

# 数据分析报告

## 摘要

本报告主要分析两份 Python 代码，它们通过模拟混合高斯分布数据，并使用直方图可视化来研究不同参数（如均值、标准差和权重）对分布形状的影响。报告将详细解释代码的实现机制，并分析生成的数据分布。

## 第一部分：混合高斯分布的模拟与可视化

### 代码解析

第一份代码定义了一系列函数，用于生成混合高斯分布样本并绘制其直方图。混合高斯模型是统计学中常用的一种模型，可以用来表示具有多个峰值的复杂分布。

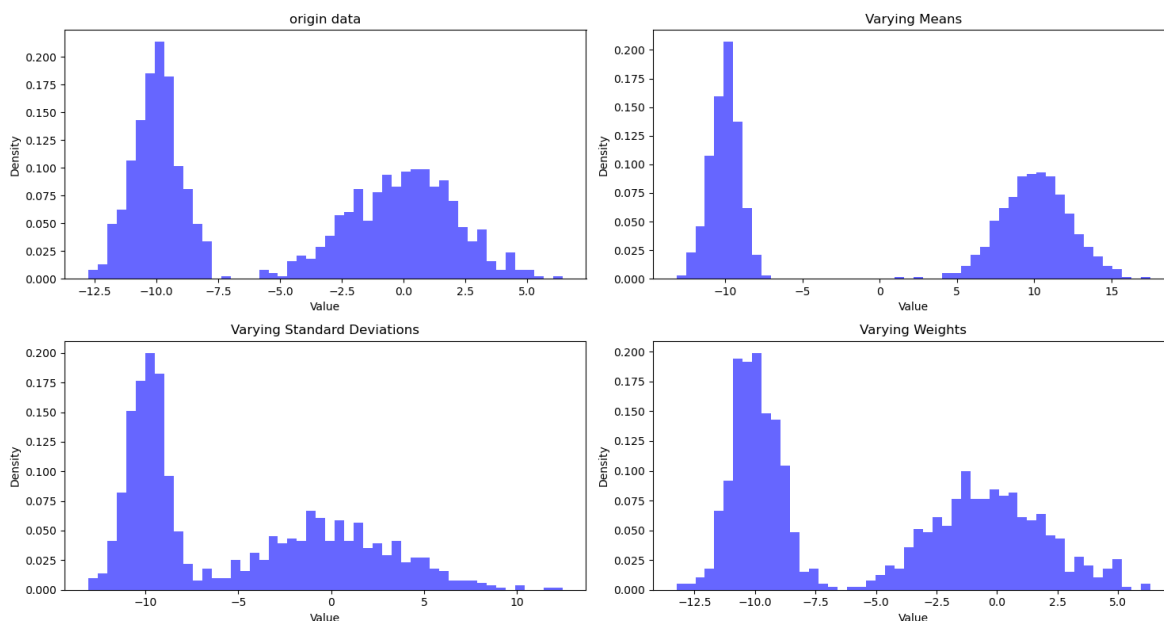
### 函数定义

- `generate_mixed_gaussian`: 此函数接受两组均值（`miu1` 和 `miu2`）、标准差（`std1` 和 `std2`）和混合参数（`yitap`），生成一个混合高斯分布的样本。
- `plot`: 该函数负责将生成的样本绘制成直方图。
- `plot_mixed_gaussian`: 它结合上述两个函数，生成样本并绘制图形。

### 参数变化

- 原始数据：设置为均值为-10和10，标准差为1和2，权重为0.8。
- 变化均值：提高第二个高斯分布的均值，观察峰值的移动。
- 变化标准差：增加第二个高斯分布的标准差，观察分布的扩散。
- 变化权重：调整混合比例，观察两个高斯分布混合后峰值的变化。

### 结果分析



上图展示了不同参数设置下，混合高斯分布的变化。可以看到，当改变均值、标准差和混合比例时，分布的峰值和形状发生了显著变化。

## 第二部分：中心极限定理的应用

### 代码解析

第二份代码演示了中心极限定理（CLT）的应用。通过多次抽取样本并计算均值，可以观察到随着样本量的增加，样本均值的分布越来越接近正态分布。

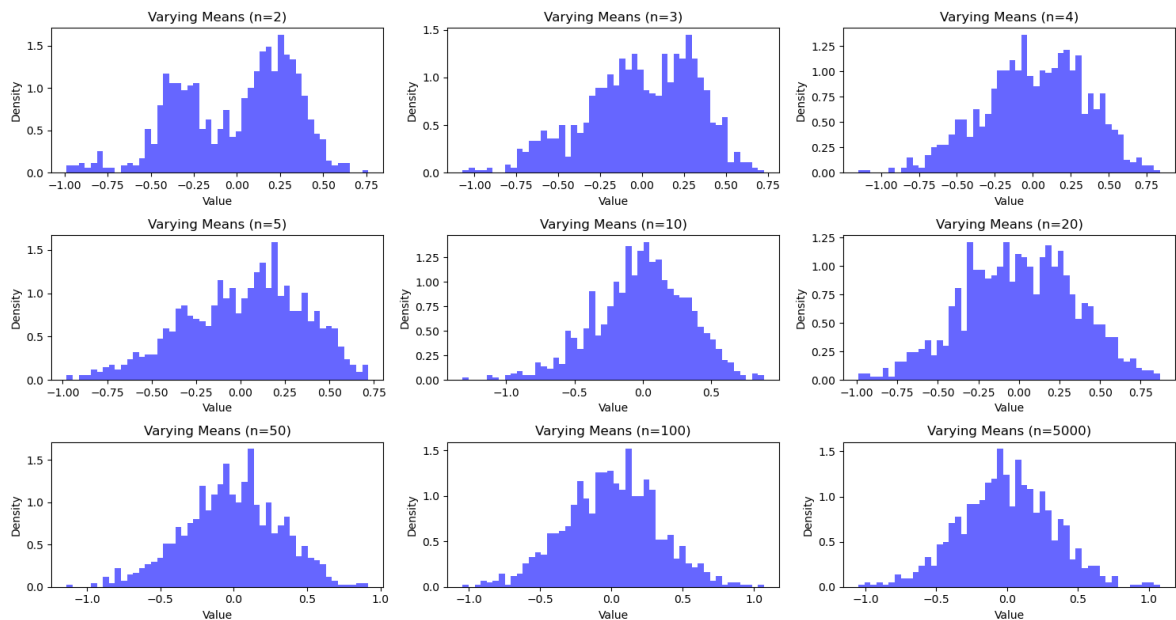
### 函数定义

- `generate_mixed_gaussian`: 与第一份代码中的同名函数功能相同。
- `problem2`: 该函数实现了 CLT 的模拟过程。它通过多次生成混合高斯分布的样本，计算样本均值，并绘制其分布的直方图。

### 参数设置

代码中预设了一个样本量列表 `nlist`，其中包含从2到5000不等的样本量。通过改变样本量，可以观察样本均值分布的变化。

### 结果分析



如上图所示，随着样本量的增加，样本均值的分布越来越趋于正态分布。这验证了中心极限定理的结论，即样本均值会趋于正态分布，无论原始数据的分布如何。

### 结论

通过两份代码的分析与结果讨论，我们可以得出以下结论：

- 混合高斯分布可以通过调整均值、标准差和混合比例来控制分布的形状。
- 中心极限定理对于样本均值的分布具有普遍适用性。

本报告详细解析了两份 Python 代码的实现，并通过图形结果深入探讨了混合高斯分布和中心极限定理的性质。这些发现对于理解复杂数据分布具有重要意义，同时也展示了统计学在数据分析中的应用。