

RESEARCH REPORT SERIES  
(*Statistics #2021-01*)

**Direct Sampling in Bayesian Regression Models with Additive  
Disclosure Avoidance Noise**

Andrew M. Raim

Center for Statistical Research & Methodology  
Research and Methodology Directorate  
U.S. Census Bureau  
Washington, D.C. 20233

Report Issued: March 15, 2021

*Disclaimer:* This report is released to inform interested parties of research and to encourage discussion. The views expressed are those of the authors and not those of the U.S. Census Bureau.

# Direct Sampling in Bayesian Regression Models with Additive Disclosure Avoidance Noise

Andrew M. Raim

Center for Statistical Research and Methodology, U.S. Census Bureau

## Abstract

Disclosure avoidance techniques are used by agencies to prepare releases of statistics and microdata when internal data contain information considered sensitive to individual subjects. Differential privacy (DP) techniques have become popular in the literature and are finding increasing use in practical applications. One fundamental DP technique to protect sensitive data is to add noise from a selected distribution in such a way that mathematical privacy criteria are satisfied. An analyst making use of such data in a statistical model may wish to account for uncertainty introduced by the added noise. This work considers Bayesian regression models which regard the agency noise—or equivalently, the unreleased sensitive data—as augmented data. Given other random variables in the model, conditional distributions of these augmented data form weighted densities, but a method of drawing from them may not be apparent. We revisit the direct sampling method proposed by Walker et al. (JCGS 2011) and explore several customizations to address issues encountered in the basic version of the algorithm. Draws from the desired conditional distributions may be then taken reliably, largely avoiding the need for rejections or manual tuning. The customized direct sampler is used to complete the specification of a Gibbs sampler to fit a Lognormal regression model where agency noise has been added to both the outcome and some of the covariates. Demonstrations compare inference using the sensitive internal data versus the privacy-protected release.

**Keywords:** Differential privacy; Gibbs sampler; Weighted distribution; Hierarchical model; Step function approximation

## 1 Introduction

This paper revisits the direct sampling method proposed by Walker et al. (2011) and demonstrates its utility in Bayesian modeling applications which account for additive noise from a disclosure avoidance mechanism. For example, consider the regression  $\log y_i = \mathbf{x}_i^\top \boldsymbol{\beta} + \gamma_i$  for  $i = 1, \dots, n$  with  $\gamma_i \stackrel{\text{iid}}{\sim} N(0, \sigma^2)$ . The outcome  $y_i \in \mathbb{R}$  and some elements of  $\mathbf{x}_i \in \mathbb{R}^d$  are collected by an agency but considered too sensitive for public release. Therefore, the agency instead releases  $\tilde{y}_i = y_i + \xi_i^y$  and  $\tilde{\mathbf{x}}_i = \mathbf{x}_i + \xi_i^x$ , along with parameters of the distributions of  $\xi_i^y$  and  $\xi_i^x$ . Our objective is to fit the desired regression model using the released data while accounting for the added noise.

Statistical agencies are entrusted to produce summaries of data collected from the public—and microdata based on data collected from the public—while ensuring that individuals' sensitive information is not

---

Disclaimer: This article is released to inform interested parties of ongoing research and to encourage discussion of work in progress. Any views expressed are those of the author and not those of the U.S. Census Bureau.

For correspondence:

Andrew M. Raim ([andrew.raim@census.gov](mailto:andrew.raim@census.gov))

Center for Statistical Research and Methodology

U.S. Census Bureau

Washington, DC, 20233, U.S.A.

disclosed. There is a large and growing literature on disclosure avoidance (e.g. [Matthews and Harel, 2011](#)). Recently, the field of differential privacy (DP) has become increasingly popular for its ability to mathematically bound risks to unwanted disclosure in the released data ([Dwork and Roth, 2014](#)). Such guarantees can be made independently of particular methods that an attacker might use to learn the sensitive data. Differential privacy is also becoming widespread in practice; for example, the U.S. Census Bureau is evaluating use of differential privacy for public release of the data collected in the 2020 Decennial Census ([Garfinkel et al., 2018](#)).

Roughly, DP proposes mechanisms to distort sensitive data into a protected form. Proof is required to show that data in the protected form meet some criteria of interest to ensure the privacy of every individual's records. One of the fundamental DP methods is to add noise from a selected distribution such as Laplace ([Dwork and Roth, 2014](#), Section 3.3), Double Geometric (alternatively, "Two-Sided Geometric") ([Ghosh et al., 2012](#); [Kuo et al., 2018](#)), and Gaussian ([Dwork and Roth, 2014](#), Appendix A). Parameters of noise distributions must be selected appropriately to ensure that DP criteria are enforced. The privacy protection mechanism, including noise parameters, is made completely known to users of the protected data ([Gong, 2020](#)). The present paper will regard the added noise simply as draws from a known distribution. For a more complete view of DP for statisticians, [Bowen and Liu \(2020\)](#) provide a recent overview including privacy criteria and a variety of other protection mechanisms.

There are many examples of literature which develop DP mechanisms to protect sensitive data while maintaining desirable statistical properties, however, some assume particular usages which may or may not be of interest to all future data users. [Dwork and Smith \(2010\)](#) consider a release mechanism based on maximum likelihood estimation which corrects for the bias due to extra randomization, producing estimates from a predetermined parametric model. [Charest \(2011\)](#) considers Bayesian modeling under privacy protection via a particular DP mechanism for binary data; here, a Metropolis-Hastings step is used to sample the sensitive data within a Gibbs sampler. [Klein and Sinha \(2019\)](#) consider generation and analysis of multiply imputed data under noise from a Laplace mechanism, taking very large and very small values to be censored. [Evans and King \(2020+\)](#) propose a modified version of ordinary least squares regression where estimators are consistent under added DP noise. [Gong \(2019\)](#) demonstrates approximate Bayesian computation (ABC) and Monte-Carlo Expectation Maximization to analyze DP releases with additive noise. For a simple linear regression model where the response and covariate have added DP noise, [Gong \(2020\)](#) shows that failing to account for the added DP noise in an analysis distorts inferences about their association, even when the sample size is taken arbitrarily large; this work emphasizes the benefit of DP that protection mechanisms are fully disclosed and can be appropriately accounted for in the analysis. [Bernstein and Sheldon \(2019\)](#) formulate a Gibbs sampler for linear regression similar to the present setting; noise from a Laplace mechanism is drawn as augmented data using the fact that a Laplace random variable is a scale mixture of Normals.

The present work considers the method of direct sampling proposed by [Walker et al. \(2011\)](#). Direct sampling does not appear to be widely adopted in subsequent literature; one exception is by [Braun and Damien \(2016\)](#) who explore it as a scalable replacement for the inherently serial Markov chain Monte Carlo (MCMC) approach. For us, direct sampling will provide a reliable way to sample agency noise—such as  $\xi_i^y$  and  $\xi_i^x$ —as augmented data within a Gibbs sampler ([Tanner and Wong, 1987](#)). In this setting, conditional distributions frequently arise which present difficulties in the basic direct sampling approach. We provide customizations to handle these cases while avoiding manual tuning and rejections; the latter being necessary only if draws from the exact distribution are required instead of an approximation.

Noise distributions are selected by the agency releasing the data, likely driven by privacy considerations rather than modeling convenience, and therefore not within the analyst's control. Furthermore, even if the agency wishes to make analysis as convenient as possible, it may be unrealistic that all possible modeling applications will be foreseen when preparing a release. Therefore, conditional distributions involved in drawing the augmented noise steps will not necessarily reduce to more familiar forms. However, given draws of the noise, the remaining steps of the Gibbs sampler can be carried out routinely.

The remainder of the paper proceeds as follows. Section 2 reviews the direct sampling method from [Walker et al. \(2011\)](#) and presents several customizations to be used in the remainder of the paper. Section 3 illustrates the customized sampler on several important special cases which are expanded upon in subsequent

sections. Section 4 presents a simple illustration of the direct sampler in a disclosure avoidance setting, where draws from the distribution of the added noise are desired. Section 5 considers a larger regression model where the response and some covariates have added agency noise. Here the direct sampler is used as a step within a Gibbs sampler. A small simulation study demonstrates the complete sampler and compares the impact of modeling the data before and after adding noise. Finally, Section 6 concludes the paper. Appendix A provides several additional technical details.

Supporting code for this paper is provided in the repository <http://github.com/andrewraim/DirectSampling>, including materials to replicate the results. Illustrations from Sections 2 and 4 are implemented purely in R ([R Core Team, 2020](#)), while more demanding computations for Section 5 are carried out in C++ and made accessible in R via the Rcpp framework ([Eddelbuettel, 2013](#)).

## 2 Direct Sampling

Lower case letters will be used to denote both random variables and particular values; the distinction should be clear from the context. Consider drawing a random variable  $x$  with support  $\Omega \subseteq \mathbb{R}$  whose density takes the form

$$f(x) = w(x)g(x)/\psi, \quad x \in \Omega, \quad \psi = \int_{\Omega} w(x)g(x)d\nu(x), \quad (1)$$

where  $\nu(\cdot)$  is a dominating measure. The distribution may be discrete, continuous, or continuous with point masses. Density  $f$  can be recognized as a weighted distribution (e.g. [Patil and Rao, 1978](#)) with  $w : \mathbb{R} \rightarrow [0, \infty)$  a nonnegative weight function that adjusts the base density  $g$  in some prescribed way. Direct sampling augments a random variable  $u$  so that the joint distribution  $[x, u]$  is easier to draw than  $x$  itself. Let  $c = \sup_{x \in \Omega} w(x)$  and  $I(\cdot)$  be the indicator function. Assume that  $[u | x] \sim \text{Uniform}(0, w(x)/c)$ , so that

$$f(u | x) = \frac{c}{w(x)} I(0 < u < w(x)/c).$$

Let us define the event  $A_u = \{x \in \Omega : w(x) > uc\}$ . The joint density of  $[x, u]$  is then

$$f(x, u) = \frac{c}{\psi} g(x) I(x \in A_u). \quad (2)$$

From (2), the marginal density of  $u$  may be obtained as

$$p(u) = \frac{c}{\psi} P(A_u), \quad u \in [0, 1], \quad P(A_u) = \int I(x \in A_u)g(x)d\nu(x).$$

The distribution of  $[x | u]$  is then

$$f(x | u) = \frac{g(x)}{P(A_u)} I(x \in A_u). \quad (3)$$

Now  $u$  is bounded in  $[0, 1]$ , with  $A_v \supseteq A_u$  if  $v \leq u$  so that  $P(A_u)$  is monotonically nonincreasing in  $u$ . At the endpoints  $u \in \{0, 1\}$ ,  $A_0$  is equivalent to the support of  $w$  with  $P(A_0) = \int_{\Omega} I(w(x) > 0)g(x)d\nu(x)$  and  $A_1$  is an empty set with  $P(A_1) = 0$ .

A draw from  $f(x)$  can be obtained by drawing  $u \sim p(u)$  then  $x \sim f(x | u)$  in the following way. For a predefined positive integer  $N$ , compute

$$q(k/N) = \frac{P(A_{k/N})}{\sum_{\ell=0}^N P(A_{\ell/N})}, \quad k = 0, 1, \dots, N. \quad (4)$$

Sample  $k$  from the values  $0, 1, \dots, N$  with respective probabilities  $q(0/N), \dots, q(N/N)$ . Given  $k$ , sample  $u \sim \text{Beta}(k+1, N-k+1)$ . The density of  $u$  is then proportional to

$$\sum_{k=0}^N \frac{u^k (1-u)^{N-k}}{B(k+1, N-k+1)} q(k/N) \propto \sum_{k=0}^N \binom{N}{k} u^k (1-u)^{N-k} q(k/N), \quad u \in (0, 1) \quad (5)$$

where  $B(a, b) = \Gamma(a)\Gamma(b)/\Gamma(a+b)$  is the beta function. Expression (5) is an approximation to  $p(u)$  by Bernstein polynomials (e.g. Rivlin, 1981). A simple way to draw  $x$  from the truncated distribution (3) is by taking candidate draws from  $x^* \sim g(x)$ , which is likely to be more straightforward, and rejecting until  $x^* \in A_u$ . This algorithm was described by Walker et al. (2011) as a basic but general implementation of the direct sampling idea; it will be referred to as the basic direct (BD) sampler.

The BD sampler encounters some challenges in practice which may lead to unreliable results. Figure 1 illustrates several difficulties involving the unnormalized density  $P(A_u)$ . The example in Figure 1a features a support point  $u_H \approx 7 \times 10^{-14}$  where  $P(A_u) = 0$  numerically for all  $u > u_H$ . Furthermore, there is a  $u_L \in (0, u_H)$  where  $P(A_{u_L}) = P(A_0)$  for all  $u < u_L$ . The interval  $[u_L, u_H]$  is an important characteristic of the function  $P(A_u)$  which contains its descent from  $P(A_0)$  to 0. It is unhelpful to approximate  $P(A_u)$  outside of this interval, which, in Figure 1a, is extremely narrow so that  $P(A_u)$  is practically a step function. Figure 1c illustrates a second example of  $P(A_u)$  where the decrease from  $P(A_0)$  to 0 is more gradual and occurs over a larger interval within  $[0, 1]$ . A second issue is more straightforward: the basic rejection sampling method described for (3) may require a very large number of draws to accept even once when the set  $A_u$  has small probability under  $g(x)$ .

