}(\ln(x)) = \frac{1}{x}$

Example:

$$f(x) = \ln(4x), \quad f'(x) = \frac{1}{x}$$

Implicit Differentiation:

When expressing a relationship implicitly, differentiate both sides with respect to x: $F(x, y) = 0$, $\frac{d}{dx}[F(x, y)] = 0$

Example:

$$xy + y^2 = 1, \quad \frac{d}{dx}(xy) + \frac{d}{dx}(y^2) = 0$$

Given the equation

$$xy + \frac{y^2}{2} = 1$$

We can find the derivative with respect to x using implicit differentiation:

$$xy + \frac{y^2}{2} = 1$$

Differentiating both sides with respect to x:

$$\frac{d}{dx}(xy) + \frac{d}{dx}\left(\frac{y^2}{2}\right) = \frac{d}{dx}(1)$$

Applying the product rule to xy:

$$y + x\frac{dy}{dx} + \frac{1}{2} \cdot 2y\frac{dy}{dx} = 0$$

Combining like terms:

$$y + x\frac{dy}{dx} + y\frac{dy}{dx} = 0$$

Factoring out $\frac{dy}{dx}$:

$$\frac{dy}{dx}(1 + x) + y = 0$$

Solving for $\frac{dy}{dx}$:

$$\frac{dy}{dx} = -\frac{y}{1 + x}$$

So, the corrected derivative of $xy + \frac{y^2}{2} = 1$ with respect to x using implicit differentiation is $\frac{dy}{dx} = -\frac{y}{1+x}$.

2 CALCULUS 2 RECARP

Integration Methods and Examples

1. Direct Integration

Example:

$$\int (3x^2 + 2x + 1)\, dx = x^3 + x^2 + x + C$$

2. Integration by Parts

Formula:

$$\int u\, dv = uv - \int v\, du$$

Example:

$$\int x\sin(x)\,dx = -x\cos(x) - \int(-\cos(x))\,dx = -x\cos(x) + \int\cos(x)\,dx$$

3. Trigonometric Integrals

List of Trigonometric Identities

Pythagorean Identity:

$$\sin^2(x) + \cos^2(x) = 1$$

Double-Angle Formulas:

$$\sin(2x) = 2\sin(x)\cos(x)$$
$$\cos(2x) = \cos^2(x) - \sin^2(x) = 2\cos^2(x) - 1 = 1 - 2\sin^2(x)$$

Half-Angle Formulas:

$$\sin\left(\frac{x}{2}\right) = \pm\sqrt{\frac{1-\cos(x)}{2}}$$
$$\cos\left(\frac{x}{2}\right) = \pm\sqrt{\frac{1+\cos(x)}{2}}$$

Sum-to-Product Formulas:

$$\sin(x) + \sin(y) = 2\sin\left(\frac{x+y}{2}\right)\cos\left(\frac{x-y}{2}\right)$$
$$\cos(x) + \cos(y) = 2\cos\left(\frac{x+y}{2}\right)\cos\left(\frac{x-y}{2}\right)$$

Product-to-Sum Formulas:

$$\sin(x)\sin(y) = \frac{1}{2}[\cos(x-y) - \cos(x+y)]$$
$$\cos(x)\cos(y) = \frac{1}{2}[\cos(x-y) + \cos(x+y)]$$

Example:

$$\int\sin^2(x)\,dx = \frac{1}{2}x - \frac{1}{4}\sin(2x) + C$$

Explanation:

This integral involves the square of the sine function, $\sin^2(x)$. The result of the integration is expressed as a sum of terms involving x and $\sin(2x)$. The constant of integration is denoted as C.

Step-by-Step Solution:

1. **Use Trigonometric Identity:**

$$\sin^2(x) = \frac{1}{2}(1 - \cos(2x))$$

2. **Integrate Term by Term:**

$$\int \sin^2(x)\,dx = \frac{1}{2}\int (1-\cos(2x))\,dx$$
$$= \frac{1}{2}\left(x - \frac{1}{2}\sin(2x)\right)$$

3. **Simplify and Add Constant of Integration:**

$$= \frac{1}{2}x - \frac{1}{4}\sin(2x) + C$$

Interpretation:

The integral represents the antiderivative of $\sin^2(x)$. The result is expressed in a form that involves both linear and sinusoidal terms. The constant of integration C accounts for the family of functions that have the same derivative.

