Kanazawa University,
Faculty of Economics and Management

Discussion Paper Series

No. 74

# Differentially Private Sampling from a Finite Population

Nobuaki HOSHINO

hoshino@kenroku.kanazawa-u.ac.jp

March 30, 2023

金沢大学経済学経営学系
〒920-1192 金沢市角間町

**Faculty of Economics and Management, Kanazawa University**
Kakumamachi, Kanazawa-shi, Ishikawa, 920-1192, Japan

**https://keikei.w3.kanazawa-u.ac.jp/research_dp.html**

# Differentially Private Sampling from a Finite Population

Nobuaki Hoshino*

March 30, 2023

**Abstract**

Sampling publishes a part of data collected, by which the disclosure of information is limited. For this purpose, simple random sampling is usually employed. However, it does not satisfy differential privacy, which is a criterion to bound the accuracy of the estimation of a population. Hence the present study introduces a dummy to privatize random sampling. The same idea is applicable to even cluster sampling, which more effectively limits the disclosure. To cover various sampling methods, a general framework of differentially private sampling is constructed, based on the theory of the Bell polynomial distribution. An instance called quasi-multinomial sampling is shown to require less dummies, which do not grow as a sample size diverges. Therefore, it is suitable for publishing microdata of a large size with less distortion.

*Keywords: Confidentiality, Histogram publishing, Overdispersed multinomial, Synthetic data*

## 1 Introduction

A statistical agency publishes data of individuals, i.e., microdata, under the pledge of confidentiality. Sampling, sometimes called subsampling or resampling, aims to achieve this goal by publishing data of selected survey respondents, which are usually a part of the whole of survey respondents.

Sampling introduces uncertainty on the presence of an individual in a published data set, even if the individual is known to have been surveyed. The presence is regarded as a prerequisite of the identification of an individual; see Marsh et al. (1991). Hence sampling prevents definite identification. However, plausible identification may still be possible.

In the literature of statistical disclosure control or limitation, this plausibility is quantified as the accuracy of the estimation of population uniqueness, i.e., non-existence of another individual of the same attributes as that of a published record. Examples vary from initial Bethlehem et al. (1990) to highly computational Rocher et al. (2019), but most assume random sampling since deterministic sampling lacks a general assessment tool of disclosure. Commonly, simple random sampling with or without replacement is utilized to simplify the assessment.

However, under these simple sampling designs, the information of samples grows proportionally to a sample size, up to a finite population correction factor. Hence these designs are fundamentally unsuitable for publishing big data where confidentiality is of primary concern.

Among non-simple random sampling, cluster sampling can lower the accuracy of population estimation. Assuming that individuals of the same attributes belong to the same cell, allocating samples to fewer cells increases variances between cells, since variances within each cell are zero. Therefore, cluster sampling can reduce the information of samples, which is advantageous to limit disclosure.

Nevertheless, the performance of cluster sampling depends on the method of clustering. Because much generality exists in the method, we are not very aware of the advantage of cluster sampling. We need a framework of cluster sampling that can be theoretically elucidated.

*School of Economics, Kanazawa University, Kakuma-machi, Kanazawa 920-1192, Japan. E-mail: hoshino@kenroku.kanazawa-u.ac.jp

Hoshino (2021) proposes a family of generalized multinomial distributions, which generalizes simple random sampling with replacement to include cluster sampling. This family is called the Bell polynomial distribution (BPD), and known distributions such as the Dirichlet-multinomial mixture, i.e., negative hypergeometric distribution (Cheng Ping, 1964), are included. The Dirichlet-multinomial mixture demonstrates that the properties of the BPD are well parameterized. Therefore, the current study investigates the BPD as a tool for statistical disclosure control.

It is important that random sampling is an instance of random masking that can be assessed with differential privacy (Dwork et al., 2006b). Differential privacy is often interpreted as a criterion to bound the certainty of the presence of an individual in a published data set, which should motivate us to evaluate the BPD with differential privacy.

Indeed, we show that any BPD can be differentially private in the sense of Machanavajjhala et al. (2008). For example, they show that the Dirichlet-multinomial mixture can be differentially private. It implies that the accuracy of the estimation of a population frequency is controlled by a privacy budget $\epsilon$ of differential privacy. More precisely, the lower bound of the variance of the unbiased estimator of a population frequency is $(\exp(\epsilon)-1)^{-2}$; see Hoshino (2020, Theorem 2).