We first address the issue of improving the approximation of  $P(A_u)$ . A bisection method described in Appendix A may be used to locate  $u_L$  and  $u_H$ . It is then possible to focus the Bernstein approximation (5) to the interval  $(u_L, u_H)$  (e.g. Rivlin, 1981), or consider other functional bases from the literature. Instead, we take an approach based on step functions, which is relatively simple with low computational burden and can be controlled for accuracy. Martino et al. (2018, Section 3.6) provide background on step functions in the context of rejection sampling; this motivates our use in approximating and drawing from the distribution  $p(u)$ . Expressions for the density, CDF, and quantile function are available, and exact draws may be taken directly via the quantile function. Through appropriate placement of knot points, a step function can directly capture jumps such as those seen in Figure 1. Simple bounds on the accuracy of the approximation can be obtained in our setting, and such bounds can be improved by placing additional knot points until a desired tolerance is achieved. Finally, a step function can serve as an envelope in rejection sampling if exact draws are required.

To approximate the unnormalized  $P(A_u)$ , let  $u_0 < \dots < u_N$  be knot points with  $u_0 = u_L$  and  $u_N = u_H$  and consider the function

$$h^*(u) = P(A_{u_0}) \cdot I(0 \leq u < u_0) + \sum_{j=0}^{N-1} P(A_{u_j}) \cdot I(u_j \leq u < u_{j+1}).$$

A density is obtained using  $h(u) = h^*(u)/a$  with

$$a = \int_0^1 h^*(u) du = P(A_{u_0}) \cdot u_0 + \sum_{j=0}^{N-1} P(A_{u_j}) \cdot (u_{j+1} - u_j),$$

The corresponding cumulative distribution function (CDF) is the piecewise linear function

$$H(u) = \begin{cases} a^{-1} P(A_{u_0}) u & \text{if } 0 \leq u < u_0, \\ a^{-1} P(A_{u_0}) u_0 + a^{-1} \sum_{j=0}^{\ell-1} P(A_{u_j}) \cdot (u_{j+1} - u_j) + a^{-1} P(A_{u_\ell}) \cdot (u - u_\ell) & \text{if } u_\ell \leq u < u_{\ell+1} \end{cases}$$

for  $\ell \in \{0, \dots, N-1\}$ , with  $H(u) = 0$  if  $u \leq 0$  and  $H(u) = 1$  if  $u \geq u_N$ . The quantile function is also a piecewise linear function,

$$H^{-1}(\varphi) = u_\ell + (u_{\ell+1} - u_\ell) \frac{\varphi - H(u_\ell)}{H(u_{\ell+1}) - H(u_\ell)}, \quad \text{if } H(u_\ell) \leq \varphi < H(u_{\ell+1}) \quad (6)$$

for  $\ell \in \{0, \dots, N-1\}$ . A draw from  $h$  is now given by  $u = H^{-1}(v)$  where  $v \sim \text{Uniform}(0, 1)$ .

One way to characterize the closeness of  $h(u)$  to  $p(u)$  is by comparing probabilities computed under the two densities using total variation distance. Let  $\mathcal{R}_j$  represent the rectangle in  $\mathbb{R}^2$  whose upper-left point is  $(u_{j-1}, P(A_{u_{j-1}}))$  and lower-right point is  $(u_j, P(A_{u_j}))$ , for  $j = 1, \dots, N$ . The area of  $\mathcal{R}_j$  is  $|\mathcal{R}_j| = [P(A_{u_{j-1}}) - P(A_{u_j})](u_j - u_{j-1})$ . A proof of the following result is given in Appendix A.

**Result 2.1.** Let  $\mathcal{B}$  denote the collection of measurable subsets of  $[0, 1]$ ; then

$$\sup_{B \in \mathcal{B}} \left| \int_B h(u) du - \int_B p(u) du \right| \leq \frac{c}{\psi} \sum_{j=1}^N |\mathcal{R}_j|. \quad (7)$$

There are a number of possible choices for the knots  $u_1, \dots, u_{N-1}$ . Equally-spaced knots  $u_j = u_L + (j/N)(u_H - u_L)$  provide simplicity but can fail to capture regions of  $[u_L, u_H]$  with sudden changes in  $P(A_u)$ . Result 2.1 motivates placement of knots to ensure that no  $\mathcal{R}_j$  is too large. Namely, given  $u_0, u_1, \dots, u_k$  with associated  $\mathcal{R}_1, \dots, \mathcal{R}_k$  we consider placing a new knot at the midpoint of  $[u_{j-1}, u_j]$  which has the largest  $|\mathcal{R}_j|$ . This replaces  $\mathcal{R}_j$  with rectangles  $\mathcal{R}_j^{(1)}$  and  $\mathcal{R}_j^{(2)}$ , yielding an improvement  $|\mathcal{R}_j^{(1)}| + |\mathcal{R}_j^{(2)}| < |\mathcal{R}_j|$  in regions where  $P(A_u)$  is decreasing; otherwise,  $|\mathcal{R}_j^{(1)}| + |\mathcal{R}_j^{(2)}| = |\mathcal{R}_j|$  so that the bound in (7) is no worse. Stated as Algorithm 1, this method often provides a better selection of knots under a fixed  $N$  than equally-spaced points, at the cost of increased computation. Use of a data structure such as a priority queue (Cormen et al., 2009, Section 6.5) can help to avoid repeated sorting of  $|\mathcal{R}_1|, \dots, |\mathcal{R}_k|$  in Algorithm 1.

Figure 1 displays two examples using both equally-spaced knots and knot selection by Algorithm 1. The example in Figures 1a and 1b is effectively a step function; here the step is correctly identified by both methods, but with only  $N = 5$  knots, Algorithm 1 requires more knots to address the large rectangle near the bottom of the step. Figures 1c displays a case where equal spacing with  $N = 5$  knots fails to capture most of the change in the density, which occurs between the first two knots. Figure 1d shows that Algorithm 1 gives this area more attention.

**Remark 2.2** (Rejection sampling). The step function  $h^*(u)$  can be used to formulate a rejection sampler to take exact draws from  $p(u)$  (e.g. Martino et al., 2018). For this application, a suitable upper bound for the unnormalized ratio of densities is required. From the definition of  $\mathcal{R}_j$  we obtain

$$P(A_{u_{j-1}}) = P(A_{u_j}) + |\mathcal{R}_j|/(u_j - u_{j-1}) = \dots = \sum_{\ell=j}^N |\mathcal{R}_\ell|/(u_\ell - u_{\ell-1}),$$

noting that  $P(A_{u_N}) \equiv 0$ . Now,

$$\begin{aligned} \frac{P(A_u)}{h^*(u)} &\leq \frac{P(A_{u_{j-1}})}{P(A_{u_j})} = \frac{\sum_{\ell=j}^N |\mathcal{R}_\ell|/(u_\ell - u_{\ell-1})}{\sum_{\ell=j+1}^N |\mathcal{R}_\ell|/(u_\ell - u_{\ell-1})}, \quad \text{for } u \in [u_{j-1}, u_j) \\ &\implies \frac{P(A_u)}{h^*(u)} \leq \bigvee_{j=1}^N \frac{\sum_{\ell=j}^N |\mathcal{R}_\ell|/(u_\ell - u_{\ell-1})}{\sum_{\ell=j+1}^N |\mathcal{R}_\ell|/(u_\ell - u_{\ell-1})} =: M, \quad \text{for } u \in [0, 1] \end{aligned} \quad (8)$$

Taking  $v \sim \text{Uniform}(0, 1)$ , the candidate  $u \sim h(u)$  may be accepted as a draw from  $p(u)$  if  $v < P(A_u)/\{M \cdot h^*(u)\}$ ; otherwise, the process is repeated. Normalizing the ratio in (8) to a ratio of densities yields  $p(u)/h(u) \leq cM/(\psi a)$  so that the probability of accepting each proposed  $u$  is  $\psi a/(cM)$  and expected number of proposals before one acceptance is  $cM/(\psi a)$ . With some additional bookkeeping, a rejected  $u$  may be added to the set of knot points to increase the probability of acceptance in subsequent proposals.

The remainder of this paper utilizes the step function  $h(u)$  with Algorithm 1 to select knots so that draws from  $h(u)$  are regarded as approximate draws from  $p(u)$ .

We now turn to the second issue identified in the BD sampler, which is efficiently drawing from (3). In particular, suppose  $w$  is a unimodal weight function with mode  $x^*$ , where  $w(x)$  is nondecreasing on

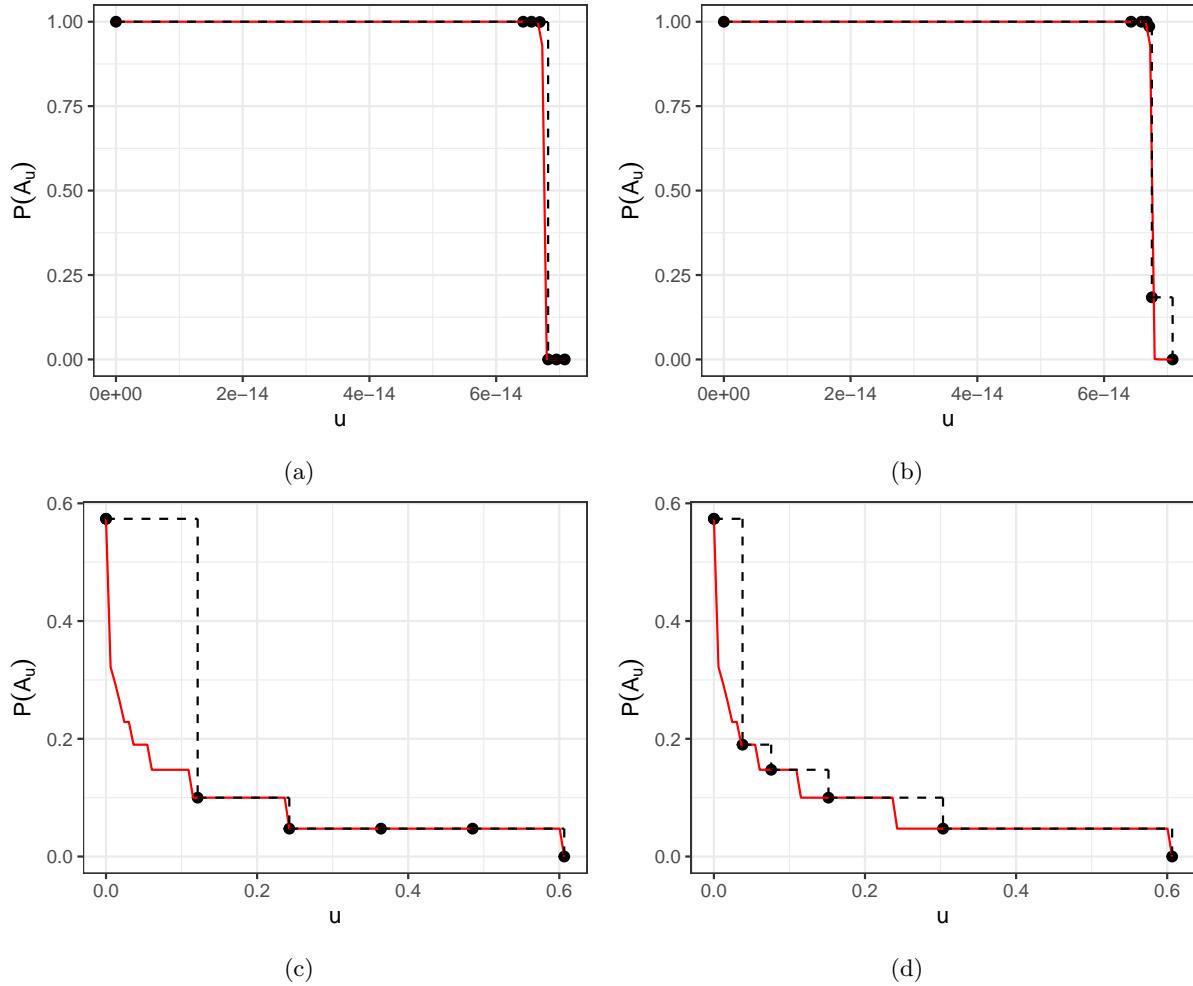


Figure 1: Two realizations of the function  $P(A_u)$ , shown as solid red lines, using a Lognormal weight function and Double Geometric base distribution as described in Section 4. (a) and (b) are based on  $z = 200000$ ,  $\mu = -2$ ,  $\sigma^2 = 8.5$ , and  $\rho = 0.01$ , while (c) and (d) are based on  $z = 2$ ,  $\mu = 0$ ,  $\sigma^2 = 1$ , and  $\rho = 0.1$ . (a) and (c) display knots selected by equal steps ( $\bullet$ ) with  $N = 5$ , while (b) and (d) use Algorithm 1. Dashed black lines highlight the upper-right portion of the rectangles  $\mathcal{R}_j$ .

**Algorithm 1** Select knots  $u_1, \dots, u_{N-1}$  to reduce  $\sum_{j=1}^N |\mathcal{R}_j|$ .

Let  $u^{(0)} = u_L$ , and  $u^{(1)} = u_H$ .

**for**  $i = 1, \dots, N - 1$  **do**

Let  $u_0 < \dots < u_i$  be sorted  $u^{(0)}, \dots, u^{(i)}$ .

Let  $|\mathcal{R}_j| = \{\Pr(A_{u_{j-1}}) - \Pr(A_{u_j})\}(u_j - u_{j-1})$  for  $j = 1, \dots, i$ .

Let  $j^* = \operatorname{argmax}_{j=1,\dots,i} |\mathcal{R}_j|$ .