4. Partial Fractions

Example:

$$\int \frac{1}{x^2+x}\,dx = \int \left(\frac{1}{x} - \frac{1}{x+1}\right)\,dx = \ln|x| - \ln|x+1| + C$$

5. Substitution

Example:

$$\int e^{2x}\,dx = \frac{1}{2}e^{2x} + C$$

6. Improper Integrals

Example:

$$\int_1^\infty \frac{1}{x^2}\,dx = \lim_{a\to\infty}\int_1^a \frac{1}{x^2}\,dx = 1$$

7. Double Integrals

Example:

Consider the double integral of the function $f(x,y) = x^2 + y^2$ over the region R defined by $0 \le x \le 1$ and $0 \le y \le 1$:

$$\begin{aligned}
\iint_R (x^2+y^2)\,dA &= \int_0^1 \int_0^1 (x^2+y^2)\,dy\,dx \\
&= \int_0^1 \left[\frac{1}{3}x^2 y + \frac{1}{3}y^3\right]_0^1 dx \\
&= \int_0^1 \left(\frac{1}{3}x^2 + \frac{1}{3}\right) dx \\
&= \left[\frac{1}{9}x^3 + \frac{1}{3}x\right]_0^1 \\
&= \frac{1}{9} + \frac{1}{3} = \frac{4}{9}.
\end{aligned}$$

So, the value of the double integral over the region R is $\frac{4}{9}$.

3 Discrete And Continous Data

Discrete Data:

Discrete data consists of distinct, separate values. These values are often counted in whole numbers and have clear boundaries between them. Discrete data cannot take on every possible value within a given range. Examples of discrete data include counts of objects, the number of students in a class, or the outcomes of rolling a die.

Example:

Let X represent the outcomes of rolling a fair six-sided die:

$$X = \{1, 2, 3, 4, 5, 6\}$$

Continuous Data:

Continuous data can take on any value within a given range. It is not restricted to distinct, separate values and can be measured with great precision. Continuous data is often associated with measurements such as height, weight, temperature, and time.

Example:

Let X represent a continuous range of values for temperature in Celsius:

$$X = [0, 10]$$

In this example, X represents a continuous range of values from 0 to 10, inclusive.

Discrete Probability Distributions:

1. Bernoulli Distribution:

- **Type:** Discrete
- **Description:** Models a binary outcome (success/failure) with a single trial.
- **Notation:** $X \sim \text{Bernoulli}(p)$ where p is the probability of success.

2. Binomial Distribution:

- **Type:** Discrete
- **Description:** Represents the number of successes in a fixed number of independent Bernoulli trials.
- **Notation:** $X \sim \text{Binomial}(n, p)$ where n is the number of trials and p is the probability of success in each trial.

3. Poisson Distribution:

- **Type:** Discrete
- **Description:** Models the number of events occurring in fixed intervals of time or space.
- **Notation:** $X \sim \text{Poisson}(\lambda)$ where λ is the average rate of occurrence.

Continuous Probability Distributions:

1. Normal (Gaussian) Distribution:

- **Type:** Continuous
- **Description:** Symmetric, bell-shaped distribution widely used in statistics.
- **Notation:** $X \sim \mathcal{N}(\mu, \sigma^2)$ where μ is the mean and σ^2 is the variance.

2. Uniform Distribution:

- **Type:** Continuous
- **Description:** All outcomes are equally likely over a range.
- **Notation:** $X \sim \text{Uniform}(a, b)$ where a and b are the minimum and maximum values.

3. Exponential Distribution:

- **Type:** Continuous
- **Description:** Models the time between events in a Poisson process.
- **Notation:** $X \sim \text{Exponential}(\lambda)$ where λ is the rate parameter.

4. Chi-square Distribution:

- **Type:** Continuous
- **Description:** Distribution of the sum of the squares of independent standard normal random variables.
- **Notation:** $X \sim \chi^2(k)$ where k is the degrees of freedom.

5. Gamma Distribution:

- **Type:** Continuous
- **Description:** Generalization of the exponential distribution.
- **Notation:** $X \sim \text{Gamma}(\alpha, \beta)$ where α is the shape parameter and β is the rate parameter.

4 Bernoulli Distribution

Bernoulli distribution is a discrete probability distribution, meaning it's concerned with discrete random variables. A discrete random variable is one that has a finite or countable number of possible values—the number of heads you get when tossing three coins at once, or the number of students in a class.

So, a discrete probability distribution describes the probability that each possible value of a discrete random variable will occur—for example, the probability of getting a six when rolling a die. When dealing with discrete variables, the probability of each value falls between 0 and 1, and the sum of all the probabilities is equal to 1. So, in the die example, assuming we're using a standard die, the probability of rolling a six is 0.167, or 16.7%. This is based on dividing 1 (the sum of all probabilities) by 6 (the number of possible outcomes).

That's discrete probability distribution in a nutshell. So what about Bernoulli distribution?

