0.4pt=0pt

# Mechanisms for Global Differential Privacy under Bayesian Data Synthesis

*Vassar College, RTI International and U.S. Bureau of Labor Statistics*

## Supplementary Materials

This Supplementary Materials document contains: 1) a detailed review of the perturbed histogram method in Section 2.3, 2) the Stan script for the censoring method for the beta synthesizer used in Section 4, 3) additional privacy comparison results from Section 4.2, 4) additional utility comparison results from Section 4.3, and 5) privacy and additional utility comparison results from Section 4.4.

# S1 Detailed review of the perturbed histogram method in Section 2.3

We present the commonly-used perturbed histogram mechanism for simulating synthetic microdata that achieves $\epsilon-$DP guarantee (Dwork et al., 2006; Wasserman and Zhou, 2010) as a comparison. Under the required strong assumption of a *bounded* and continuous variable, one first discretizes it into a histogram with a selected number of bins. One induces a formal $\epsilon-$DP privacy guarantee into the histogram by adding Laplace noise. The $\epsilon-$DP guarantee is only global to the extent that one assumes the data space of datasets of size $n$, $\mathcal{X}^n$ is absolutely bounded, which is highly unlikely in practice. Finally, one simulates microdata from the private histogram under $\epsilon-$DP, which is a post-processing step in a similar fashion as generating synthetic data under the pseudo posterior mechanism (given the privacy protected parameter draws) reviewed in Section 2.1.

We describe the perturbed histogram synthesizer for univariate data, $\mathbf{x}$, of size $n$, with a bound of size $L$. We select the number of bins, $m$, that we use to partition $\mathbf{x}$ into $m$ bins, $\{B_1, \cdots, B_m\}$, where each bin, $B_j$, is of length $L/m$. The choice of $m$ should be independent of the data $\mathbf{x}$; e.g., $m = \ln(n)$ or $m = \sqrt{n}$. Let $C_j = \sum_{i=1}^{n} I(x_i \in B_j)$, where $I(\cdot)$ is the

indicator function; i.e., $C_j$ is the number of observations in bin $B_j$.

To privatize the resulting histogram, let $D_j = C_j + \eta_j$, where $\eta_1, \cdots, \eta_m \overset{i.i.d.}{\sim} \text{Laplace}(0, 2/\epsilon)$. In other words, each bin count $C_j$ has added noise from a Laplace distribution with mean 0 and scale $2/\epsilon$, where $\epsilon$ is the targeted privacy budget and 2 is the global sensitivity of a histogram (to reflect the move of a unit from one bin to another). This noise addition process guarantees $\epsilon-$DP for $\boldsymbol{D} = (D_1, \cdots, D_m)$ (Dwork et al., 2006; Wasserman and Zhou, 2010).

Finally, to create synthetic microdata from private $\boldsymbol{D}$, define $\tilde{D}_j = \max(D_j, 0)$. Calculate $\hat{q}_j = \tilde{D}_j / \sum_s \tilde{D}_s$, which is the probability of membership in each privatized bin $B_j$. To simulate a synthetic microdata value for record $i$, we first take a multinomial draw under probabibilities $(\hat{q}_1, \cdots, \hat{q}_m)$, resulting in a bin indicator $b_i \in (1, \cdots, m)$. Next, given the sampled bin indicator $b_i$, we take a uniform draw from that bin to generate synthetic value $x_i^*$ for record $i$. We repeat this process for all records, obtaining a synthetic dataset $\mathbf{x}^*$ from the perturbed histogram synthesizer.