Let  $u^{(i+1)} = \text{mid}(u_{j^*-1}, u_{j^*})$ .

end for

Let  $u_0 < \dots < u_N$  be sorted  $u^{(0)}, \dots, u^{(N)}$ .

**return**  $(u_0, \dots, u_N)$ .

$\{x \in \Omega : x < x^*\}$  and nonincreasing on  $\{x \in \Omega : x > x^*\}$ . Such weight functions will be of interest for the remainder of this work. Furthermore, suppose density  $g$  is associated with CDF and quantile functions

$$G(x) = \int_{-\infty}^x g(s)d\nu(s) \quad \text{and} \quad G^-(\varphi) = \inf\{x \in \Omega : G(x) \geq \varphi\},$$

respectively. In this case,  $A_u$  is an interval  $(x_1(u), x_2(u))$  whose endpoints are identified by the roots of the equation  $w(x) = cu$ . Some examples of  $w$  where  $x_1(u)$  and  $x_2(u)$  have a closed form are given in Section 3; in other cases, numerical root finding may be necessary. Now, (3) represents the base distribution  $g$  truncated to the interval  $(x_1(u), x_2(u))$ , with

$$f(x | u) = \frac{g(x) I(x_1(u) < x < x_2(u))}{G(x_2(u)-) - G(x_1(u))}$$

where  $G(x-) = \lim_{t \uparrow x} G(t)$  and  $\lceil x \rceil$  and  $\lfloor x \rfloor$  represent the ceiling and floor functions of  $x$ , respectively. The associated CDF is

$$F(x | u) = \frac{G(x) - G(x_1(u))}{G(x_2(u)-) - G(x_1(u))}, \quad x_1(u) < x < x_2(u), \quad (9)$$

with  $F(x | u) = 0$  for  $x < x_1(u)$  and  $F(x | u) = 1$  for  $x > x_2(u)$ . We may invert  $F(x | u)$  to obtain the quantile function. To reduce clutter, let us write

$$s = G(x_1(u)) \quad \text{and} \quad t = G(x_2(u)-),$$

so that the  $\varphi \in (0, 1)$  quantile of the distribution  $[x | u]$  is given by

$$\begin{aligned} F^-(\varphi | u) &= \inf\{x \in A_u : F(x | u) \geq \varphi\} \\ &= \inf\{x \in \Omega : F(x | u) \geq \varphi\} \end{aligned} \quad (10)$$

$$\begin{aligned} &= \inf\{x \in \Omega : [G(x) - s]/(t - s) \geq \varphi\} \\ &= \inf\{x \in \Omega : G(x) \geq (t - s)\varphi + s\} \\ &= G^-((t - s)\varphi + s). \end{aligned} \quad (11)$$

To justify step (10),  $x \in \Omega \setminus A_u$  implies  $x \leq x_1(u)$  or  $x \geq x_2(u)$  so that either  $F(x | u) = 0$  and does not satisfy the criteria  $F(x | u) \geq \varphi$ , or  $F(x | u) = 1$  and there is a smaller  $x' \in (x_1(u), x_2(u))$  with  $F(x' | u) \geq \varphi$ . Therefore, including  $\Omega \setminus A_u$  does not change the infimum. Now, an exact draw can be obtained via the inverse CDF method (e.g. Lange, 2010, Section 22.3) using  $x = F^-(v | u)$  with  $v \sim \text{Uniform}(0, 1)$ .

To distinguish the BD sampler from the modified version we have just described, we will refer to the latter as the customized direct (CD) sampler. To summarize, the CD sampler is feasible for weight functions  $w$  with maximum value  $c$  and where  $A_u = \{x \in \Omega : w(x) > uc\}$  is an interval with endpoints  $(x_1(u), x_2(u))$ ; ideally, all can be readily identified without much computation. The sampler is also facilitated by a base distribution  $g$  which is easy draw from and whose quantiles are easily computed.

### 3 Example Weight Functions and Base Distributions

We now demonstrate some specific distributions which are relevant to the disclosure avoidance application and/or convenient as  $w$  or  $g$  functions in the CD sampler. Several of the examples will be expanded upon in Sections 4 and 5.

**Example 3.1** (Normal weight function). Suppose  $w(x)$  is based on the density of Normal distribution  $N(\mu, \sigma^2)$ . The normalizing constant  $(\sigma\sqrt{2\pi})^{-1}$  cancels from (1), and can therefore be omitted, yielding

$$w(x | \mu, \sigma^2) = \exp \left\{ -\frac{1}{2\sigma^2}(x - \mu)^2 \right\}, \quad x \in \mathbb{R}.$$

The mode of  $w(x | \mu, \sigma^2)$  is  $x^* = \mu$  so that  $c = w(x^* | \mu, \sigma^2) = (\sigma\sqrt{2\pi})^{-1}$ . To find the endpoints of the interval  $A_u$ , consider finding the roots of the equation  $w(x) - cu = 0$ , or equivalently,

$$0 = \log w(x) - \log(cu) \iff 0 = -\frac{1}{2\sigma^2}(x - \mu)^2 - \log(cu),$$

which is a quadratic form in  $x$  with roots

$$\{x_1(u), x_2(u)\} = \mu \pm [-2\sigma^2 \log(cu)]^{1/2}.$$

**Example 3.2** (Normal weight function with transformation). Let  $h : \mathbb{R} \rightarrow \mathbb{R}$  be a bijection, and consider extending the weight function from Example 3.1 to

$$w(x | \mu, \sigma^2) = \exp \left\{ -\frac{1}{2\sigma^2}[h(x) - \mu]^2 \right\}, \quad x \in \mathbb{R}.$$

Here, the mode is  $x^* = h^{-1}(\mu)$  so that  $c = w(x^* | \mu, \sigma^2) = 1$ , and the endpoints of the interval  $A_u$  are

$$\{x_1(u), x_2(u)\} = h^{-1} \left( \mu \pm [-2\sigma^2 \log(cu)]^{1/2} \right).$$

**Example 3.3** (Lognormal weight function). Suppose  $w(x)$  is based on the density of Lognormal distribution  $\text{LN}(\mu, \sigma^2)$ ,

$$w(x | \mu, \sigma^2) = \frac{1}{x} \exp \left\{ -\frac{1}{2\sigma^2}[\log(x) - \mu]^2 \right\} \cdot \mathbf{I}(x > 0).$$

The mode is  $x^* = \exp\{\mu - \sigma^2\}$  so that  $c = w(x^* | \mu, \sigma^2) = \exp\{-(\mu - \sigma^2/2)\}$ . To find the endpoints of the interval  $A_u$ , consider finding the roots of the equation  $w(x) - cu = 0$ , or equivalently,

$$0 = \log w(x) - \log(cu) \iff 0 = -y - \frac{1}{2\sigma^2}(y - \mu)^2 - \log(cu),$$

taking  $y = \log(x)$ . This is a quadratic form in  $y$  with roots

$$y = (\mu - \sigma^2) \pm [\sigma^4 - 2\mu\sigma^2 + 2\sigma^2 \log(cu)]^{1/2};$$

therefore,

$$\{x_1(u), x_2(u)\} = \exp \left\{ (\mu - \sigma^2) \pm [\sigma^4 - 2\mu\sigma^2 + 2\sigma^2 \log(cu)]^{1/2} \right\}.$$

**Example 3.4** (Student's t weight function). Suppose  $w(x)$  is based on the unnormalized density of the t-distribution with  $\nu > 0$  degrees of freedom, center  $\mu \in \mathbb{R}$ , and scale  $\sigma > 0$ ,

$$w(x | \nu, \mu, \sigma^2) = \left[ 1 + \frac{(x - \mu)^2}{\nu\sigma^2} \right]^{-(\nu+1)/2}.$$

The mode  $x^* = \mu$  yields  $c = w(x^* | \nu, \mu, \sigma) = 1$ . After some algebra, the roots of  $w(x^* | \nu, \mu, \sigma) = cu$  are found to be

$$\{x_1(u), x_2(u)\} = \mu \pm \sigma \left[ \nu(cu)^{-2/(\nu+1)} - 1 \right]^{1/2}.$$

**Example 3.5** (Normal base distribution). Suppose  $g(x)$  is the density of  $N(0, \sigma^2)$ . The associated CDF and quantile functions are  $G(x | \sigma^2) = \Phi(x/\sigma)$  and  $G^-(\varphi | \sigma^2) = \sigma\Phi^{-1}(\varphi)$ , respectively, where  $\Phi$  is the CDF of the standard Normal distribution. These may be used to compute (9) and (11).

**Example 3.6** (Laplace base distribution). Let  $g(x | \lambda) = \frac{1}{2\lambda} e^{-|x|/\lambda}$  for  $x \in \mathbb{R}$  be the density of the Laplace distribution  $\text{Lap}(0, \lambda)$ . The associated CDF and quantile functions are

$$\begin{aligned} G(x | \lambda) &= \frac{1}{2} + \frac{1}{2} \text{sgn}(x)[1 - e^{-|x|/\lambda}], \quad \text{and} \\ G^-(\varphi | \lambda) &= -\lambda \text{sgn}(\varphi - 1/2) \log(1 - 2|\varphi - 1/2|), \end{aligned}$$

respectively, where  $\text{sgn}(x) = I(x > 0) - I(x < 0)$  denotes the sign function. Again, these may be used to compute (9) and (11), with  $G(x^- | \lambda) = G(x | \lambda)$  in the continuous case.

**Example 3.7** (Double Geometric base distribution). Let  $D\text{Geom}(\rho)$  denote the Double Geometric distribution, whose density is  $g(x | \lambda) = \frac{\rho}{2-\rho}(1-\rho)^{|x|} \cdot I(x \in \mathbb{Z})$  with  $\mathbb{Z}$  the set of integers. The Double Geometric distribution can be obtained from the transformation  $x = x_1 - x_2$  where  $x_1$  and  $x_2$  are independent  $\text{Geom}(\rho)$  random variables with density  $h(x | \rho) = \rho(1-\rho)^x \cdot I(x \in \{0, 1, 2, \dots\})$ . In some differential privacy literature such as Ghosh et al. (2012),  $D\text{Geom}(\rho)$  appears with the parameterization  $\rho = 1 - e^{-\alpha}$ . Using expressions for the geometric series, it can be shown that the CDF of  $x$  is

$$G(x | \rho) = \begin{cases} \frac{1}{2-\rho}(1-\rho)^{-\lfloor x \rfloor} & \text{if } x < 0, \\ 1 - \frac{1}{2-\rho}(1-\rho)^{\lfloor x \rfloor + 1} & \text{if } x \geq 0. \end{cases}$$

Because  $G$  is a discrete distribution,  $G(x^- | \rho) = G(\lceil x \rceil - 1 | \rho)$  is used in (9). Finally,  $G(x | \rho)$  may be inverted to obtain the quantile function for use in (11). If  $x < 0$ ,

$$\varphi \leq G(x) \equiv \frac{1}{2-\rho}(1-\rho)^{-x} \iff x \geq -\frac{\log[\varphi(2-\rho)]}{\log(1-\rho)}.$$

The smallest integer  $x$  that satisfies the last inequality is  $G^-(\varphi | \rho) = \lceil -\log(\varphi(2-\rho)) / \log(1-\rho) \rceil$ . If  $x \geq 0$ ,

$$\varphi \leq G(x) \equiv 1 - \frac{1}{2-\rho}(1-\rho)^{x+1} \iff x \geq \frac{\log[(1-\varphi)(2-\rho)]}{\log(1-\rho)} - 1$$

so that  $G^-(\varphi | \rho) = \lceil \log[(1-\varphi)(2-\rho)] / \log(1-\rho) - 1 \rceil$ .

## 4 Direct Sampler Illustration

Suppose an agency releases a sensitive random variable  $y \geq 0$  by adding Double Geometric noise with a known parameter  $\rho$ . Using the released data, we would like to model  $\log y$  by a  $N(\mu, \sigma^2)$  distribution and account for the added noise. This scenario can be described by

$$z = y + \xi, \quad \log y \sim N(\mu, \sigma^2), \quad \xi \sim D\text{Geom}(\rho),$$

with  $z$  and  $\rho$  observed. Here  $z$  may be negative, but we must have  $z > \xi$  because  $y$  is nonnegative. Here we are concerned only with the direct sampling step; Section 5 will extend to a more complete setting with multiple noisy observations,  $\mu$  specified by a regression, and a prior assumed for unknown parameters. Augmented data approaches such as Expectation Maximization (Dempster et al., 1977) and Gibbs sampling (Tanner and Wong, 1987) make use of the distribution  $[\xi | z, \mu, \sigma^2]$ . The joint distribution of  $[\xi, z | \mu, \sigma^2]$  is

$$f(z, \xi | \mu, \sigma^2) \propto w(z - \xi | \mu, \sigma^2)g(\xi),$$

where  $w$  is the Lognormal weight function in Example 3.3 and  $g$  is the Double Geometric density in Example 3.7. The distribution of  $[\xi | z, \mu, \sigma^2]$  is then

$$f(\xi | z, \mu, \sigma^2) = \frac{w(z - \xi | \mu, \sigma^2)g(\xi)}{\sum_{s=-\infty}^{\infty} w(z - s | \mu, \sigma^2)g(s)} = \frac{\frac{1}{z-\xi} \exp\left\{-\frac{1}{2\sigma^2}(\log(z - \xi) - \mu)^2\right\} (1-\rho)^{|\xi|} \cdot I(z > \xi)}{\sum_{s=-\infty}^{\lfloor z \rfloor} \frac{1}{z-s} \exp\left\{-\frac{1}{2\sigma^2}(\log(z - s) - \mu)^2\right\} (1-\rho)^{|s|}}. \quad (12)$$

Consider drawing  $\xi$  from (12) using the direct sampler described in Section 2. To compute the endpoints of the interval  $A_u$ , we must first extend Example 3.3 in the following way. Let  $w^*(\xi) = w(z - \xi | \mu, \sigma^2)$ . The mode of  $w^*$  is

$$\xi^* = \underset{\xi: z > \xi}{\operatorname{argmax}} w(z - \xi | \mu, \sigma^2) = z - \exp\{\mu - \sigma^2\},$$

which attains a value of  $c = w(z - \xi^* | \mu, \sigma^2) = \exp\{-(\mu - \sigma^2/2)\}$ . The roots of  $w^*(\xi) - cu = 0$  are also modified to

$$\{\xi_1(u), \xi_2(u)\} = z - \exp\left\{(\mu - \sigma^2) \pm [\sigma^4 - 2\mu\sigma^2 + 2\sigma^2 \log(cu)]^{1/2}\right\}.$$

We can now provide several examples of the BD and CD samplers under this setting.