The Dirichlet-multinomial mixture, however, does not generate analytically useful data under a moderate privacy budget. Especially, the Dirichlet-multinomial mixture requires distortion proportional to the size of a published file, by which publishing big data is impractical. Hence Machanavajjhala et al. (2008) abandon the original differentialy privacy, and they allow "small" violation. Similarly, the U.S. census bureau employs approximate differential privacy (Dwork et al., 2006a), which also allows "small" violation, to publish useful 2020 census data. These examples suggest that the original differential privacy is hard to achieve in practice, where users of data do not accept much distortion.

Even under the original differential privacy, an instance of the BPD requires less distortion. We show that distortion necessary for sampling with the quasi-multinomial distribution (type 2) (Consul and Mittal, 1977) converges to a constant as the size of a published file grows when $\epsilon > 1$. Therefore, the quasi-multinomial distribution (type 2) is far less perturvative than the Dirichlet-multinomial distribution. In addition, moment formulae (Hoshino, 2021) and a sampling algorithm (Hu and Hoshino, 2018) are available for the quasi-multinomial distribution (type 2). It should be promising for publishing a large data set.

The remaining part of the present article is organized as follows. Section 2 consists of two subsections, where notation to use differential privacy and the Bell polynomial is prepared. Section 3 constructs the theory of differentially private sampling. The first subsection privatizes simple random sampling, where dummies are added to attain differential privacy. The second subsection employs the same idea, by which sampling with the general Bell polynomial distribution is privatized. Section 4 introduces sampling with the quasi-multinomial distribution as an instance of sampling with the Bell polynomial distribution. The superiority of the quasi-multinomial sampling is demonstrated both theoretically and empirically. Section 5 concludes with some discussion on the discrete Laplace distribution. All the proofs of theorems are collected in Appendix.

# 2 Preliminaries

Throughout the present article, we denote the set of positive integers and the set of nonnegative integers by $\mathbb{N}$ and $\mathbb{N}_0$, respectively. The set of integers is denoted by $\mathbb{Z}$. The set of positive real numbers is denoted by $\mathbb{R}_+$. Similarly, $\mathbb{R}_-$ denotes the set of negative real numbers, and $\mathbb{R}$ denotes the set of real numbers. Also, $\mathbb{R}_{0+} := \{0\} \cup \mathbb{R}_+$, and $\mathbb{R}_{0-} := \{0\} \cup \mathbb{R}_-$.

## 2.1 Differential privacy

As stated in Section 1, we employ the differential privacy of Machanavajjhala et al. (2008). This subsection defines it together with notation and motivation.

If the coding of attributes is common for all individuals, then a microdata set is equivalent to the frequencies of cells, where each cell corresponds to a unique combination of the values of attributes. For example, suppose that individuals are classified with respect to Sex and Height, where the coding of Sex is binary as $F$ or $M$ and the coding of Height is binary as $T$ or $S$. Then there are four cells, each of which is designated by one of $\{F, M\} \times \{T, S\}$. The set of frequencies $(1, 1, 0, 0)$ of the cells of $(FT, FS, MT, MS)$ is equivalent to a microdata set of $\{\{Sex = F, Height = T\}, \{Sex = F, Height = S\}\}$.

Therefore, we consider protecting the frequency vector of a population by publishing a frequency vector of samples as a sanitized data set. This situation looks similar to histogram publishing, but random sampling differs in that sample frequencies are always nonnetative integers and sum up to a fixed constant.

Let us denote the frequency vector of a population by $\vec{n} = (n_1, n_2, \ldots, n_J)$, where $J$ denotes the number of cells. We denote the size of a population by $n = \sum_{j=1}^{J} n_j$. We publish $\vec{m} = (m_1, m_2, \ldots, m_J)$, where $m_j$ denotes the sample frequency of the $j$th cell. Correspondingly, $m = \sum_{j=1}^{J} m_j$ denotes the size of a published file.

The space of a frequency vector of distributing $n$ individuals over $J$ cells is denoted by

$$\mathcal{F}_{n,J} := \{\boldsymbol{f}_J | f_j \in \mathbb{N}_0, j \in [J], \sum_{j=1}^{J} f_j = n\}.$$

The size of a population, $n$, is fixed under sampling from a finite population. This situation is different from standard differential privacy that conceals the addition or deletion of one individual. It is more natural to conceal the move of one individual to a different cell, as originally considered by Dwork et al. (2006b).