Bernoulli Distribution and Bernoulli Trials Explained

Bernoulli distribution applies to events that have one trial and two possible outcomes. These are known as Bernoulli trials. Think of any kind of experiment that asks a yes or no question—for example, will this coin land on heads when I flip it? Will I roll a six with this die? Will I pick an ace from this deck of cards? Will voter X vote “yes” in a political referendum? Will student Y pass their math test?

You get the idea. In Bernoulli trials, the two possible outcomes can be thought of in terms of “success” or “failure”—but these labels are not to be taken literally. In this context, “success” simply means getting a “yes” outcome (for example, rolling a six, picking an ace, and so on).

The Bernoulli distribution is, essentially, a calculation that allows you to create a model for the set of possible outcomes of a Bernoulli trial. So, whenever you have an event that has only two possible outcomes, Bernoulli distribution enables you to calculate the probability of each outcome.

Difference Between Bernoulli Distribution and Binomial Distribution

What’s the difference between Bernoulli distribution and binomial distribution? While grappling with Bernoulli distribution, you’ve likely come across another term: binomial distribution. So what’s the difference between the two, and how do they relate to one another?

In very simplistic terms, a Bernoulli distribution is a type of binomial distribution. We know that Bernoulli distribution applies to events that have one trial ($n = 1$) and two possible outcomes—for example, one coin flip (that’s the trial) and an outcome of either heads or tails. When we have more than one trial—say, we flip a coin five times—binomial distribution gives the discrete probability distribution of the number of “successes” in that sequence of independent coin flips (or trials).

So, to continue with the coin flip example: Bernoulli distribution gives you the probability of “success” (say, landing on heads) when flipping the coin just once (that’s your Bernoulli trial). If you flip the coin five times, binomial distribution will calculate the probability of success (landing on heads) across all five coin flips.

Bernoulli Distribution Example: Tossing a Coin

The coin toss example is perhaps the easiest way to explain Bernoulli distribution. Let’s say that the outcome of “heads” is a “success,” while an outcome of “tails” is a “failure.” In this instance:

The probability of a successful outcome (landing on heads) is written as p. The probability of a failure (landing on tails), written as q, is calculated as $1 - p$. With a standard coin, we know that there’s a 50/50 chance of landing on either heads or tails. So, in this case:

$p = 0.5$

$q = 1 - 0.5$

So, in our coin toss example, both p and q are 0.5. On a graph, you’d represent the probability of a failure as “0” and the probability of success as “1,” both on the y-axis.

Conditions for Bernoulli Distribution

To help you understand when and how Bernoulli distribution applies, it’s useful to consider the conditions for Bernoulli trials. An event or experiment can only be considered a Bernoulli trial (and thus be relevant for Bernoulli distribution) if it meets these criteria:

1. There are only two possible outcomes from the trial. Another way to think of this is in terms of “success” or “failure”—in other words, does your experiment ask a “yes or no” question? Think back to our previous examples, such as “Will student X pass their math test?” or “Will patient Y be cured when they take this drug?”

2. Each of the two outcomes has a fixed probability of occurring. In other words, no matter how many times you flip a coin, the probability of landing on heads is fixed. In mathematical terms, the probability of success is always p, and the probability of failure is always $1-p$.

3. Trials are entirely independent of each other. The result of one trial (say, the first coin flip) has absolutely no bearing on the outcome of any subsequent coin flips.

If a scenario meets all three of those criteria, it can be considered a Bernoulli trial. Now that we're familiar with Bernoulli distribution, let's consider where it comes into play in the broader fields of data analytics, data science, and machine learning. The PMF (Probability Mass Function) is a concept in probability theory and statistics that describes the probability distribution of a discrete random variable. It provides the probabilities of all possible outcomes of the random variable.

For a discrete random variable X, the PMF is denoted as $P(X=x)$, where x is a specific value that X can take. The PMF satisfies two properties:

Non-negativity: $P(X=x) \geq 0$ for all possible values of x.

Summation: $\sum P(X=x) = 1$ over all possible values of X.

In the context of the Bernoulli distribution, the PMF describes the probabilities of the two possible outcomes, typically denoted as 0 and 1 (representing failure and success, or vice versa). The Bernoulli PMF is often expressed as:

$$P(X=k) = \begin{cases} p & \text{if } k=1, \\ q=1-p & \text{if } k=0. \end{cases}$$

Here, p is the probability of success (e.g., getting a head in a coin toss), and q is the probability of failure (not getting a head).

In summary, the PMF provides a concise way to express the probabilities associated with different values of a discrete random variable, capturing the likelihood of each possible outcome. The probability that a discrete random variable X takes on a particular value x, that is, $P(X=x)$, is frequently denoted $p(x)$. The function $p(x)$ is typically called the probability mass function, although some authors also refer to it as the probability function, the frequency function, or probability density function. We will use the common terminology—the probability mass function—and its common abbreviation—the p.m.f.

