Number System and Arithmetic   Algebra   Set Theory   Probability   Statistics   Geometry   Calculus

Marginal Distribution

Last Updated : 02 Dec, 2024

Marginal distribution is a fundamental concept in statistics and probability theory that refers to the distribution of a subset of variables within a larger set. Imagine you have a dataset with multiple variables; the marginal distribution focuses on just one of those variables, ignoring the others. This is useful for understanding the overall behavior of a single variable without considering its relationship to other variables.

For example, in a survey where respondents are asked about their favorite sports and their gender, the marginal distribution of sports would tell you the overall popularity of each sport regardless of gender. If 36 people like baseball, 31 like basketball, and 33 like football out of 100 respondents, these figures represent the marginal distribution of sports. Similarly, the marginal distribution of gender would tell you the overall number of males and females in the survey, without linking it to their sports preferences.

Marginal Distribution

A marginal distribution is a probability distribution of a subset of variables from a larger set, focusing on just one or a few variables while ignoring the others.

Mathematical Definition of Marginal Distribution

In a joint distribution, where multiple variables are considered simultaneously, the marginal distribution of one of those variables is obtained by summing (or integrating, in the case of continuous variables) over the possible values of the other variable(s).

This process essentially "marginalizes" the other variables, reducing the multi-dimensional distribution to a single dimension.

For a joint distribution of two random variables, X and Y, represented as P(X, Y), the **marginal distribution** of X is found by summing P(X, Y) over all possible values of Y:

- **For discrete variables:** $P(X) = \sum_Y P(X, Y)$
- **For continuous variables:** $f_X(x) = \int_{-\infty}^{\infty} f(x,y)\, dy$

This process eliminates the dependence on Y, giving the distribution of X alone.

Examples of Marginal Distribution

- **Two-Way Table Example:** Consider a two-way table showing the preferences of 100 people for different sports (baseball, basketball

people like baseball, 31 like basketball, and 33 like football, these counts represent the marginal distribution of sports.

- **Survey Data:** In survey data where respondents' favorite movie genres and their ages are recorded, the marginal distribution of movie genres can be found by summing the counts or probabilities across all age groups.

Properties of Marginal Distribution

Some of the important properties of marginal distribution are:

- Marginal distributions simplify multi-dimensional data by reducing the dimensions of the joint distribution. By summing or integrating over the other variables, the marginal distribution focuses on a single variable or a subset of variables

Normalization

- Marginal distributions, like all probability distributions, must satisfy the property of normalization. This means that the total probability across all possible values of the variable must equal 1.

  - For a discrete variable X: $\sum_x P(X = x) = 1$

- In the case of joint distributions, the marginal distribution is obtained by summing (for discrete variables) or integrating (for continuous variables) over the other variables.
    - For discrete variables X and Y: $P(X = x) = \sum_y P(X = x, Y = y)$
    - For continuous variables X and Y: $f_X(x) = \iint_{-\infty}^{\infty} f(x, y)\, dy$

Independence

- If two variables are independent, their joint distribution can be expressed as the product of their marginal distributions.
- For discrete variables X and Y: $P(X = x, Y = y) = P(X = x) \cdot P(Y = y)$
- For continuous variables X and Y: $f(x, y) = f_X(x) \cdot f_Y(y)$

Calculating Marginal Distribution

There are two common methods for different cases to calculate marginal distribution i.e.,

- Marginal Distribution from a Joint Probability Table
- Marginal Distribution in Continuous Random Variables

Marginal Distribution from a Joint Probability Table

Consider a joint distribution given by a two-way table showing the number of people who prefer different sports (baseball, basketball, football) across two genders (male, female).

Gender\Sport	Baseball	Basketball	Football	Total
Male	12	25	26	63
Female	12	13	12	37

- **Marginal Distribution of Sports:**
    - Sum the counts for each sport across all genders.
    - Baseball: 12 + 12 = 24
    - Basketball: 25 + 13 = 38
    - Football: 26 + 12 = 38

The marginal distribution of sports is: P(Baseball) = 24/100 = 0.24, P(Basketball) = 38/100 = 0.38, P(Football) = 38/100 = 0.38

- **Marginal Distribution of Gender:**
    - Sum the counts for each gender across all sports.
    - Male: 15 + 10 + 23 = 63
    - Female: 21 + 21 + 10 = 37

The marginal distribution of gender is: P(Male) = 63/100 = 0.63, P(Female) = 37/100 = 0.37

Marginal Distribution in Continuous Random Variables

Assume we have a joint probability density function (PDF) f(x, y) of two continuous variables X and Y.