**Example 4.1.** First we describe a situation where the BD sampler works well. Suppose  $z = 10$  is observed with  $\mu = 3$ ,  $\sigma^2 = 2$ , and  $\rho = 0.7$ . We draw  $n = 1,000$  samples from the distribution  $[\xi | z, \mu, \sigma^2]$  in (12) using  $N = 100$  in the Bernstein polynomial approximation of  $p(u)$ . When drawing each of the  $n$  samples from (3) by rejection sampling, we set an upper limit of 10,000 attempts. Figure 2 plots the results of the direct sampler. Figure 2a shows that the empirical distribution of the draws is close to the density, which indicates that the sampler is functioning correctly. Figure 2b plots the number of attempts required for rejection sampling in draws  $i = 1, \dots, n$ . Most draws required a small number of attempts before an acceptance, but several required more than 450 attempts.

**Example 4.2.** Here is a situation where the BD sampler runs into difficulties. Again, consider Example 4.1, but suppose  $z = 244,388$  is observed with  $\mu = 3.5$ , and  $\sigma^2 = 8.5$ . With  $N = 100$ , the BD sampler is unable to accept any draws of the rejection sampler within the limit of 10,000 attempts. Upon closer inspection, we find that  $u_L \approx 1.8179 \times 10^{-8}$  and  $u_H \approx 1.8188 \times 10^{-8}$ . The BD sampler is not designed to make use of this fact, basing its approximation of  $p(u)$  on  $q(0/N) = 1$  and  $q(1/N) = \dots = q(N/N) = 0$ , and failing to capture the region of interest  $[u_L, u_H]$ .

**Example 4.3.** The CD sampler overcomes the issues in Example 4.2. Figure 3a plots the empirical distribution of  $n = 50,000$  draws with  $N = 100$ , showing close agreement with the target density  $f(x)$ . The remaining panels of Figure 3 display several variations of the distribution along with the empirical distribution of draws from the CD sampler. Figure 3b reduces  $\rho$  to 0.01, which results in the agency noise being increased by several orders of magnitude. Figures 3c and 3d take  $z = -10$ ,  $\mu = 3.5$ ,  $\sigma^2 = 8.5$ , with respective values of  $\rho \in \{0.7, 0.01\}$ . Here the shape of the distribution is heavily influenced by the lower bound  $z$ , and is more obviously not a symmetric distribution about zero than Examples 4.1 and 4.2.

In this setting, the weight function  $w$  corresponds to the distribution of the outcome and the base distribution  $g$  corresponds to the added agency noise. Using the examples in Section 3, we could also formulate a CD sampler with outcomes following Normal, transformed Normal, or shifted/scaled Student's t distributions, or with Normal or Laplace agency distributions assumed for agency noise.

## 5 Regression Model Applications

We now build on Section 4 and show the direct sampler—specifically, the CD sampler—in the context of a more complete Bayesian regression model. The following notation will be used. The  $d$ -variate Normal distribution with density function  $f_N(\mathbf{x} | \boldsymbol{\mu}, \boldsymbol{\Sigma}) \propto \exp\{-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu})\}$  is denoted by  $\mathbf{x} \sim N_d(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ . The Inverse Gamma distribution with density  $f_{IG}(x) \propto x^{-a-1} e^{-b/x} \cdot I(x > 0)$  is denoted by  $x \sim IG(a, b)$ . The Laplace distribution with density  $f_{Lap}(x | \mu, \lambda) \propto e^{-|x-\mu|/\lambda}$  is denoted by  $x \sim Lap(\mu, \lambda)$ . The Double Geometric distribution with density  $f_{DGeom}(x | \rho) \propto (1 - \rho)^{|x|} I(x \in \mathbb{Z})$  is denoted  $x \sim DGeom(\rho)$ . The  $n \times n$  identity matrix is  $\mathbf{I}_n$ . The vector of outcomes will be denoted by  $\mathbf{y} = (y_1, \dots, y_n)$  and the matrix of covariates (without agency noise) by  $\mathbf{X} = (x_{ij})$ , with row  $\mathbf{x}_{i \cdot} = (x_{i1}, \dots, x_{id})$  corresponding to the  $i$ th

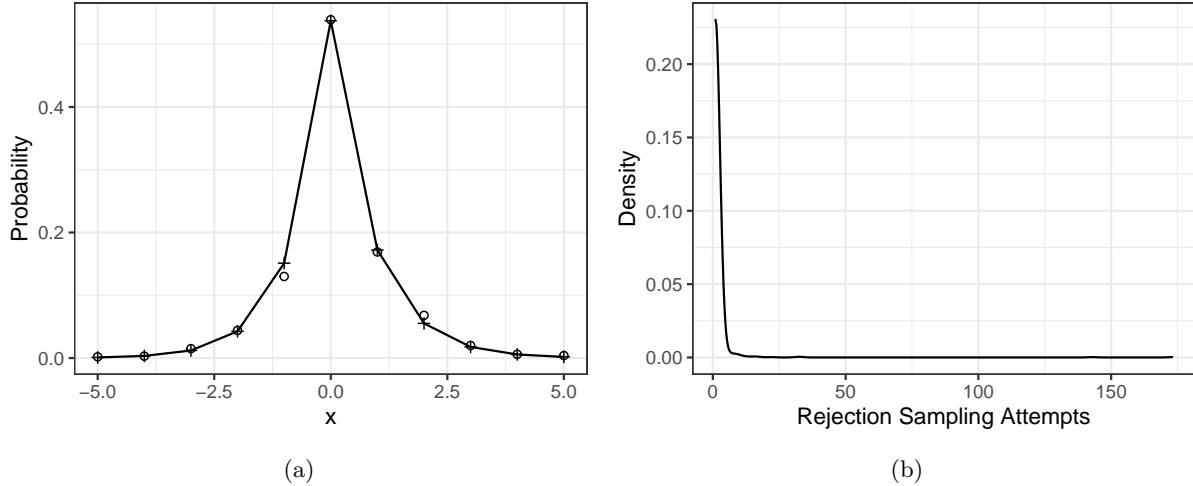


Figure 2: Outcome of BD sampler for Example 4.1. (a) empirical distribution of the sample (○) along with the density value (+). Connecting lines between density values are shown only for visual reference—the target distribution is defined on  $\mathbb{Z}$ . (b) density of the number of rejections needed to acquire each of the  $n = 1,000$  samples.

subject and column  $\mathbf{x}_{\cdot j} = (x_{1j}, \dots, x_{nj})$  corresponding to the  $j$ th variable. Similar conventions will be used for other matrix and vector-valued quantities.

Consider a non-negative response  $y_i$  and covariates  $x_{ij} \in \mathbb{R}$  for subjects  $i = 1, \dots, n$  and variables  $j = 1, \dots, d$ . Suppose  $\mathbf{y}$  and  $\mathbf{x}_{\cdot j}$  are considered sensitive for  $j = 1, \dots, d_1$  and  $d_1 \leq d$ ; these data are procured by an agency which release noise-infused versions to protect privacy. Namely,  $\tilde{y}_i = y_i + \xi_i^y$  and  $\tilde{x}_{ij} = x_{ij} + \xi_{ij}^x$  are released using  $\xi_i^y \stackrel{\text{ind}}{\sim} \text{DGeom}(\rho_i^y)$  and  $\xi_{ij}^x \stackrel{\text{ind}}{\sim} \text{Lap}(0, \lambda_{ij}^x)$ .<sup>1</sup> Parameters  $\rho_i^y$  and  $\lambda_{ij}^x$  are provided so that the noise-generating mechanism is fully known to the analyst. Covariates labeled  $j = d_1 + 1, \dots, d$  are not considered sensitive and available to the analyst without noise. To study a Lognormal regression relationship which accounts for the added noise, we consider the hierarchical model

$$\begin{aligned} \tilde{y}_i &= y_i + \xi_i^y, \quad \xi_i^y \stackrel{\text{ind}}{\sim} \text{DGeom}(\rho_i^y), \\ \tilde{x}_{ij} &= x_{ij} + \xi_{ij}^x, \quad \xi_{ij}^x \stackrel{\text{ind}}{\sim} \text{Lap}(0, \lambda_{ij}^x), \quad j = 1, \dots, d_1, \\ \log y_i &= \mathbf{x}_{\cdot i}^\top \boldsymbol{\beta} + \gamma_i, \quad \gamma_i \stackrel{\text{iid}}{\sim} \text{N}(0, \sigma^2), \\ \boldsymbol{\beta} &\sim \text{N}_d(\mathbf{0}, \sigma_\beta^2 \mathbf{I}_d), \quad \sigma^2 \sim \text{IG}(a_\sigma, b_\sigma). \end{aligned} \tag{13}$$

As analysts outside of the agency, we therefore observe noisy responses  $\tilde{\mathbf{y}}$  and covariate  $\tilde{\mathbf{X}}$  whose first  $d_1$  columns have been protected with agency noise. We are also provided the noise parameters  $\rho_i^y$  and  $\lambda_{ij}^x$  and know that they are associated with Double Geometric and Laplace noise mechanisms, respectively. Hyperparameters  $\sigma_\beta^2$ ,  $a_\sigma$ , and  $b_\sigma$  are chosen by us to facilitate the analysis. For demonstration purposes, the assumed Lognormal regression will be the correct data-generating mechanism; of course, such a mechanism usually is not known with real data so that model specification becomes an important component of the analysis.

A Gibbs sampler to fit model (13) is given in Algorithm 2. The notation  $[\phi | \dots]$  is used to denote the distribution of  $\phi$  conditional on all other random variables. A conjugate prior has been assumed for the model parameters  $\boldsymbol{\theta} = (\boldsymbol{\beta}, \sigma^2)$  so that drawing from their conditionals is routine. This formulation of

<sup>1</sup>As described later in this section, covariates will be drawn from a standard Normal distribution. A Double Geometric noise mechanism produces integer noise which would be large relative to the data; therefore, we instead consider a Laplace noise mechanism which allows finer control over magnitude of the noise.

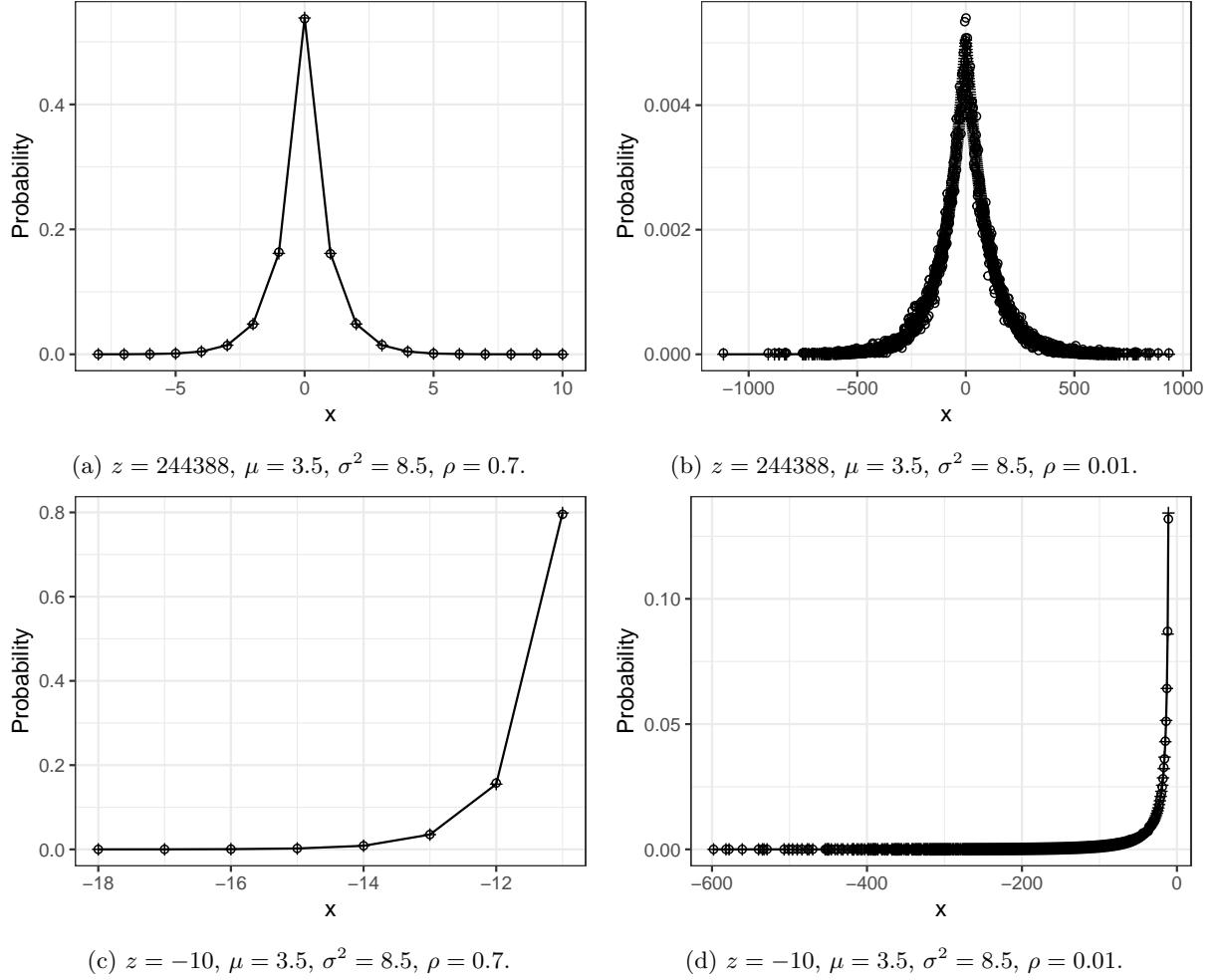


Figure 3: Outcome of CD sampler for Example 4.3 under four realizations. Empirical distribution of the sample ( $\circ$ ) is shown with the density value (+). Connecting lines between density values are shown only for visual reference—the target distribution is defined on  $\mathbb{Z}$ .