Probability Mass Function

The probability mass function, $p(x)$, of a discrete random variable X is a function that satisfies the following properties:

1. $p(x) \geq 0$, if x belongs to the support of X. Note that if x does not belong to the support, then $p(x)=0$.

2. $\sum p(x) = 1$ over all possible values in the support of X.

3. To determine the probability associated with the event A, you just sum up the probabilities of the values in A.

Since $p(x)$ is a function, it can be presented:

- In tabular form
- In graphical form
- As a formula

Binomial vs Bernoulli Distribution: A Quick Comparison

Finally, let's dive into a side-by-side comparison of the Binomial and Bernoulli distributions to understand their unique characteristics.

Aspect	Bernoulli	Binomial
Use	Single trial	Multiple trials
Notation	p	n (trials), p (success probability)
Mean Formula	p	np
Variance Formula	$p(1-p)$	$np(1-p)$
Example	A lightbulb working or not.	Number of successes in 10 lightbulb tests.

Example 1

The prevalence of a certain disease in the general population is 10%.

If we randomly select a person from this population, we can have only two possible outcomes (diseased or healthy person). We call one of these outcomes (diseased person) success and the other (healthy person), a failure.

The probability of success (p) or diseased person is 10% or 0.1. So, the probability of failure (q) or healthy person is $1 - p = 1 - 0.1 = 0.9$.

Bernoulli Distribution Conclusion

Suppose X is a random variable that can only take on the values 1 or 0 with probabilities $P(X = 1) = p$ and $P(X = 0) = 1 - p$. Then X is said to have a Bernoulli distribution with the probability of "success" p, denoted $X \sim \text{Bernoulli}(p)$.

Examples:

- We're flipping a coin once, and our random variable is $X_1 = 1$ if the outcome is heads and $X_1 = 0$ if the outcome is tails. Then, $X_1 \sim \text{Bernoulli}(1/2)$.
- We're rolling a die, and our random variable takes on the value $X_2 = 1$ if the outcome is strictly greater than 4 and $X_2 = 0$ otherwise. Then, $X_2 \sim \text{Bernoulli}(1/3)$.

Binomial Distribution: A Simple Definition

The binomial distribution models the number of successes in a fixed number of independent experiments (trials), each with the same probability of success (p). It is denoted as $X \sim \text{Binomial}(n, p)$, where n is the number of trials and p is the probability of success in each trial.

Probability Mass Function (PMF):

The probability mass function of a binomial distribution is given by:

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

where k is the number of successes.

Complex Binomial Example:

Suppose we have a biased coin that lands on heads with a probability of 0.6. We flip this coin 10 times ($n = 10$). The probability of getting exactly 7 heads is:

$$P(X = 7) = \binom{10}{7} (0.6)^7 (0.4)^3$$

- Suppose we flip a fair coin ($p = 0.5$) 5 times ($n = 5$). The probability of getting exactly 3 heads ($k = 3$) is:

$$P(X = 3) = \binom{5}{3}(0.5)^3(0.5)^2$$

- In an experiment with 8 trials ($n = 8$) and a success probability of $p = 0.3$, the probability of observing at least 5 successes ($k \geq 5$) is:

$$P(X \geq 5) = \sum_{k=5}^{8} \binom{8}{k}(0.3)^k(0.7)^{8-k}$$

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

To find the probability of getting exactly 3 heads in 5 coin flips with a biased coin that has a probability of heads $p = 0.6$, we use the binomial probability mass function (PMF):

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

In this case, $n = 5$ (number of coin flips), $k = 3$ (number of heads), and $p = 0.6$ (probability of heads).

1. **Binomial Coefficient:**
$$\binom{5}{3} = \frac{5!}{3!(5-3)!} = \frac{5 \times 4}{2 \times 1} = 10$$

2. **Probability of Success (p^k):**
$$(0.6)^3 = 0.216$$

3. **Probability of Failure ($(1-p)^{n-k}$):**
$$(0.4)^2 = 0.16$$

4. **Multiplying the Values:**

$$P(X = 3) = \binom{5}{3} \times (0.6)^3 \times (0.4)^2 = 10 \times 0.216 \times 0.16 \approx 0.3456$$

So, the probability of getting exactly 3 heads in 5 coin flips is approximately 0.3456 or 34.56%.

5 Hypergeometric Distribution

To understand the hypergeometric distribution, which models the probability of k successes in a sample of size N drawn without replacement from a population of size K containing S successes, we can use the probability mass function (PMF):

$$P(X = k) = \frac{\binom{S}{k}\binom{K-S}{n-k}}{\binom{N}{n}}$$

where:

- N is the population size,
- K is the number of successes in the population,
- n is the sample size,
- S is the number of successes in the sample,

- k is the number of successes we are interested in.