Let $\vec{n}' = (n'_1, n'_2, \ldots, n'_J)$ denote the result of moving one individual of $\vec{n}$ to a different cell. When one individual of the $k$th cell moves to the $j$th cell, $j \neq k$, it results in that $n'_j = n_j + 1, n'_k = n_k - 1$. Specifically, we adopt differential privacy defined below.

**Definition 1 (Machanavajjhala et al., 2008)** *Let $(m, n, J) \in \mathbb{N}^3$ and $\epsilon \in \mathbb{R}_+$. A random mask that generates $\vec{m}$ is $\epsilon$-differentially private ($\epsilon$-DP) if and only if for all $(\vec{m}, \vec{n}, \vec{n}') \in \mathcal{F}_{m,J} \times \mathcal{F}_{n,J} \times \mathcal{F}_{n,J}$,*

$$\mathrm{P}(\vec{m}; \vec{n})/\mathrm{P}(\vec{m}; \vec{n}') \leq \exp(\epsilon). \tag{1}$$

## 2.2 Bell polynomials

The current study considers random sampling induced by the Bell polynomial distribution (Hoshino, 2021). This family of distributions is named after the Bell polynomials (Comtet, 1974), which appear in the probility mass function. For a reader who is unfamiliar with the Bell polynomial, we review it in this subsection.

For an infinite sequence of real numbers $\boldsymbol{w} := (w_1, w_2, \ldots)$, $n \in \mathbb{N}$ and $k \in [n] := \{1, 2, \ldots, n\}$, the partial Bell polynomial is defined as

$$B_{n,k}(\boldsymbol{w}) := n! \sum_{\boldsymbol{s} \in \mathcal{P}_{n,k}} \prod_{i=1}^{n} \Big(\frac{w_i}{i!}\Big)^{s_i} \frac{1}{s_i!},$$

where

$$\mathcal{P}_{n,k} := \{(s_1, \ldots, s_n) : s_i \in \mathbb{N}_0, i \in \mathbb{N}, \sum_{i=1}^{n} i s_i = n, \sum_{i=1}^{n} s_i = k\}.$$

Considering this definition, $B_{n,1}(\boldsymbol{w}) = w_n$, and $B_{n,n}(\boldsymbol{w}) = {w_1}^n$.

The (total) Bell polynomial $B_n(\boldsymbol{w})$ is defined as the sum of partial Bell polynomials over $k$. Namely,

$$B_n(\boldsymbol{w}) := n! \sum_{\boldsymbol{s} \in \mathcal{P}_n} \prod_{i=1}^{n} \Big(\frac{w_i}{i!}\Big)^{s_i} \frac{1}{s_i!},$$

where

$$\mathcal{P}_n := \cup_{k=1}^{n} \mathcal{P}_{n,k}.$$

It is important that $B_0(\boldsymbol{w}) = B_{0,0}(\boldsymbol{w}) \equiv 1$ for all $\boldsymbol{w}$; see, e.g., Riordan (1958). It has to be $B_{n,0}(\boldsymbol{w}) = 0$ for $n \in \mathbb{N}$.

We extensively use the following expression:

$$B_n(\lambda, \boldsymbol{w}) := \sum_{k=1}^{n} \lambda^k B_{n,k}(\boldsymbol{w}). \tag{2}$$

According to, e.g., Charalambides (2002, eq. 11.15),

$$B_n(\lambda, \boldsymbol{w}) = B_n(\lambda w_1, \lambda w_2, \ldots, \lambda w_n).$$

For computation, the following recurrence formula is convenient.

$$B_n(\lambda, \boldsymbol{w}) = \sum_{i=0}^{n-1} \binom{n-1}{i} B_i(\lambda, \boldsymbol{w}) \lambda w_{n-i};$$

see, e.g., Charalambides (2002, eq. 11.10).

# 3 Theory of Differentially Private Sampling

## 3.1 Simple random sampling

We begin with considering simple random sampling without replacement or the hypergeometric distribution. This random mask is defined by

$$\mathrm{P}(\vec{m}; \vec{n}) = \binom{n_1}{m_1} \binom{n_2}{m_2} \cdots \binom{n_J}{m_J} \Big/ \binom{n}{m}. \tag{3}$$

Simple random sampling without replacement (3) can not be $\epsilon$-DP because $\mathrm{P}(\vec{m}; \vec{n})/\mathrm{P}(\vec{m}; \vec{n'})$ is not finite when $\mathrm{P}(\vec{m}; \vec{n'}) = 0$. For example, let $n_k = 1$. Moving one individual of the corresponding population from the $k$th cell to the $j$th cell results in $n'_k = 0$. Then $\mathrm{P}(\vec{m}; \vec{n'}) = 0$ when $m_k \geq 1$.