## S2 Stan script for the censoring method for the beta synthesizer used in Section 4

Below we include the Stan script for our censoring method for the beta synthesizer used in our simulation study in Section 4. Note that the value `M` in the script is set as $M = \epsilon/2$.

```
functions{

  real betawt_lpdf(vector outcome, real beta1, real beta2, vector alpha, int n, real M)
  {
    real check_term;
    real update_term;
    check_term = 0.0;
    for( i in 1:n )
    {
        update_term = alpha[i] * beta_lpdf(outcome[i] | beta1, beta2);
            check_term = check_term + fmax(fmin(update_term, M), -M);
    }
    return check_term;
  }

  real betawt_i_lpdf(real outcome_i, real beta1, real beta2, real alpha_i, real M){
    real check_term;
    real update_term;
            update_term = alpha_i * beta_lpdf(outcome_i | beta1, beta2);
            check_term = fmax(fmin(update_term, M), -M);
    return check_term;
  }

  real betawt_i_noM_lpdf(real outcome_i, real beta1, real beta2, real alpha_i){
    real check_term;
            check_term = alpha_i * beta_lpdf(outcome_i | beta1, beta2);
    return check_term;
  }

} /* end function{} block */
```

```
data {
    int<lower=1> n; // number of observations
    vector[n] outcome; // Response variable
    vector<lower=0>[n] alpha; // observation−indexed (privacy) weights
    real<lower=0> M; //censoring threshold
}

parameters{
  real<lower=0,upper=1> phi;
  real<lower=0.1> lambda;
}

transformed parameters{
  real<lower=0> beta1 = lambda * phi;
  real<lower=0> beta2 = lambda * (1 - phi);
}

model{
  phi ~ beta(1, 1); // uniform on phi, could drop
  lambda ~ pareto(0.1, 1.5);

  target += betawt_lpdf(outcome | beta1, beta2, alpha, n, M);
} /* end model{} block */

generated quantities{
  vector[n] log_lik;
  vector[n] log_lik_noM;

  for (i in 1:n) {
        log_lik[i] = betawt_i_lpdf(outcome[i] | beta1, beta2, alpha[i], M);
        log_lik_noM[i] = betawt_i_noM_lpdf(outcome[i] | beta1, beta2, alpha[i]);
        }
}
```

# S3 Additional privacy comparison results from Section 4.2

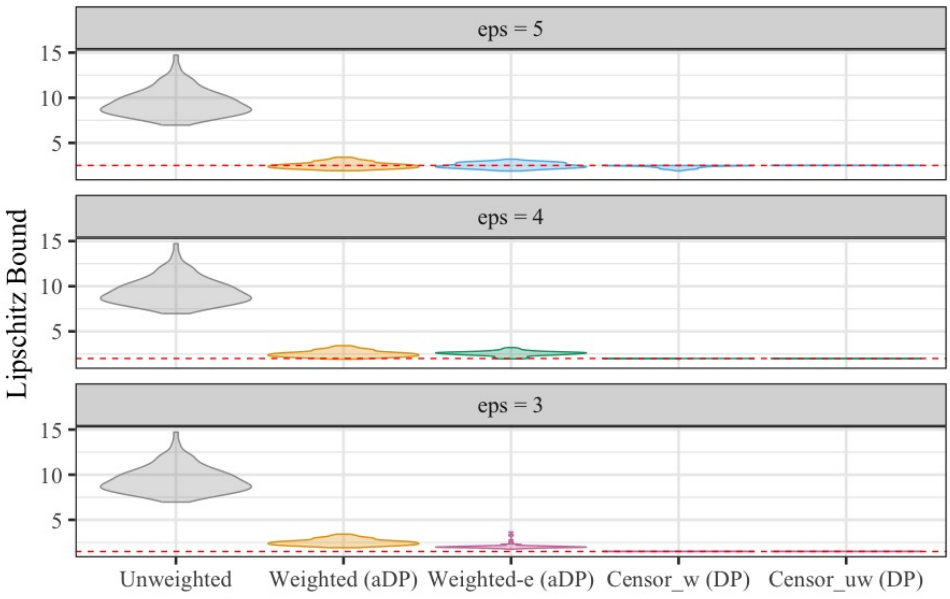


Figure 1: Violin plots of Lipschitz bounds over $R = 100$ replicates under the Unweighted, the Weighted (aDP), the Weighted-e (aDP), the Censor_w (DP), and the Censor_uw (aDP), with $\epsilon$ values of $\{5, 4, 3\}$. A dashed horizontal line at $\epsilon/2$ is included in each panel.