1. **Selecting k Successes in the Sample:**

$$\binom{S}{k}$$

2. **Selecting $n - k$ Failures in the Sample:**

$$\binom{K-S}{n-k}$$

3. **Selecting n Items in the Sample:**

$$\binom{N}{n}$$

4. **Calculating the Hypergeometric Probability:**

$$P(X = k) = \frac{\binom{S}{k}\binom{K-S}{n-k}}{\binom{N}{n}}$$

This formula represents the probability of observing k successes in a sample of size n drawn without replacement from a population of size N.

Example: Hypergeometric Distribution

Suppose there are 20 marbles in a bag, and 8 of them are red. We want to draw 5 marbles from the bag without replacement and are interested in the probability of getting exactly 3 red marbles.

Parameters:

- Population size (N): 20
- Number of successes in the population (K): 8
- Sample size (n): 5
- Number of successes in the sample (S): 3

Probability Mass Function (PMF):

The hypergeometric PMF is given by:

$$P(X = k) = \frac{\binom{S}{k}\binom{K-S}{n-k}}{\binom{N}{n}}$$

Step-by-Step Explanation:

1. **Selecting k Successes in the Sample ($\binom{S}{k}$):**

$$\binom{8}{3} = 56$$

2. **Selecting $n - k$ Failures in the Sample ($\binom{K-S}{n-k}$):**

$$\binom{20-8}{5-3} = \binom{12}{2} = 66$$

3. **Selecting n Items in the Sample ($\binom{N}{n}$):**

$$\binom{20}{5} = 15504$$

4. **Calculating the Hypergeometric Probability:**

$$P(X = 3) = \frac{\binom{8}{3}\binom{12}{2}}{\binom{20}{5}} = \frac{56 \times 66}{15504} \approx 0.238$$

Result:

So, the probability of drawing exactly 3 red marbles out of 5 from the bag is approximately 0.238 or 23.8%.

The Poisson Distribution

The Poisson distribution is a discrete probability distribution that models the number of events that occur in a fixed interval of time or space.

Probability Mass Function (PMF):

The probability mass function of the Poisson distribution is given by:

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

where:

- X is the random variable representing the number of events,
- k is a non-negative integer (the number of events),
- λ is the average rate of events per unit time or space,
- e is the mathematical constant approximately equal to 2.71828,
- $k!$ is the factorial of k, the product of all positive integers up to k.

Key Features:

1. **Independence:** Events must occur independently.
2. **Constant Rate:** Events occur at a constant average rate (λ).
3. **Discreteness:** The number of events is a non-negative integer ($k = 0, 1, 2, \ldots$).
4. **Memorylessness:** The probability of future events is not influenced by past events.
5. **Rare Events:** Often used for modeling rare events.

Example:

Suppose we are observing the number of emails received per hour, and on average, we receive 4 emails per hour ($\lambda = 4$). We want to find the probability of receiving exactly 3 emails in a given hour.

$$P(X = 3) = \frac{e^{-4} \times 4^3}{3!} \approx 0.195$$

So, the probability of receiving exactly 3 emails in an hour, given an average rate of 4 emails per hour, is approximately 0.195 or 19.5%.

Example: Poisson Distribution

Suppose we are observing the number of emails received per hour, and on average, we receive 4 emails per hour ($\lambda = 4$). We want to find the probability of receiving exactly 3 emails in a given hour.

Probability Mass Function (PMF):

The Poisson PMF is given by:

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

Step-by-Step Explanation:

1. **Calculate $e^{-\lambda}$:**

$$e^{-4} \approx 0.0183$$

2. **Calculate λ^k:**

$$4^3 = 64$$

3. **Calculate $k!$:**

$$3! = 6$$

4. **Substitute into the Poisson PMF:**

$$P(X = 3) = \frac{0.0183 \times 64}{6} \approx 0.195$$

Result:

The probability of receiving exactly 3 emails in an hour, given an average rate of 4 emails per hour, is approximately 0.195 or 19.5%.

