

Preliminary Research Idea (1): Employing Reversible Dirichlet Processes for Diffusion-Model-Like Tasks

Professor Richard Yi Da Xu
HKBU and TadReamk Limited

Introduction

The Dirichlet Process can be viewed as a random discrete measure, typically denoted as:

$$G \sim \text{DP}(\alpha, H) \quad (1)$$

where H is the base measure (often regarded as a “prior” or some reference data distribution). Whenever we draw a sample from G , it can be perceived as an observation obtained from a specific realization of this random discrete measure.

Conceptual Framework

From this perspective, a natural question arises: Can we interpret H as the true data distribution, p_{data} , and interpret G as the random measure corresponding to the batch of training samples we actually possess, which were generated by p_{data} ?

The Challenge of Reversibility

A key difficulty lies in the fact that the Dirichlet Process does not possess “reversibility” in the same manner as diffusion models.

In diffusion models, we explicitly learn a reverse process to gradually restore noisy observations into clean samples, thereby obtaining a parameterized approximation of the true data distribution p_{data} .

However, in the context of a Dirichlet Process, even given a large number of samples drawn from G (or given a complete realization of G), we lack a direct “inversion” mechanism—*independent* of specific model assumptions—to recover the underlying base distribution H .

Core Research Question

This leads to a fundamental question: Is it meaningful to investigate how to “reverse” a Dirichlet Process? In other words, starting from G (observed indirectly through samples), can we infer H and model H as a parameterized data distribution?