# S4 Additional utility comparison results from Section 4.3

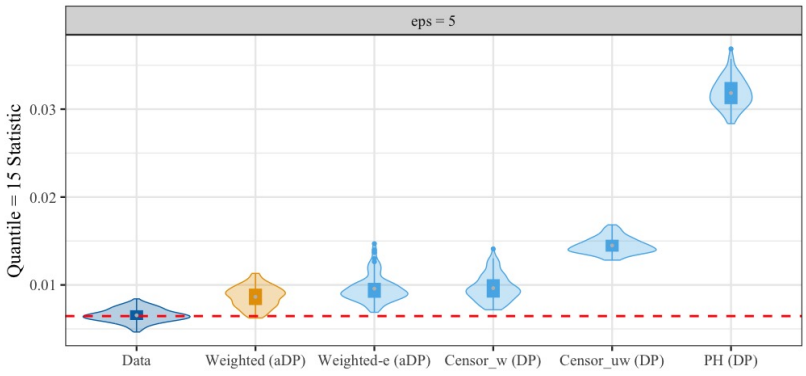


Figure 2: Violin plots of the 15th quantile over $R = 100$ replicates for the Weighted (aDP), the Weighted-e (aDP), the Censor_w (DP), the Censor_uw (DP), and the PH (DP), at $\epsilon = 5$. A dashed horizontal line at the analytical 15th quantile from Beta(0.5, 3) is included.

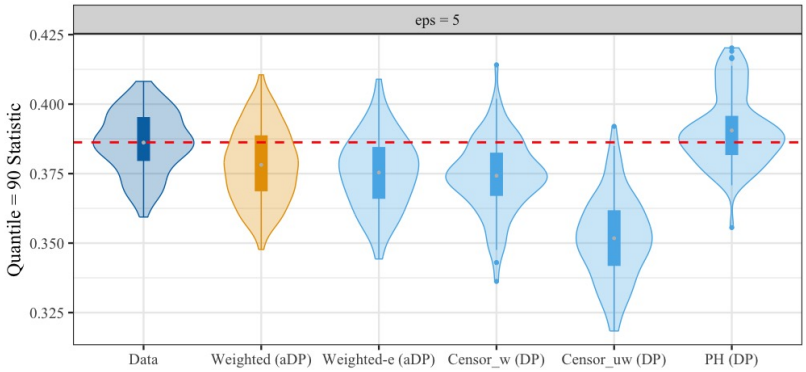


Figure 3: Violin plots of the 90th quantile over $R = 100$ replicates for the Weighted (aDP), the Weighted-e (aDP), the Censor_w (DP), the Censor_uw (DP), and the PH (DP), at $\epsilon = 5$. A dashed horizontal line at the analytical 90th quantile from Beta(0.5, 3) is included.

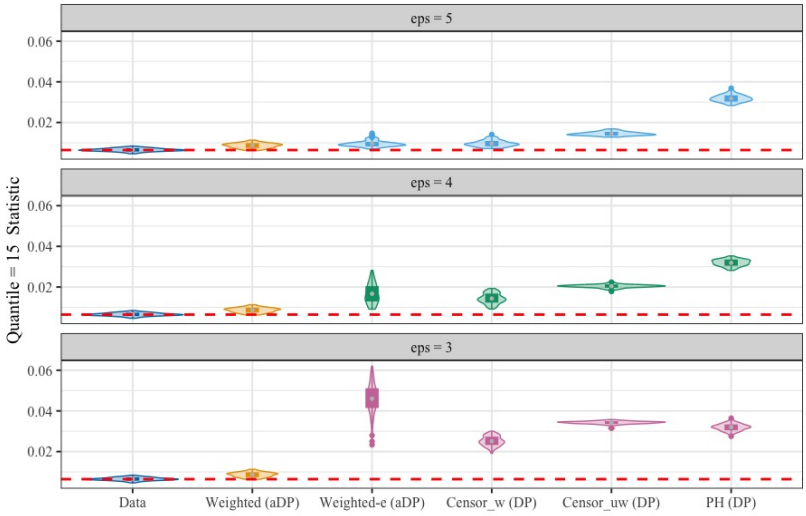


Figure 4: Violin plots of the 15th quantile over $R = 100$ replicates, for the Weighted (aDP), the Weighted-e (aDP), the Censor_w (DP), the Censor_uw (DP), and the PH (DP), with $\epsilon$ values of $\{5, 4, 3\}$. A dashed horizontal line at the analytical 15h quantile from Beta(0.5, 3) is included in each panel.