6 Continuous Probability Distributions

Normal Distribution

The normal distribution is a continuous probability distribution characterized by its bell-shaped curve. The probability density function (PDF) of the standard normal distribution is given by:

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}$$

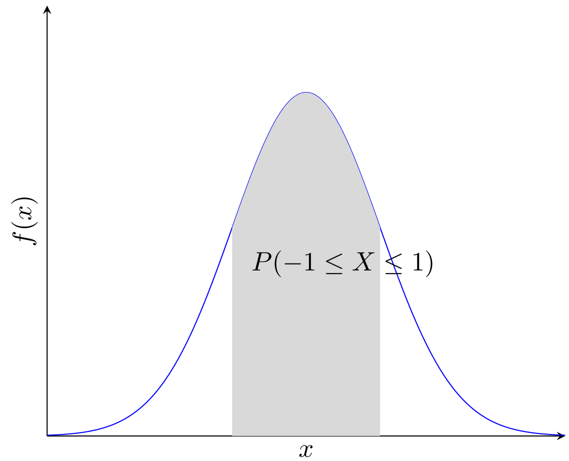
Example: Probability Calculation

Suppose we have a normally distributed variable X with a mean (μ) of 0 and a standard deviation (σ) of 1. We want to find the probability that X falls between -1 and 1.

$$P(-1 \leq X \leq 1) = \int_{-1}^{1} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}\, dx$$

Using numerical methods or tables, we find this probability to be approximately 0.6827, meaning there is a 68.27% chance that X falls within one standard deviation of the mean.

Graphical Representation:



Grading and Z-Scores

Suppose the test scores follow a normal distribution with a mean (μ) of 75 and a standard deviation (σ) of 10.

Z-Score Calculation:

The z-score for a student with a score of X is calculated using the formula:

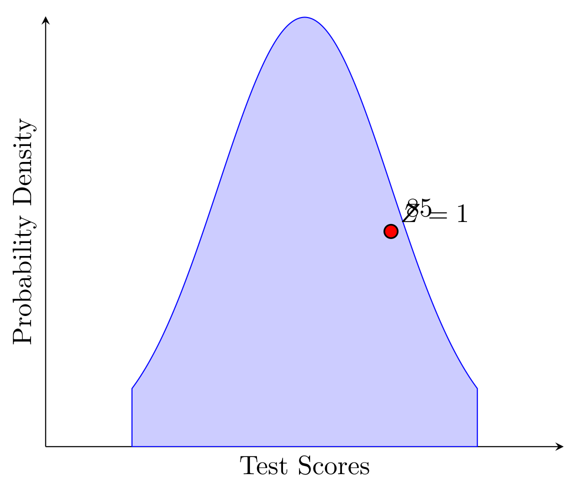
$$Z = \frac{X - \mu}{\sigma}$$

Let's say a student scored 85 on the test. We can calculate the z-score for this student:

$$Z = \frac{85 - 75}{10} = 1$$

A z-score of 1 means the student's score is 1 standard deviation above the mean.

Graphical Representation:



Importance of Mean and Standard Deviation in Normal Distribution

Shape of the Distribution:

The mean (μ) determines the central point and symmetry of the distribution. The standard deviation (σ) controls the spread or width of the distribution.

68-95-99.7 Rule:

Approximately 68% of the data falls within one standard deviation of the mean, 95% within two standard deviations, and 99.7% within three standard deviations.

Z-Scores:

Z-scores, calculated using the mean and standard deviation, provide a standardized measure of how far a data point is from the mean in terms of standard deviations.

Probability Calculations:

The mean and standard deviation are critical for calculating probabilities associated with specific ranges of values in the normal distribution's probability density function (PDF).

Central Limit Theorem:

The Central Limit Theorem states that the sum or average of a large number of independent, identically distributed random variables approaches a normal distribution. The mean and standard deviation play a key role in this theorem.

Statistical Inference:

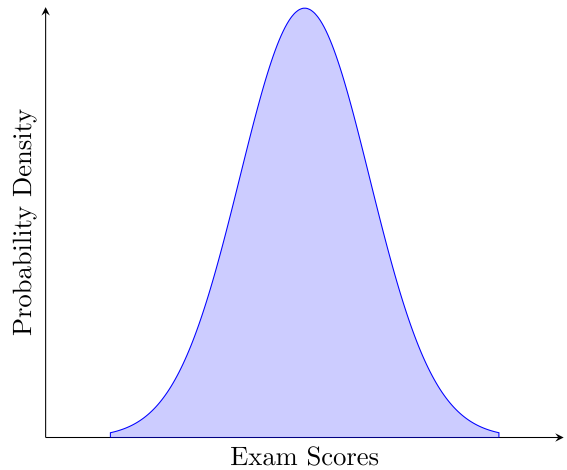
In statistical inference, the mean and standard deviation are crucial for hypothesis testing, confidence intervals, and other statistical methods.

Modeling Real-world Phenomena:

The mean and standard deviation are fundamental for modeling natural phenomena and processes that follow or can be approximated by a normal distribution.

Normal Distribution of Exam Scores

Drawing the Normal Distribution Curve:



Calculating Probability of Obtaining a Score:

Let's say we want to calculate the probability of a student obtaining a score of 80 or higher.