Bayesian linear regression is a special case of the one discussed in Hoff (2009, Section 9.2); other formulations would work equally well in this demonstration (e.g. Walter and Augustin, 2010; Gelman et al., 2013).

To derive a Gibbs sampler, consider the joint distribution of  $[\tilde{\mathbf{y}}, \mathbf{y}, \tilde{\mathbf{X}}, \boldsymbol{\xi}^x, \boldsymbol{\theta}]$ , factorized as

$$\begin{aligned} f(\tilde{\mathbf{y}}, \mathbf{y}, \tilde{\mathbf{X}}, \boldsymbol{\xi}^x, \boldsymbol{\theta}) &= f(\tilde{\mathbf{y}} | \mathbf{y}, \tilde{\mathbf{X}}, \boldsymbol{\xi}^x, \boldsymbol{\theta}) \cdot f(\mathbf{y} | \tilde{\mathbf{X}}, \boldsymbol{\xi}^x, \boldsymbol{\theta}) \cdot f(\tilde{\mathbf{X}} | \boldsymbol{\xi}^x, \boldsymbol{\theta}) \cdot f(\boldsymbol{\xi}^x | \boldsymbol{\theta}) \cdot f(\boldsymbol{\theta}) \\ &= f(\tilde{\mathbf{y}} | \mathbf{y}) \cdot f(\mathbf{y} | \tilde{\mathbf{X}}, \boldsymbol{\xi}^x, \boldsymbol{\theta}) \cdot f(\boldsymbol{\xi}^x) \cdot f(\boldsymbol{\theta}) \end{aligned}$$

with

$$\begin{aligned} f(\tilde{\mathbf{y}} | \mathbf{y}) &= \prod_{i=1}^n f_{\text{DGeom}}(\tilde{y}_i - y_i | \rho_i^y), \quad f(\boldsymbol{\xi}^x) = \prod_{i=1}^n \prod_{j=1}^{d_1} f_{\text{Lap}}(\xi_{ij}^x | 0, \lambda_{ij}^x), \\ f(\mathbf{y} | \tilde{\mathbf{X}}, \boldsymbol{\xi}^x, \boldsymbol{\theta}) &= \prod_{i=1}^n f_{\text{LN}}(y_i | \mathbf{x}_{i \cdot}^\top \boldsymbol{\beta}, \sigma^2), \quad f(\boldsymbol{\theta}) = f_{\text{N}}(\boldsymbol{\beta} | \mathbf{0}, \sigma_\beta^2 \mathbf{I}_d) \cdot f_{\text{IG}}(\sigma^2 | a_\sigma, b_\sigma). \end{aligned}$$

For the term  $f(\tilde{\mathbf{y}} | \mathbf{y})$ , the variables  $\tilde{y}_i = y_i + \xi_i^y$  are shifts of  $\xi_i^y \sim \text{DGeom}(\rho_i^y)$  by constants  $y_i$  which are independent for  $i = 1, \dots, n$ , and make an appropriate transformation to obtain the density function. This term is free of model parameters  $\boldsymbol{\theta}$ . From here, we may obtain the desired conditional distributions. From the conjugate prior, we routinely obtain that

$$[\boldsymbol{\beta} | \cdots] \sim \text{N}_d(\boldsymbol{\vartheta}, \boldsymbol{\Omega}^{-1}), \quad \boldsymbol{\Omega} = \sigma^{-2} \mathbf{X}^\top \mathbf{X} + \sigma_\beta^{-2} \mathbf{I}_d, \quad \boldsymbol{\vartheta} = \boldsymbol{\Omega}^{-1} \left( \sigma^{-2} \sum_{i=1}^n \mathbf{x}_{i \cdot} \log y_i \right), \quad (14)$$

$$[\sigma^2 | \cdots] \sim \text{IG}(a^*, b^*), \quad a^* = a_\sigma + \frac{n}{2}, \quad b^* = b_\sigma + \frac{1}{2} \sum_{i=1}^n (\log y_i - \mathbf{x}_{i \cdot}^\top \boldsymbol{\beta})^2, \quad (15)$$

which are obtained routinely. For the unobserved outcomes  $\mathbf{y}$ , we have

$$\begin{aligned} f(\mathbf{y} | \cdots) &\propto \prod_{i=1}^n f_{\text{LN}}(y_i | \mathbf{x}_{i \cdot}^\top \boldsymbol{\beta}, \sigma^2) \cdot f_{\text{DGeom}}(\tilde{y}_i - y_i | \rho_i^y) \\ &\propto \prod_{i=1}^n \left[ \frac{1}{y_i} \exp \left\{ -\frac{1}{2\sigma^2} [\log y_i - \mathbf{x}_{i \cdot}^\top \boldsymbol{\beta}]^2 \right\} (1 - \rho_i^y)^{|\tilde{y}_i - y_i|} \mathbb{I}(y_i \geq 0) \right]. \end{aligned} \quad (16)$$

whose product form indicates that  $y_1, \dots, y_n$  may be drawn independently. These distributions match (12) with  $(\tilde{y}_i - \xi_i^y, \xi_i^y)$  relabeled as  $(y_i, \tilde{y}_i - y_i)$ ; therefore,  $y_i$  can be sampled by first drawing  $\xi_i^y$  from  $[\xi_i^y | \tilde{y}_i, \boldsymbol{\beta}, \sigma^2]$  via (12) using the direct sampler in Section 4, then taking  $y_i = \tilde{y}_i - \xi_i^y$ . Finally, to sample  $\boldsymbol{\xi}^x$ , consider drawing the  $j$ th coordinate  $\boldsymbol{\xi}_{\cdot j}^x = (\xi_{1j}^x, \dots, \xi_{nj}^x)$  conditionally on all other random variables:

$$\begin{aligned} f(\boldsymbol{\xi}_{\cdot j}^x | \cdots) &\propto \prod_{i=1}^n f_{\text{Lap}}(\xi_{ij}^x | 0, \lambda_{ij}^x) \prod_{i=1}^n f_{\text{LN}}(y_i | \mathbf{x}_{i \cdot}^\top \boldsymbol{\beta}, \sigma^2) \\ &\propto \prod_{i=1}^n \left[ e^{-|\xi_{ij}^x|/\lambda_{ij}^x} \exp \left\{ -\frac{1}{2\sigma^2} \left[ \log y_i - \sum_{\ell \neq j} x_{i\ell} \beta_\ell - (\tilde{x}_{ij} - \xi_{ij}^x) \beta_j \right]^2 \right\} \right] \\ &\propto \prod_{i=1}^n \left[ e^{-|\xi_{ij}^x|/\lambda_{ij}^x} \exp \left\{ -\frac{1}{2\tau_{ij}^2} [(\tilde{x}_{ij} - \xi_{ij}^x) - \vartheta_{ij}]^2 \right\} \right], \end{aligned} \quad (17)$$

where  $\tau_{ij}^2 = \sigma^2 / \beta_j^2$  and  $\vartheta_{ij} = \beta_j^{-1} (\log y_i - \sum_{\ell \neq j} x_{i\ell} \beta_\ell)$ . Because (17) factors into a product,  $\xi_{1j}^x, \dots, \xi_{nj}^x$  may be drawn independently within this step. A direct sampler for each coordinate can be obtained following the same procedure as in Section 4, using the expressions in Examples 3.1 and 3.6 corresponding to a Normal

weight function and Laplace base distribution. Note that if  $\beta_j = 0$ , (17) simplifies to a product of Laplace densities because the Lognormal weight functions are free of  $\xi_{\cdot j}^x$ .

A Gibbs sampler using these conditional distributions is stated as Algorithm 2. The special case  $d_1 = 0$  with only noise-free covariates is given as Algorithm 3. Finally, Algorithm 4 states an equivalent sampler when all  $\mathbf{y}$  and  $\mathbf{x}_{\cdot j}$  are observed without noise.

**Remark 5.1.** For comparison, it would be interesting to consider a model which uses the protected data and simply ignores the noise. However, a Lognormal regression model can not immediately be applied as some observations become negative with the added noise. To avoid ad hoc adjustments to the model, along with comparisons to the many alternative sampling approaches which exist in the literature, Algorithm 4 represents an “ideal” situation for comparison.

**Remark 5.2.** Transformations of  $\mathbf{x}$  may be of interest in the regression portion of (13), rather than  $\mathbf{x}$  itself. When  $\tilde{\mathbf{x}}$  is observed with added noise, care must be taken to incorporate such a transformation into (17) to determine an appropriate draw of  $\xi_{\cdot j}^x$ . For example, consider  $h(\mathbf{x}_{\cdot j}) = (h_1(\mathbf{x}_{\cdot j}), \dots, h_n(\mathbf{x}_{\cdot j}))$  which operates only on the  $j$ th variable. A logarithmic transformation  $h_i(\mathbf{x}_{\cdot j}) = \log x_{ij}$  operates coordinate-wise on elements of  $\mathbf{x}_{\cdot j}$  and manifests in (17) as a transformed normal weight function as in Example 3.2. This differs from Example 3.3 where the Jacobian of the transformation appears. On the other hand, standardization via  $h_i(\mathbf{x}_{\cdot j}) = (x_{ij} - \bar{x}_{\cdot j})/s(\mathbf{x}_{\cdot j})$ , with  $\bar{x}_{\cdot j} = \frac{1}{n} \sum_{i=1}^n x_{ij}$  and  $s^2(\mathbf{x}_{\cdot j}) = \frac{1}{n-1} \sum_{i=1}^n (x_{ij} - \bar{x}_{\cdot j})^2$ , involves data from each subject in all coordinates of  $h_i(\mathbf{x}_{\cdot j})$ . Here, (17) no longer factorizes into independent terms by subject.

---

**Algorithm 2** Gibbs sampler for model (13).

---

Repeat the following steps for  $r = 1, \dots, R$  to obtain the desired number of draws  $R$ .

1. Draw  $y_i$  from  $[y_i | \dots]$  for  $i = 1, \dots, n$  using (16).
  2. For each noisy covariate  $j = 1, \dots, d_1$ ,
    - a. Draw  $\xi_{1j}^x, \dots, \xi_{nj}^x$  from  $[\xi_{1j}^x, \dots, \xi_{nj}^x | \dots]$  using (17), conditioning on all  $x_{i\ell}$  for  $\ell \neq j$ .
    - b. Let  $x_{ij} = \tilde{x}_{ij} - \xi_{ij}^x$  for  $i = 1, \dots, n$ .
  3. Draw  $\boldsymbol{\beta}$  from  $[\boldsymbol{\beta} | \dots]$  using (14).
  4. Draw  $\sigma^2$  from  $[\sigma^2 | \dots]$  using (15).
  5. Let  $\mathbf{y}^{(r)} = \mathbf{y}$ ,  $\mathbf{X}^{(r)} = \mathbf{X}$ ,  $\boldsymbol{\beta}^{(r)} = \boldsymbol{\beta}$ , and  $\sigma^{2(r)} = \sigma^2$ .
- 

---

**Algorithm 3** Gibbs sampler for model (13) with  $d_1 = 0$ .

---

Repeat the following steps for  $r = 1, \dots, R$  to obtain the desired number of draws  $R$ .