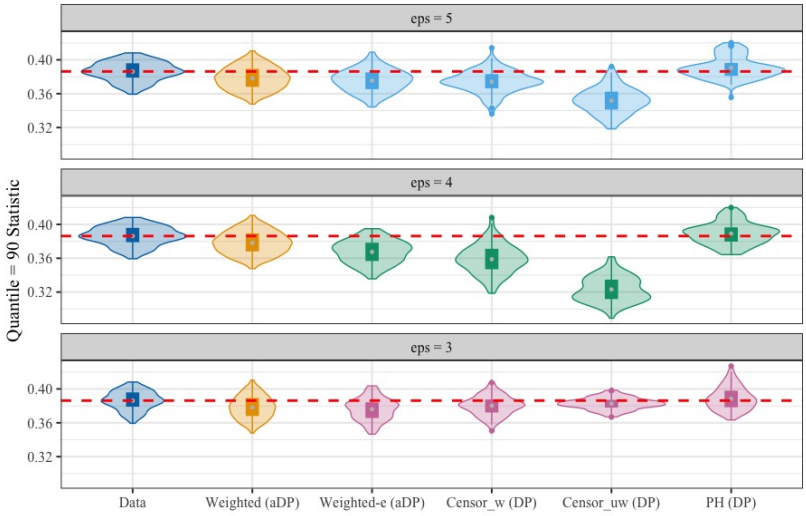


Figure 5: Violin plots of the 90th quantile over $R = 100$ replicates, for the Weighted (aDP), the Weighted-e (aDP), the Censor_w (DP), the Censor_uw (DP), and the PH (DP), with $\epsilon$ values of $\{5, 4, 3\}$. A dashed horizontal line at the analytical 90th quantile from Beta(0.5, 3) is included in each panel.

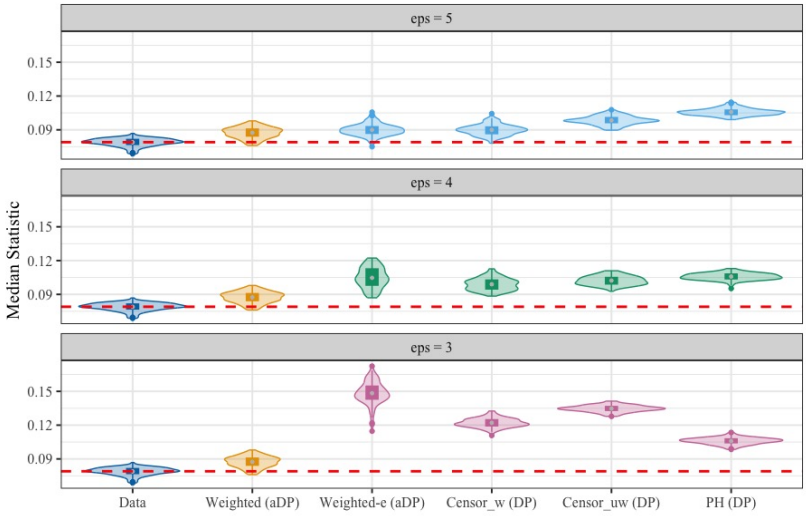


Figure 6: Violin plots of the median over $R = 100$ replicates, for the Weighted (aDP), the Weighted-e (aDP), the Censor_w (DP), the Censor_uw (DP), and the PH (DP), with $\epsilon$ values of $\{5, 4, 3\}$. A dashed horizontal line at the analytical median from Beta(0.5, 3) is included in each panel.