The probability is given by the area under the curve to the right of the score. Using the standard normal distribution table or a calculator, we find the z-score for 80:

$$Z = \frac{80 - 70}{10} = 1$$

Now, we find the probability using the standard normal distribution table:

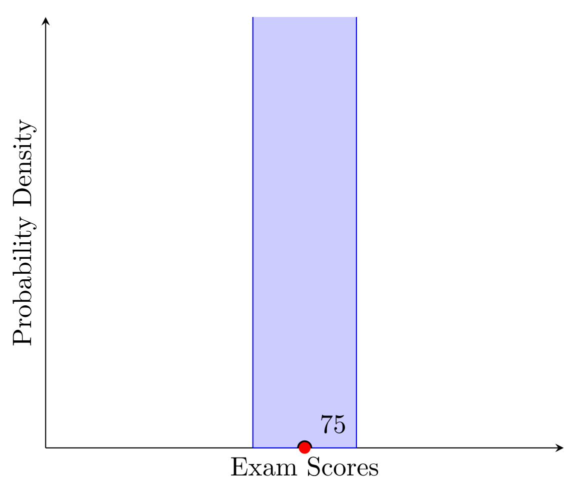
$$P(X \geq 80) = 1 - P(X < 80)$$

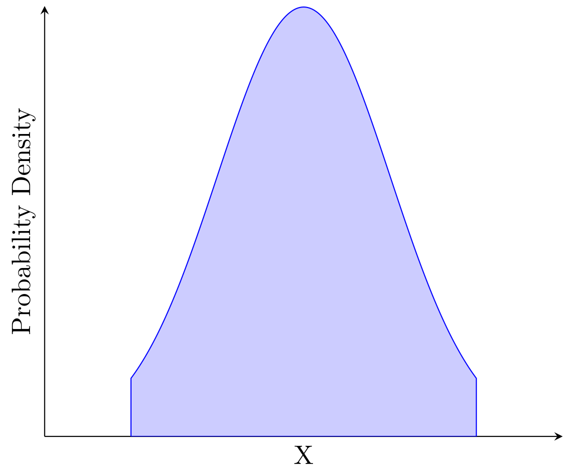
Using the table, $P(X < 80) \approx 0.8413$, so

$$P(X \geq 80) = 1 - 0.8413 \approx 0.1587$$

Therefore, the probability of a student obtaining a score of 80 or higher is approximately 15.87%.

Drawing the Normal Distribution Curve:





Calculating Probability of Scoring Between 70 and 80:

To calculate the probability of a student scoring between 70 and 80, we need to find the area under the curve between these two scores.

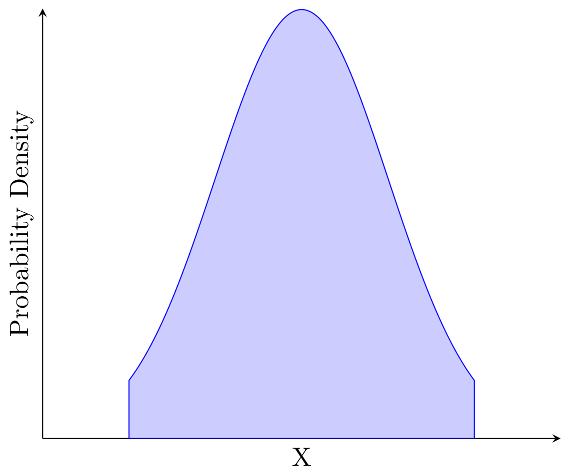
$$P(70 \leq X \leq 80) = \int_{70}^{80} \frac{1}{8\sqrt{2\pi}} e^{-\frac{(x-75)^2}{2\cdot 8^2}} \, dx$$

This integral can be evaluated using numerical methods or standard normal distribution tables.

Drawing a Normal Distribution Curve

Step-by-Step Explanation:

1. **Understand the Parameters:**
   - Mean $(\mu) = 70$
   - Standard Deviation $(\sigma) = 10$
2. **Choose a Range:**
   - Choose x-values between 50 and 90.
3. **Calculate PDF:**
   - Use the formula:
     $$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$
   - Calculate the value for each x in your range.
4. **Plot the Curve:**
   - Use a graphing tool to connect the points smoothly.
5. **Shade Areas (Optional):**
   - Shade the area under the curve to show probability.



Uniform Distribution

The uniform distribution is the simplest statistical distribution, where all outcomes are equally likely. It plays a fundamental role in statistical analysis and probability theory.

Example:

If you randomly hand a \$100 bill to any passerby on the street corner, and every passerby has an equal chance of receiving the money, that's an example of a uniform distribution. The probability is \$1 divided by the total number of outcomes (number of passersby). Favoring certain groups would violate the uniform probability concept.

A deck of cards also exhibits a uniform distribution. Drawing a spade, heart, club, or diamond has an equal chance. Similarly, when tossing a fair coin, the likelihood of getting a tail or head is the same.