1. Draw  $y_i$  from  $[y_i | \dots]$  for  $i = 1, \dots, n$  using (16).
  2. Draw  $\boldsymbol{\beta}$  from  $[\boldsymbol{\beta} | \dots]$  using (14).
  3. Draw  $\sigma^2$  from  $[\sigma^2 | \dots]$  using (15).
  4. Let  $\mathbf{y}^{(r)} = \mathbf{y}$ ,  $\boldsymbol{\beta}^{(r)} = \boldsymbol{\beta}$ , and  $\sigma^{2(r)} = \sigma^2$ .
- 

## 5.1 Fixed Covariates

First, consider  $d = 2$  covariates with  $d_1 = 0$  so that both are observed without added agency noise. Here, it may be of interest to study the effects on inference when using noisy responses  $\tilde{\mathbf{y}}$  with Algorithm 3 versus the sensitive  $\mathbf{y}$  with Algorithm 4. We take  $n = 200$ ,  $\boldsymbol{\beta} = (5, -1)$ ,  $\rho_i^y \equiv \rho = 0.01$ , and consider  $\sigma \in \{0.25, 5\}$  corresponding to relatively accurate and imprecise regression models. To give an idea of the magnitude of the noise, the 0.005 and 0.995 quantiles of DGeom(0.01) are -458 and 458 respectively. We generate

**Algorithm 4** Noise-free Gibbs sampler for model (13).

---

Repeat the following steps for  $r = 1, \dots, R$  to obtain the desired number of draws  $R$ .

1. Draw  $\beta$  from  $[\beta | \text{---}]$  using (14).
  2. Draw  $\sigma^2$  from  $[\sigma^2 | \text{---}]$  using (15).
  3. Let  $\beta^{(r)} = \beta$  and  $\sigma^{2(r)} = \sigma^2$ .
- 

covariates  $x_{ij} \sim N(0, 1)$  independently for  $j = 1, 2$  and  $i = 1, \dots, n$ . For each value of  $\sigma$ , we generate observations according to (13) and run Algorithms 3 and 4 for 2,000 iterations, discarding the first 1,000 draws as burn-in. Figures 4 and 5 display the saved 1,000 draws of  $\beta$  and  $\sigma^2$ , respectively. When the data are generated more precisely about the regression with  $\sigma = 0.25$ , we notice that the posterior distribution of  $\beta$  is more dispersed under Algorithm 3 than Algorithm 4. Furthermore, the distribution of  $\sigma^2$  is notably larger and more dispersed under Algorithm 3. When  $\sigma = 5$ , the two algorithms produce results which are much more similar. The data in this setting have also yielded posteriors centered on the true value of  $\sigma^2$ , but this may vary with draws of  $y$ .

The initial results support intuition that agency noise will have a larger effect on inference when the proposed regression model holds more precisely. A simulation study over  $S = 500$  generated datasets confirms the behavior over many realizations. We now consider  $n = 200$ ,  $\beta = (5, -1)$ , and vary  $\rho_i^y \equiv \rho \in \{0.01, 0.1, 0.4\}$  and  $\sigma \in \{0.25, 1, 5\}$ . Covariates  $x_{ij} \sim N(0, 1)$  are generated independently for  $j = 1, 2$  and  $i = 1, \dots, n$  and fixed for the remainder of the study. For each pair of  $\rho$  and  $\sigma$ , we generate  $n$  observations according to (13) and run Algorithms 3 and 4 for 2,000 iterations, discarding the first 1,000 draws as burn-in. This process is repeated to produce  $S$  realizations  $\mathbf{y}^{(s)}, \tilde{\mathbf{y}}^{(s)}, \mathbf{X}^{(s)}$ , and  $\tilde{\mathbf{X}}^{(s)}$ <sup>2</sup> and saved MCMC draws  $\theta^{(r,s)} = (\beta^{(r,s)}, \sigma^{2(r,s)})$  for  $r = 1, \dots, R = 1,000$  of simulation  $s = 1, \dots, S$  from each algorithm. To summarize the posterior distribution of  $\theta$  relative to the true data-generating  $\theta_0$ , we compute the mean-squared error

$$\text{MSE}^{(s)} = \frac{1}{R} \sum_{r=1}^R \|\theta^{(r,s)} - \theta_0\|^2 \approx \int \|\theta - \theta_0\|^2 f(\theta | \tilde{\mathbf{y}}^{(s)}, \tilde{\mathbf{X}}^{(s)}) d\theta \quad (18)$$

for  $s = 1, \dots, S$  under both algorithms. The resulting empirical distributions of  $\text{MSE}^{(1)}, \dots, \text{MSE}^{(S)}$  are displayed in Figure 6. This supports our initial intuition that the loss of precision, in using  $\mathbf{y}$  with Algorithm 4 to using  $\tilde{\mathbf{y}}$  with Algorithm 3, is most pronounced when the regression model is more precise and there is more agency noise. Conversely, inference from the two algorithms tends to become more similar as agency noise is reduced or the data-generating regression model becomes less precise.

## 5.2 Noisy Covariate

Now, consider  $d = 2$  covariates with  $d_1 = 1$  so that  $\tilde{x}_{.1}$  is released with Laplace noise and  $\mathbf{x}_{.2}$  is unprotected. Here we illustrate the effects on inference when using Algorithm 2 with the privacy-protected releases versus Algorithm 4 with the sensitive data. We take  $n = 200$ ,  $\beta = (5, -1)$ ,  $\rho_i^y \equiv \rho = 0.01$ ,  $\sigma = 1$ , and consider  $\lambda_{ij}^x \equiv \lambda \in \{0.05, 0.20\}$ . For reference, (0.005, 0.995) quantiles for the  $\text{Lap}(0, 0.05)$  distribution are  $(-0.230, 0.230)$ , and are  $(-0.921, 0.921)$  for the  $\text{Lap}(0, 0.20)$  distribution; therefore  $\lambda = 0.20$  represents a larger magnitude of noise added to covariates. We generate covariates  $x_{ij} \sim N(0, 1)$  independently for  $j = 1, 2$  and  $i = 1, \dots, n$ . For each value of  $\lambda$ , we generate observations according to (13) and Algorithms 2 and 4 are used to produce a chain of 2,000 draws of  $\theta$ , discarding the first 1,000 draws as burn-in. In particular, at each step of the simulation, Algorithms 2 and 4 make use of common realizations of the generated data. Figures 7 and 8 display the saved 1,000 draws of  $\beta$  and  $\sigma^2$ , respectively. As may be anticipated, observations generated with larger  $\lambda$  result in increased bias and uncertainty in the posterior of  $\beta$ .

---

<sup>2</sup>To emphasize,  $\mathbf{X}^{(s)}$  and  $\tilde{\mathbf{X}}^{(s)}$  are equivalent in Section 5.1 and different in Section 5.2.

To empirically assess the behavior over many samples, a simulation study similar to Section 5.1 was carried out. We now consider  $n = 200$ ,  $\beta = (5, -1)$  and  $\sigma = 1$ , and vary  $\rho_i^y \equiv \rho \in \{0.01, 0.1, 0.4\}$  and  $\lambda_{ij}^x \equiv \lambda \in \{0.05, 0.10, 0.20\}$ . Covariates  $x_{ij} \stackrel{\text{ind}}{\sim} N(0, 1)$  are fixed for the duration of the study. For each combination of  $\rho$  and  $\lambda$ , observations are generated according to (13) and Algorithms 2 and 4 are each run for 2,000 iterations, discarding the first 1,000 draws as burn-in. This process is repeated  $S = 500$  times, yielding  $\boldsymbol{\theta}^{(r,s)}$  and  $\text{MSE}^{(s)}$  as defined in (18).

The empirical distributions of  $\text{MSE}^{(1)}, \dots, \text{MSE}^{(S)}$  are displayed in Figure 9. This supports the intuition that more agency noise or more precise data generation leads to less precise inference from Algorithm 2 than Algorithm 4. Conversely, inference from the two algorithms tends to become more similar as agency noise is reduced or the data-generating regression model becomes less precise. Larger magnitudes of  $\xi_i^y$  or  $\xi_{ij}^x$ , from smaller  $\rho$  and larger  $\lambda$  respectively, decrease the precision of the posterior distribution of  $\boldsymbol{\theta}$  from Algorithm 2 compared to Algorithm 4 where privacy is ignored.

## 6 Discussion and Conclusions

In this work, we revisited the direct sampling approach proposed by Walker et al. (2011), which appears to be underappreciated in the literature. To formulate Gibbs sampling steps involving pairs of commonly used but non-conjugate distributions, we considered several customizations to the basic direct sampling approach for robustness in a variety of encountered situations, and to avoid rejecting proposed samples. First, a step function approximation helped to capture the shape of the density  $p(u)$ , which is monotone and nonincreasing on  $[0, 1]$  but subject to sudden jumps that may all occur within a very narrow interval. Second, for unimodal weight functions and base distributions with a readily computed quantile function, draws from the required truncated base distribution may be taken without the need for rejections.

The resulting sampler provided a means to draw latent agency noise in a disclosure avoidance setting, where noise distributions are selected to enforce privacy protection and not necessarily for convenience in a Gibbs sampler. This was illustrated in a small empirical study, comparing models fit using noise-free and noise-infused data, corresponding to sensitive data before and after disclosure avoidance measures have been taken. As we might expect, inference based on the noise-infused data becomes more distorted when the accuracy of the underlying regression model and the amount of noise are increased. The primary objective here was to demonstrate the sampler; a more in-depth study of the relationship between inference and DP noise is provided by Gong (2020).

Draws of the latent agency noise were by far the slowest step of the otherwise straightforward Gibbs sampler in our illustration, but might still be considered reasonable. For example, a run of Algorithm 3 to produce draws for Figure 4b took 77.34 seconds to complete on an Intel Core i7-2600 3.40 GHz workstation with four CPU cores; of this, 76.11 seconds were spent in drawing the latent  $\mathbf{y}$ . A run of Algorithm 2 to produce draws for Figure 7b completed in 129.29 seconds on the same workstation; of this, 70.29 seconds were spent drawing  $\mathbf{y}$  and 57.67 seconds were spent drawing  $\xi^x$ . With multiple CPUs and some additional programming, elapsed time might be considerably reduced by parallel computing: elements of  $\mathbf{y}$  and  $\mathbf{x}_{\cdot j}$  may be drawn in parallel as their respective steps are encountered in the overall sampler. Care with floating point operations was also required for a resilient implementation of the sampler.

The illustration in Section 5 represents a simplified setting where subject-level data are released. Real data releases have a number of interesting features which may require nontrivial extensions. For example, sensitive data may consist of tabulations where noise variates are added to each cell. Sensitive data may also represent estimates from a survey and include both point estimates and margins of error. Such extensions may be of increasing interest as disclosure avoidance becomes more widely adopted and data users seek to include this source of variability in analyses.

## Acknowledgements

The author is grateful to Drs. Scott Holan, Kyle Irimata, Ryan Janicki, and James Livsey at the U.S. Census Bureau for discussions on statistical modeling of releases from a disclosure avoidance system which motivated this work.

## References

- Garrett Bernstein and Daniel R Sheldon. Differentially private Bayesian linear regression. In H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett, editors, *Advances in Neural Information Processing Systems 32*, pages 525–535. Curran Associates, Inc., 2019.
- Claire McKay Bowen and Fang Liu. Comparative study of differentially private data synthesis methods. *Statistical Science*, 35(2):280–307, 2020.
- Michael Braun and Paul Damien. Scalable rejection sampling for Bayesian hierarchical models. *Marketing Science*, 35(3):427–444, 2016.
- Anne-Sophie Charest. How can we analyze differentially-private synthetic datasets? *Journal of Privacy and Confidentiality*, 2(2), 2011.
- Thomas H. Cormen, Charles E. Leiserson, Ronald L. Rivest, and Clifford Stein. *Introduction to Algorithms*. The MIT Press, 3rd edition, 2009.
- A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society: Series B (Methodological)*, 39(1):1–22, 1977.
- Cynthia Dwork and Aaron Roth. *The Algorithmic Foundations of Differential Privacy*. Now Publishers Inc, 2014.
- Cynthia Dwork and Adam Smith. Differential privacy for statistics: What we know and what we want to learn. *Journal of Privacy and Confidentiality*, 1(2), 2010.
- Dirk Eddelbuettel. *Seamless R and C++ Integration with Rcpp*. Springer, New York, 2013. doi: 10.1007/978-1-4614-6868-4. ISBN 978-1-4614-6867-7.
- Georgina Evans and Gary King. Statistically valid inferences from differentially private data releases, with application to the Facebook URLs dataset, 2020+. <https://gking.harvard.edu/dpd>.
- Simson L. Garfinkel, John M. Abowd, and Sarah Powazek. Issues encountered deploying differential privacy. In *Proceedings of the 2018 Workshop on Privacy in the Electronic Society*, WPES’18, pages 133–137, New York, NY, USA, 2018. Association for Computing Machinery.
- Andrew Gelman, John B. Carlin, Hal S. Stern, David B. Dunson, Aki Vehtari, and Donald B. Rubin. *Bayesian Data Analysis*. Chapman and Hall/CRC, 3rd edition, 2013.
- Arpita Ghosh, Tim Roughgarden, and Mukund Sundararajan. Universally utility-maximizing privacy mechanisms. *SIAM Journal on Computing*, 41(6):1673–1693, 2012.
- Ruobin Gong. Exact inference with approximate computation for differentially private data via perturbations, 2019. <https://arxiv.org/abs/1909.12237>.
- Ruobin Gong. Transparent privacy is principled privacy, 2020. <https://arxiv.org/abs/2006.08522>.
- Peter D. Hoff. *A First Course in Bayesian Statistical Methods*. Springer, 2009.
- Martin Klein and Bimal Sinha. Multiple imputation for parametric inference under a differentially private laplace mechanism. Technical Report Statistics #2019-05, Center for Statistical Research and Methodology, U.S. Census Bureau, 2019. <https://www.census.gov/library/working-papers/2019/adrm/RRS2019-05.html>.
- Yu Hsuan Kuo, Cho Chun Chiu, Daniel Kifer, Michael Hay, and Ashwin Machanavajjhala. Differentially private hierarchical count-of-counts histograms. In *Proceedings of the VLDB Endowment*, volume 11, pages 1509–1521. Very Large Data Base Endowment Inc., 2018. doi: 10.14778/3236187.3236202.
- Kenneth Lange. *Numerical Analysis for Statisticians*. Springer, 2nd edition, 2010.
- Luca Martino, David Luengo, and Joaquín Míguez. *Accept–Reject Methods*, pages 65–113. Springer International Publishing, Cham, 2018.
- Gregory J. Matthews and Ofer Harel. Data confidentiality: A review of methods for statistical disclosure limitation and methods for assessing privacy. *Statistics Surveys*, 5:1–29, 2011.