Steps to Calculate Marginal Distribution:
- **Marginal Distribution of X:**
    - Integrate the joint PDF over all possible values of Y.

$$f_X(x) = \int_{-\infty}^{\infty} f(x, y) \, dy$$

- **Marginal Distribution of Y:**
    - Integrate the joint PDF over all possible values of X.

$$f_Y(y) = \int_{-\infty}^{\infty} f(x, y) \, dx$$

The key differences between marginal distribution and conditional distribution are listed in the following table:

Aspect	Marginal Distribution	Conditional Distribution
Definition	The probability distribution of a subset of variables within a larger set is obtained by summing or integrating the other variables.	The probability distribution of a variable given that another variable is known or fixed.
Purpose	To understand the overall distribution of a single variable without considering the influence of other variables.	To understand the distribution of a variable under the condition that another variable is known or fixed.
Calculation (Discrete)	Summing the joint probabilities over the other variables.	Dividing the joint probability by the marginal probability of the given variable.
Calculation (Continuous)	Integrating the joint density over the other variables.	Dividing the joint density by the marginal density of the given variable.
Normalization	Must sum or integrate to 1.	Must sum or integrate to 1 for each fixed value of the given variable.
Independence	When variables are independent, the joint distribution is the product	Not applicable directly. Independence is tested using

Aspect	Marginal Distribution	Conditional Distribution
Example (Discrete)	In a table of students' grades and study hours, the marginal distribution of grades is obtained by summing across study hours.	In the same table, the conditional distribution of grades given study hours is obtained by dividing the joint probabilities by the marginal probability of study hours.
Use Cases	Summarizing data, simplifying analysis, and initial data exploration.	Predictive modeling, understanding relationships between variables, statistical inference.

Applications of Marginal Distribution

Some of the common applications of marginal distribution are:

- **Descriptive Statistics:** Marginal distributions help in summarizing the overall characteristics of a single variable in a dataset.

- **Exploratory Data Analysis (EDA):** In EDA, marginal distributions are used to visualize and understand the distribution of individual variables before diving into more complex analyses.

- **Financial Risk Management:** Marginal distributions are used to assess the risk associated with individual financial assets. By understanding the marginal distribution of returns, risk managers can make informed decisions on portfolio allocation and risk mitigation strategies.

- **Insurance:** Marginal distributions help in understanding the risk of individual events, such as the likelihood of natural disasters or accidents, which is crucial for setting premiums and reserves.

individual features helps in detecting anomalies, scaling data, and improving model performance.

- **Bayesian Networks:** Marginal distributions are fundamental in constructing and inferring Bayesian networks, which are used for probabilistic reasoning and decision-making under uncertainty.

- **Disease Prevalence:** Marginal distributions are used to estimate the prevalence of diseases in a population. This helps in understanding the overall health status of a population and planning healthcare interventions accordingly.

- **Clinical Trials:** In clinical trials, marginal distributions of treatment outcomes are analyzed to assess the effectiveness of different treatments or interventions.

Read More,

- *Probability in Maths*
- *Conditional Probability*
- *Bayes Theorem*
- *Probability Distribution*
- *Probability Distribution Function*
- *Binomial Probability Distribution*
- *Continuous and Discrete Uniform Distribution*

| Comment | More info | Advertise with us |

Next Article

Marginal Distribution

Similar Reads

GeeksforGeeks
Sanchhaya Education Private Limited

Corporate & Communications Address:

A-143, 7th Floor, Sovereign Corporate
Tower, Sector- 136, Noida, Uttar Pradesh
(201305)

Registered Address:

K 061, Tower K, Gulshan Vivante
Apartment, Sector 137, Noida, Gautam
Buddh Nagar, Uttar Pradesh, 201305

GET IT ON Google Play   Download on the App Store

Advertise with us

Company

About Us

Legal

Privacy Policy

Careers

In Media

Explore

Job-A-Thon

Offline Classroom Program

DSA in JAVA/C++

Master System Design

Master CP

Tutorials

Python

Java

C++

PHP

GoLang

SQL

R Language

Android

DSA

Data Structures

Algorithms

DSA for Beginners

Basic DSA Problems

DSA Roadmap

DSA Interview Questions

Competitive Programming

Data Science & ML

Data Science With Python

Machine Learning

ML Maths

Data Visualisation

Pandas

NumPy

NLP

Deep Learning

Web Technologies

HTML

CSS

JavaScript

TypeScript

ReactJS

NextJS

NodeJs

Bootstrap

Tailwind CSS

Python Tutorial

Python Examples

Django Tutorial

Python Projects

Python Tkinter

Web Scraping

OpenCV Tutorial

Python Interview Question

Computer Science

GATE CS Notes

Operating Systems

Computer Network

Database Management System

Software Engineering

Digital Logic Design

Engineering Maths

DevOps

Git

AWS

Docker

Kubernetes

Azure

GCP

DevOps Roadmap

System Design

High Level Design

Low Level Design

UML Diagrams

Interview Guide

Design Patterns

OOAD

System Design Bootcamp

Interview Questions

School Subjects

Mathematics

Databases

SQL

We use cookies to ensure you have the best browsing experience on our website. By using our site, you acknowledge that you have read and understood our Cookie Policy & Privacy Policy

English Grammar

Preparation Corner

Company-Wise Recruitment Process

Aptitude Preparation

Puzzles

Company-Wise Preparation

More Tutorials

Software Development

Software Testing

Product Management

Project Management

Linux

Excel

All Cheat Sheets

Courses

IBM Certification Courses

DSA and Placements

Web Development

Data Science

Programming Languages

DevOps & Cloud

Programming Languages

C Programming with Data Structures

C++ Programming Course

Java Programming Course

Python Full Course

Clouds/Devops

DevOps Engineering

AWS Solutions Architect Certification

Salesforce Certified Administrator Course

GATE 2026

GATE CS Rank Booster

GATE DA Rank Booster

GATE CS & IT Course - 2026

GATE DA Course 2026

GATE Rank Predictor