# S5 Privacy and additional utility comparison results from Section 4.4

| | | Min | Q1 | Median | Mean | Q3 | Max | sd |
|---|---|---|---|---|---|---|---|---|
| $\epsilon = 4$ | Weighted-e (aDP) | 0 | 130 | 219 | 195 | 272 | 364 | 103.48 |
| | Censor_w (DP) | 0 | 404 | 425 | 408 | 440 | 549 | 92.56 |
| $\epsilon = 4$ | Weighted-e (aDP) | 0 | 0 | 93 | 91 | 165 | 269 | 86.49 |
| (downscale) | Censor_w (DP) | 0 | 0 | 0 | 144 | 328 | 413 | 165.66 |

Table 1: Summaries of the number of records (out of $n = 2000$) receiving truncated weight at $\alpha_i = 0$ in Weighted-e (aDP) and censored likelihood at $\epsilon/2$ in Censor_w (DP). The number of Monte Carlo simulations is $R = 100$.

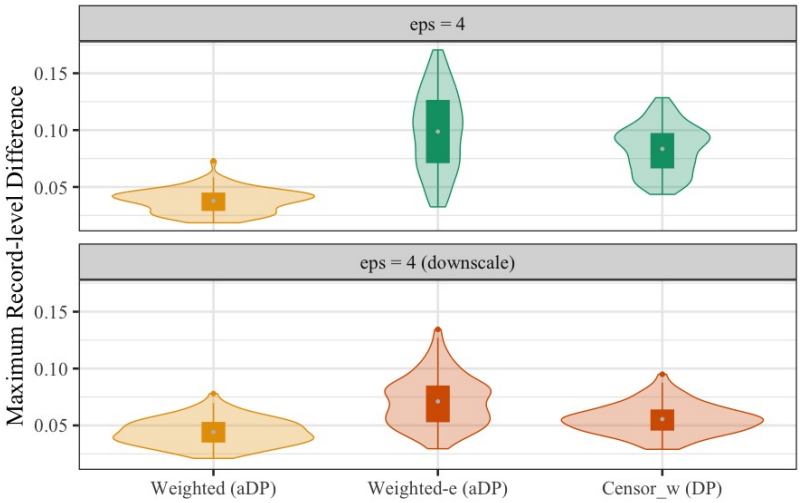


Figure 7: Violin plots of max-ECDF utility over $R = 100$ replicates, for the Weighted (aDP), the Weighted-e (aDP), and the Censor_w (DP) at $\epsilon = 4$, without downscaling (top) and with downscaling (bottom).

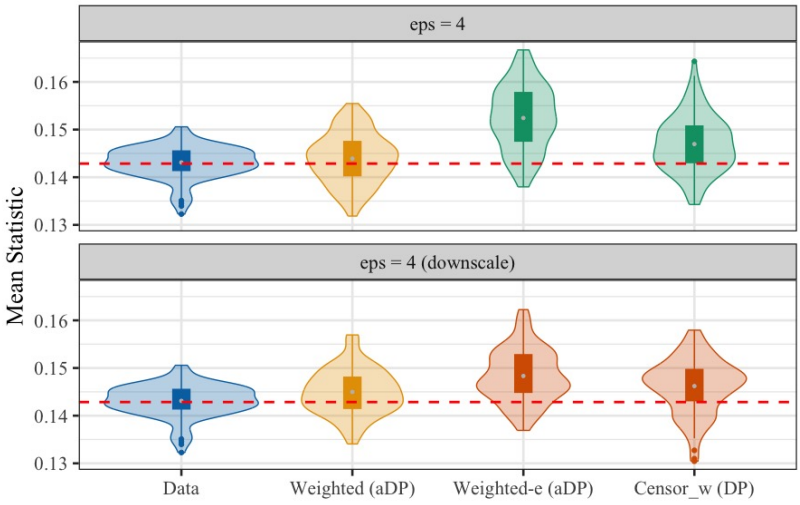


Figure 8: Violin plots of mean over $R = 100$ replicates for the Weighted (aDP), the Weighted-e (aDP), and the Censor_w (DP) at $\epsilon = 4$, without (top) and with (bottom) downscaling. A dashed horizontal line at the analytical mean from Beta(0.5, 3) is included in each panel.