Types of Uniform Distribution:

1. Discrete Uniform Distribution

In a discrete uniform distribution, outcomes are equally likely and finite. For instance, rolling a fair 6-sided die has six equally probable outcomes: 1, 2, 3, 4, 5, or 6, each with a probability of 1/6.

2. Continuous Uniform Distribution

In a continuous uniform distribution, outcomes are equally likely and infinite. A classic example is an idealized random number generator, where any real value within a specified range has an equal chance of occurring.

Example: Uniform Distribution

The data in Table 5.3.1 are 55 smiling times, in seconds, of an eight-week-old baby.

The sample mean = 11.49 and the sample standard deviation = 6.23.

We assume that the smiling times, in seconds, follow a uniform distribution between zero and 23 seconds, inclusive. This means that any smiling time from zero to and including 23 seconds is equally likely. The histogram that could be constructed from the sample is an empirical distribution that closely matches the theoretical uniform distribution.

Let X = length, in seconds, of an eight-week-old baby's smile.

The notation for the uniform distribution is $X \sim U(a, b)$ where a is the lowest value of x and b is the highest value of x.

10.4	19.6	18.8	13.9	17.8	16.8
21.6	17.9	12.5	11.1	4.9	12.8
14.8	22.8	20.0	15.9	16.3	13.4
17.1	14.5	19.0	22.8	1.3	0.7
8.9	11.9	10.9	7.3	5.9	3.7
17.9	19.2	9.8	5.8	6.9	2.6
5.8	21.7	11.8	3.4	2.1	4.5
6.3	10.7	8.9	9.4	9.4	7.6
10.0	3.3	6.7	7.8	11.6	13.8
18.6					

Table 1: Smiling times of an eight-week-old baby (in seconds)

The probability density function is $f(x) = \frac{1}{b-a}$ for $a \le x \le b$.

For this example, $X \sim U(0, 23)$ and $f(x) = \frac{1}{23-0}$ for $0 \le X \le 23$.

Formulas for the theoretical mean and standard deviation are $\mu = \frac{a+b}{2}$ and $\sigma = \frac{b-a}{\sqrt{12}}$.

For this problem, the theoretical mean and standard deviation are $\mu = \frac{0+23}{2} = 11.50$ seconds and $\sigma = \frac{23-0}{\sqrt{12}} = 6.64$ seconds.

Notice that the theoretical mean and standard deviation are close to the sample mean and standard deviation in this example.

Uniform Distribution in Statistics

In statistics, the uniform distribution is a type of probability distribution in which all possible outcomes are equally likely. A deck of cards has a uniform distribution since the probability of drawing a heart, club, diamond, or spade is equally possible. Similarly, a coin also has a uniform distribution since the probability of getting either heads or tails in a coin toss is the same. The uniform distribution can be visualized as a straight horizontal line, representing equal probabilities for each outcome.

There are two types of uniform distributions: discrete and continuous. In a discrete uniform distribution, each possible outcome is discrete, while in a continuous distribution, outcomes are continuous and infinite. In this lesson, we will explore what a uniform distribution is, the uniform distribution formula, the mean of a uniform distribution, the density of a uniform distribution, and examine some uniform distribution examples.

Uniform Probability Distribution

A continuous probability distribution called the uniform distribution is related to events that are equally possible to occur. It is defined by two parameters, x and y, where x is the minimum value and y is the maximum value. It is generally represented as $U(x, y)$.

If the probability density function (PDF) of the uniform distribution with a continuous random variable X is $f(b) = \frac{1}{y-x}$, it is denoted by $X \sim U(a, b)$ where x and y are constants such that $x < a < y$.

$$X \sim U(a, b) \tag{1}$$

Figure 1: Uniform Distribution

Theoretical Mean of Uniform Distribution

The theoretical mean (μ) of the uniform distribution is given by:

$$\mu = \frac{x+y}{2} \tag{2}$$

Standard Deviation Formula of Uniform Distribution

The standard deviation (σ) formula of the uniform distribution is given by:

$$\sigma = \frac{y - x}{\sqrt{12}} \tag{3}$$

Uniform Distribution Examples

Let's consider a couple of examples to better understand the uniform distribution.

Example 1

Suppose the average weight gained by a person over winter months is uniformly distributed and ranges from 0 to 30 lbs. Find the probability that a person will gain between 10 and 15 lbs.

$$\text{Probability} = 5 \times \frac{1}{30} = \frac{1}{6} \tag{4}$$

Example 2

Determine $P(X \leq 10)$ for the given question.

$$\text{Probability} = 10 \times \frac{1}{30} = \frac{1}{3} \tag{5}$$

Maths Introduction for Data Science

Resource file from www.kncmap.com.