- G. P. Patil and C. R. Rao. Weighted distributions and size-biased sampling with applications to wildlife populations and human families. *Biometrics*, 34(2):179–189, 1978.
- R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2020. <https://www.R-project.org/>.
- Theodore J. Rivlin. *An Introduction to the Approximation of Functions*. Dover, 1981.
- Martin A. Tanner and Wing Hung Wong. The calculation of posterior distributions by data augmentation. *Journal of the American Statistical Association*, 82(398):528–540, 1987.
- Stephen G. Walker, Purushottam W. Laud, Daniel Zantedeschi, and Paul Damien. Direct sampling. *Journal of Computational and Graphical Statistics*, 20(3):692–713, 2011.
- Gero Walter and Thomas Augustin. *Bayesian Linear Regression—Different Conjugate Models and Their (In)Sensitivity to Prior-Data Conflict*, pages 59–78. Physica-Verlag HD, Heidelberg, 2010.

## A Appendix

A bisection search method (e.g. Lange, 2010, Section 5) is useful in several computations in the CD sampler. Suppose  $\mathcal{S} \subseteq \mathbb{R}$  and  $\zeta(x) : \mathcal{S} \rightarrow \{0, 1\}$  is a step function which increases from 0 to 1 at a point  $x^*$ . The objective of Algorithm 5 is to identify  $x^*$  by supplying lower and upper bounds  $x_L < x_H$  such that  $\zeta(x_L) = 0$  and  $\zeta(x_H) = 1$ , a function  $\text{mid}(x, y) : \mathcal{S}^2 \rightarrow \mathcal{S}$  which returns a point in  $[x, y]$ , and a distance function  $\text{dist}(x, y)$ . We may therefore write  $x^* = \min\{x \in [x_L, x_H] : \zeta(x) = 1\}$ .

---

**Algorithm 5** Bisection search for  $x^* = \min\{x \in [x_L, x_H] : \zeta(x) = 1\}$ . Inputs are bounds  $x_L < x_H$ , a step function  $\zeta(x)$  with  $\zeta(x_L) = 0$  and  $\zeta(x_H) = 1$ , a midpoint function  $\text{mid}(x, y)$ , a distance function  $\text{dist}(x, y)$ , and a tolerance  $\delta > 0$ .

---

```

 $x = \text{mid}(x_L, x_H)$ 
while  $\text{dist}(x_L, x_H) > \delta$  do
     $x_L = \zeta(x) \cdot x_L + [1 - \zeta(x)] \cdot x$ 
     $x_H = \zeta(x) \cdot x + [1 - \zeta(x)] \cdot x_H$ 
     $x = \text{mid}(x_L, x_H)$ 
end while
return  $x$ 
```

---

Algorithm 5 is useful in the following computations in Section 2.

1. The interval  $[u_L, u_H]$  represents the range where  $p(u)$  decreases from  $P(A_0)$  to 0. To find the point  $u_L$ , we first locate a sufficiently small  $j^* \in \{0, 1, 2, 4, 8, \dots\}$  until  $P(A_{\exp(-j)}) = P(A_0)$ . Now,  $u_L$  represents the smallest  $u \in [0, 1]$  such that  $P(A_u) < P(A_0)$ , and may be located by Algorithm 5 using  $\zeta(x) = I\{P(A_x) < P(A_0)\}$  with  $x_L = e^{-j^*}$ ,  $x_H = 1$ ,  $\text{mid}(x, y) = (x + y)/2$  and  $\text{dist}(x, y) = y - x$ .
2. The point  $u_H$  represents the smallest  $u \in [0, 1]$  such that  $P(A_u) = 0$ . It may be located by Algorithm 5 using  $\zeta(x) = I\{P(A_x) = 0\}$ ,  $x_L = u_L$ , and  $x_H = 1$ .
3. The quantile function  $H^{-1}(\varphi)$  may be evaluated by Algorithm 5. Given precomputed values  $H(u_0), \dots, H(u_N)$  of the associated CDF, the index  $\ell$  of the interval containing  $\varphi$  can be identified using  $\mathcal{S} = \{0, 1, \dots, N\}$ ,  $x_L = 0$ ,  $x_H = N$ ,  $\text{mid}(\ell_1, \ell_2) = \lfloor (\ell_1 + \ell_2)/2 \rfloor$ , and  $\zeta(\ell) = I\{H(u_\ell) \geq \varphi\}$ . From here, linearity between  $H(u_\ell)$  and  $H(u_{\ell+1})$  yields  $H^{-1}(\varphi) = u_\ell + (u_{\ell+1} - u_\ell)\{\varphi - H(u_\ell)\}/\{H(u_{\ell+1}) - H(u_\ell)\}$ .

In practice, we work on the log-scale to seek  $\log u_L$  and  $\log u_H$ , as  $u_L$  and  $u_H$  may be numbers with extremely small magnitudes. A bisection approach is useful in this setting because smoothness is not required in  $P(A_u)$ . Both phenomena— $u_L$  and  $u_H$  with extremely small magnitudes and  $P(A_u)$  with abrupt changes—are seen in Figure 1.

*Proof of Result 2.1.* First considering the unnormalized densities,

$$\begin{aligned}
\sup_{B \in \mathcal{B}} \left| \int_B h^*(u) du - \int_B P(A_u) du \right| &= \sup_{B \in \mathcal{B}} \int_B [h^*(u) - P(A_u)] du \\
&= \int_0^1 [h^*(u) - P(A_u)] du \\
&= \sum_{j=1}^N \int_{u_{j-1}}^{u_j} [h^*(u) - P(A_u)] du \\
&\leq \sum_{j=1}^N \int_{u_{j-1}}^{u_j} [P(A_{u_{j-1}}) - P(A_{u_j})] du \\
&= \sum_{j=1}^N [P(A_{u_{j-1}}) - P(A_{u_j})](u_j - u_{j-1}) \\
&= \sum_{j=1}^N |\mathcal{R}_j|. \tag{19}
\end{aligned}$$

We have used the fact that  $h^*(u) = P(A_u)$  for  $u \in [0, u_0]$  and  $h^*(u) \geq P(A_u)$  otherwise. Next, integrating each term of the inequality  $P(A_u) \leq h^*(u) \leq 1$  over  $u \in [0, 1]$  gives  $\psi/c \leq a \leq 1$ , and inverting gives  $1 \leq 1/a \leq c/\psi$ . Combining this with (19) yields inequalities for the normalized densities

$$\int_B h(u) du - \int_B p(u) du \leq \frac{c}{\psi} \left[ \int_B h^*(u) du - \int_B P(A_u) du \right] \leq \frac{c}{\psi} \sum_{j=1}^N |\mathcal{R}_j|. \tag{20}$$

and

$$\begin{aligned}
\int_B p(u) du - \int_B h(u) du &\leq \frac{c}{\psi} \int_B h^*(u) du - \frac{1}{a} \int_B h^*(u) du \\
&= \left[ \frac{a - \psi/c}{a\psi/c} \right] \int_B h^*(u) du \\
&\leq \left[ \frac{a - \psi/c}{a\psi/c} \right] a \\
&= \frac{c}{\psi} \left[ \int_0^1 h^*(u) du - \int_0^1 P(A_u) du \right] \\
&\leq \frac{c}{\psi} \sum_{j=1}^N |\mathcal{R}_j|. \tag{21}
\end{aligned}$$

The result follows from (20) and (21).  $\square$

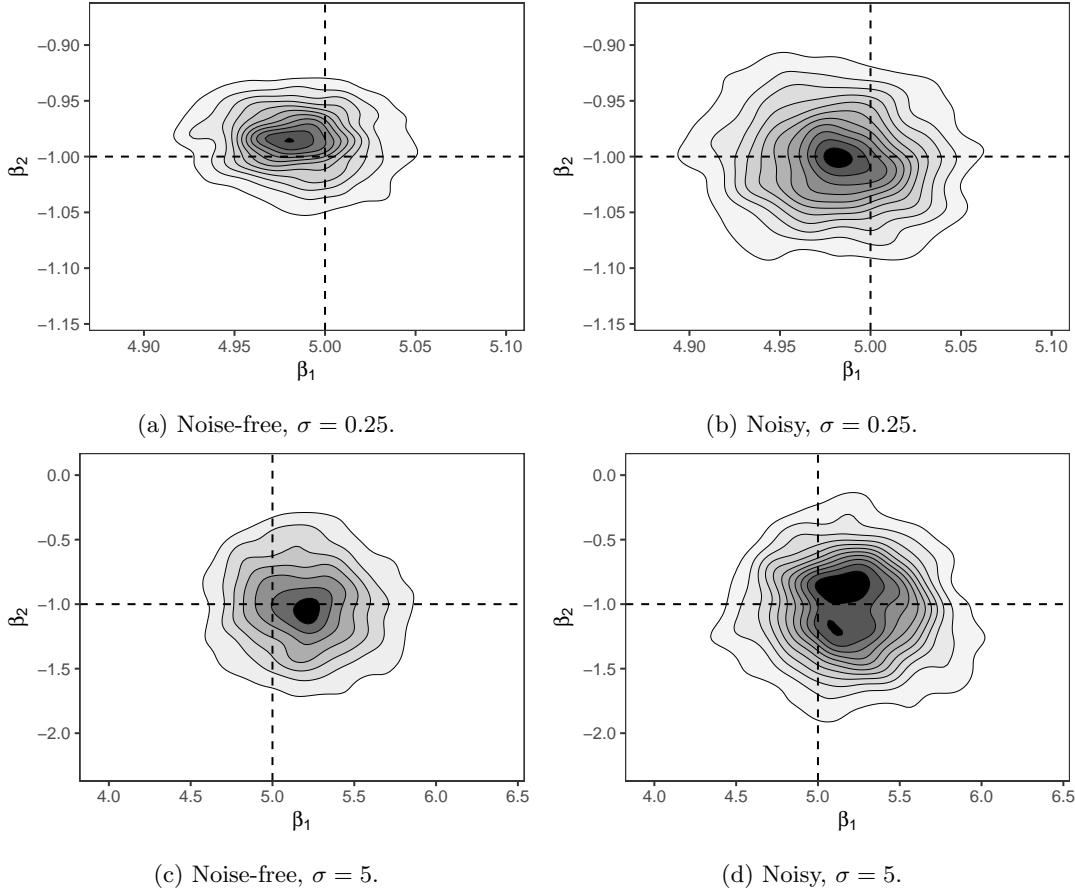


Figure 4: Empirical density—based on 1,000 draws—of the posterior distribution of  $\beta$ , for particular data realizations in Section 5.1. Data generating parameter values were  $\beta = (5, -1)$ ,  $\rho = 0.01$ , and the displayed value of  $\sigma$ . (b) and (d) adjust for agency noise via Algorithm 3, while (a) and (c) utilizes Algorithm 4 with sensitive data observed.

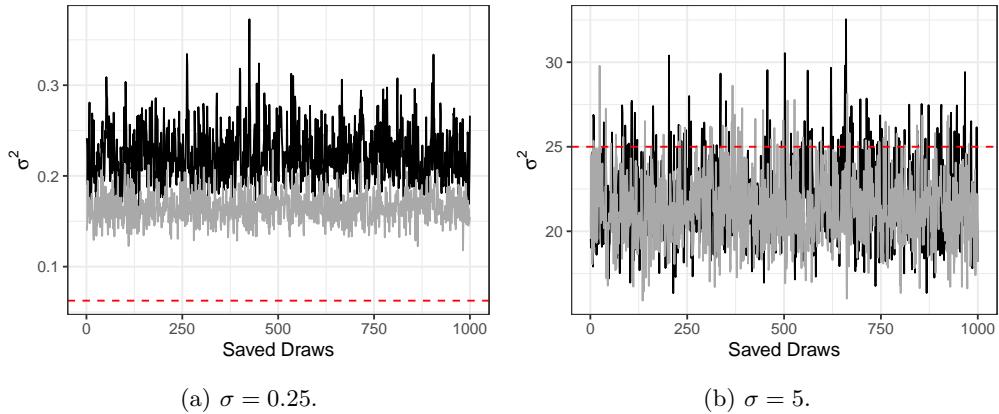


Figure 5: Traceplots of  $\sigma^2$  draws for particular data realizations in Section 5.1. Black and grey lines correspond to noisy and noise-free fits under Algorithms 3 and 4, respectively, and red dashed lines are true data-generating values of  $\sigma^2$ .