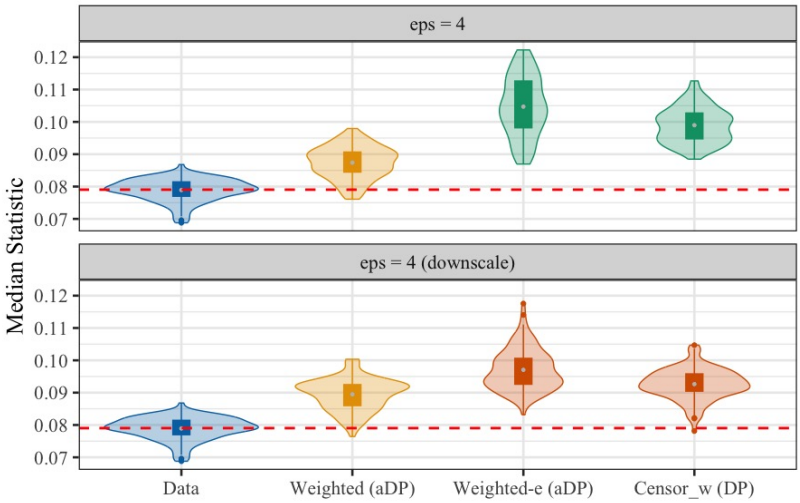


Figure 9: Violin plots of median over $R = 100$ replicates for the Weighted (aDP), the Weighted-e (aDP), and the Censor_w (DP) at $\epsilon = 4$, without downscaling (top) and with downscaling (bottom). A dashed horizontal line at the analytical median from Beta(0.5, 3) is included in each panel.

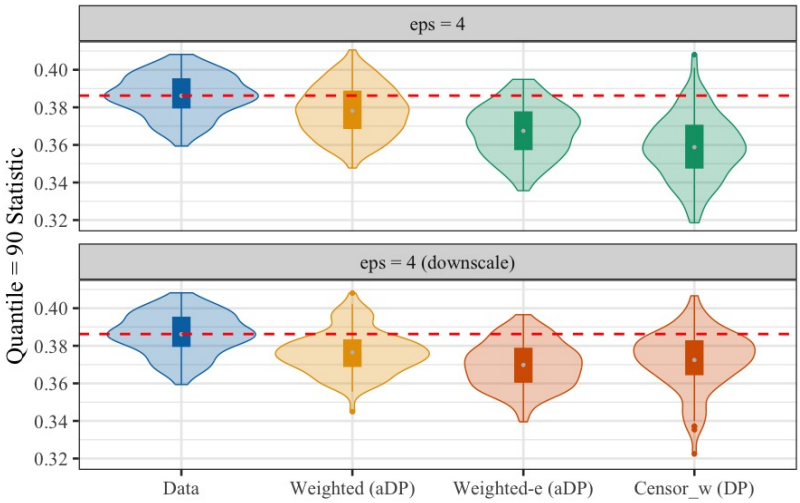


Figure 10: Violin plots of the 90th quantile over $R = 100$ replicates for the Weighted (aDP), the Weighted-e (aDP), and the Censor_w (DP) at $\epsilon = 4$, without downscaling (top) and with downscaling (bottom). A dashed horizontal line at the analytical 90th quantile from Beta(0.5, 3) is included in each panel.

## Bibliography

Dwork, C., McSherry, F., Nissim, K., and Smith, A. (2006). "Calibrating Noise to Sensitivity in Private Data Analysis." In *Proceedings of the Third Conference on Theory of Cryptography*, TCC'06, 265–284. Berlin, Heidelberg: Springer-Verlag.

Wasserman, L. and Zhou, S. (2010). "A Statistical Framework for Differential Privacy." *Journal of the American Statistical Association*, 105:

375–389.