Data Augmentation for Privacy Aware Analysis

DP Group

September 2023

Abstract

This paper serves as a reference and introduction on using the **mcDP** R package. This package provides a frame work for adapting an existing Bayesian analysis to account for privatized data. A key strength of this framework is the ability to target the exact posterior in settings where the likelihood is too complex to analytically express. The main purpose of the package is to ingest three model components and return a Markov chain targeting the posterior given the privatized data. These model components are the (1) naive posterior sampler (2) likelihood sampler and (3) privacy mechanism. Other optional components can be specified to speed up calculations. Several working examples are given.

Introduction

Blah blah introduce differential privacy.

Simple Binomial Proportion

Reference the transportation example. Discrete Gaussian distribution

Definition 0.1. Let $\mu, \sigma \in \mathbb{R}$ and $\sigma > 0$. The discrete Gaussian distribution has probability distribution

$$P(X = x) = \frac{\exp(-(x-\mu)^2/2\sigma^2)}{\sum_{y \in \mathbb{Z}} \exp(-(y-\mu)^2/2\sigma^2}$$

Linear Regression with Clamping