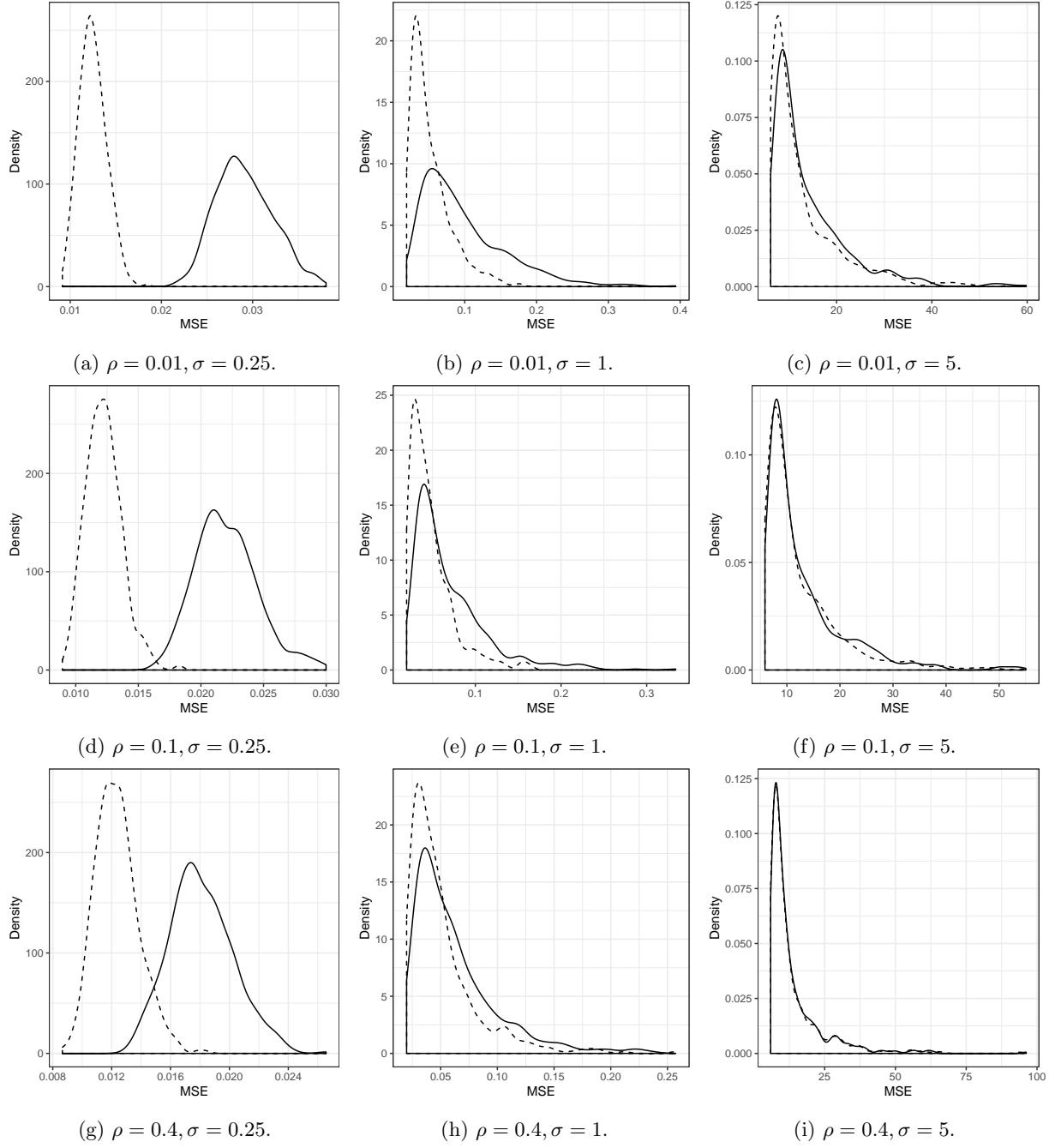


Figure 6: For Section 5.1, the empirical density of the MSE (18) over  $S = 500$  simulation repetitions. The solid line and dashed lines represent noisy and noise-free fits under Algorithms 3 and 4, respectively.

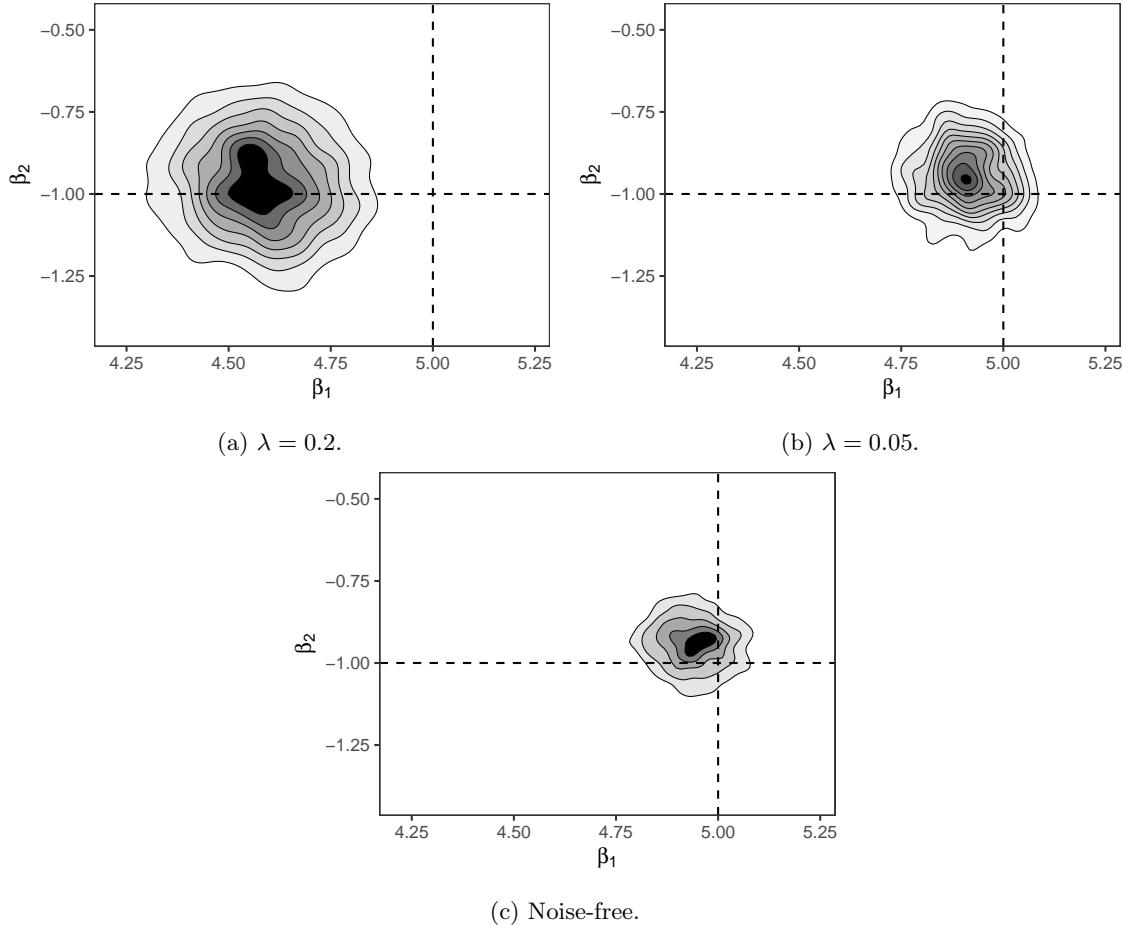


Figure 7: Empirical density—based on 1,000 draws—of the posterior distribution of  $\beta$ , for a particular data realization in Section 5.2. Data generating parameter values were  $\beta = (5, -1)$ ,  $\rho = 0.01$ , and  $\sigma = 1$ . (a) and (b) adjust for agency noise via Algorithm 2, while (c) utilizes Algorithm 4 with sensitive data observed.

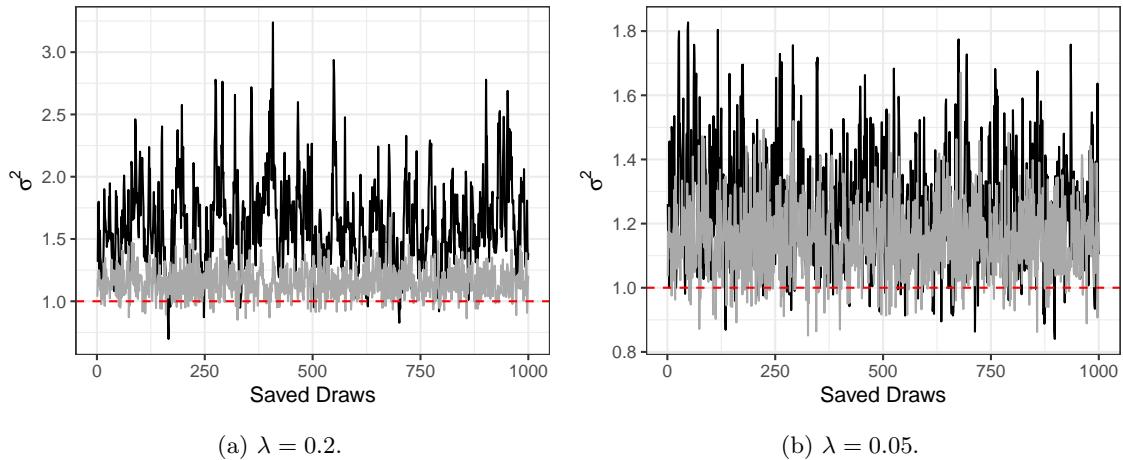


Figure 8: Traceplots of  $\sigma^2$  draws for a particular data realization in Section 5.2. The black and grey lines correspond to Algorithms 2 and 4, respectively, and red dashed lines mark true data-generating value  $\sigma^2 = 1$ .

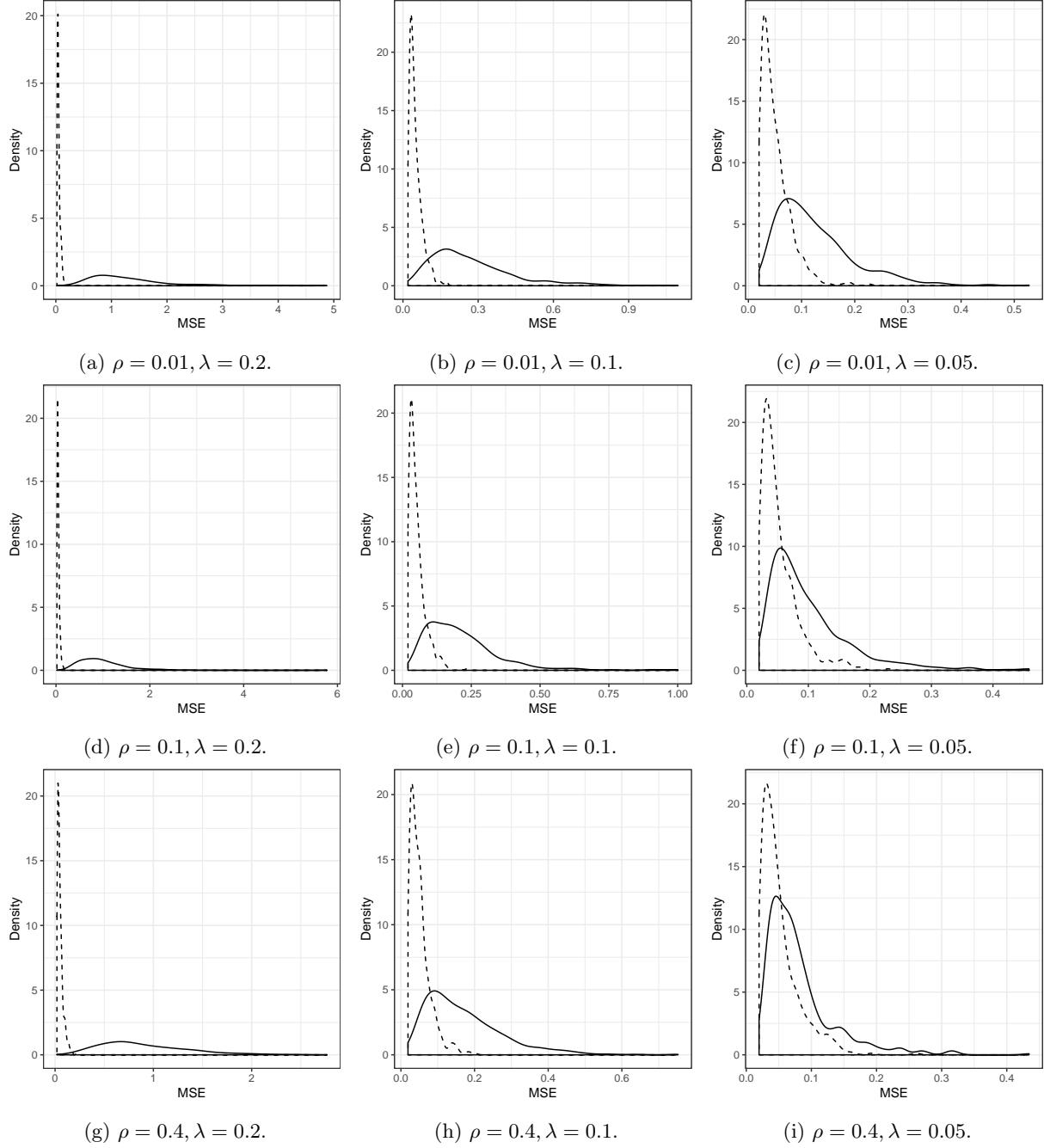


Figure 9: For Section 5.2, the empirical density of the MSE (18) over  $S = 500$  simulation repetitions. Solid line and dashed lines represent Algorithms 2 and 4, respectively.