Generally, any random sampling is not $\epsilon$-DP when the support of $\vec{m}$ depends on $\vec{n}$; see Hoshino (2020, Remark 1). To eliminate this dependence under simple random sampling without replacement, we may add $\gamma_j$ dummy individuals to the $j$th cell of the population, where $\gamma_j \geq m, j \in [J]$. Write $\vec{\gamma} = (\gamma_1, \gamma_2, \ldots, \gamma_J)$. Then

$$\mathrm{P}(\vec{m}; \vec{n}, \vec{\gamma}) = \binom{n_1 + \gamma_1}{m_1} \binom{n_2 + \gamma_2}{m_2} \cdots \binom{n_J + \gamma_J}{m_J} \Big/ \binom{n + \gamma_\cdot}{m}, \tag{4}$$

where $\gamma_\cdot := \sum_{j=1}^{J} \gamma_j$. Sampling by (4) can be $\epsilon$-DP since the support of $\vec{m}$ is not restricted by $\vec{n}$.

By introducing dummies, samples should consist of both real and imaginary individuals. This mixture is called hybrid in the literature of synthetic data. As the number of dummies $\gamma_j$ increases, samples should have less information about a population. This intuition is supported by the condition of differential privacy derived in Theorem 1.

Actually, $\gamma_j$ is not necessarily a nonnegative integer. To understand this fact, we rewrite (4) as

$$\mathrm{P}(\vec{m}; \vec{n}, \vec{\gamma}) = \binom{m}{\vec{m}} \frac{\Gamma(n - m + \gamma_\cdot + 1)}{\Gamma(n + \gamma_\cdot + 1)} \prod_{j=1}^{J} \frac{\Gamma(n_j + \gamma_j + 1)}{\Gamma(n_j + \gamma_j - m_j + 1)}. \tag{5}$$

**Theorem 1** *Hypergeometric sampling (4) or (5) is $\epsilon$-DP if and only if*

$$\min_j \gamma_j \geq m - 1 + \frac{m}{\exp(\epsilon) - 1}. \tag{6}$$

In (6), $\min_j \gamma_j$ has to diverge as $\epsilon \to 0$. Also $\min_j \gamma_j$ has to be $O(m)$ as $m$ grows. These facts suggest that hypergeometric sampling is not very promising unless $n \gg m$.

The difficulty of simple random sampling without replacement has been the dependence of the support of $\vec{m}$ on $\vec{n}$. Sampling with replacement avoids this dependence mostly, except for empty cells of a population (structural zeros). To remedy this issue, we again add $\gamma_j$ dummies to the $j$th cell of a population. Accordingly, simple random sampling with replacement or the multinomial distribution is defined by

$$\mathrm{P}(\vec{m}; \vec{n}, \vec{\gamma}) = \binom{m}{\vec{m}} \prod_{j=1}^{J} \left( \frac{n_j + \gamma_j}{n + \gamma_\cdot} \right)^{m_j}. \tag{7}$$

We are familiar with the interpretation of (7) that $(n_j + \gamma_j)/(n + \gamma_\cdot)$ is the probability of the $j$th cell to be sampled. This interpretation liberates us from regarding $\gamma_j$ as a nonnegative integer. A dummy is not necessarily an imaginary individual. It is generally a sampling weight to allow unequal probability sampling. Oversampling and downsampling are often employed in sample surveys of official statistics, where the inclusion probabilities of microdata should not be neglected. Introduced $\gamma_j$ can describe differences in inclusion probabilities.

**Theorem 2** *Multinomial sampling (7) is $\epsilon$-DP if and only if*

$$\min_j \gamma_j \geq \frac{1}{\exp(\epsilon/m) - 1}. \tag{8}$$

The right hand side of (8) is approximately $m/\epsilon$. Therefore, necessary dummies are reduced from hypergeometric sampling, but they are still $O(m)$. The essential reason of dummies to grow is that multinomial sampling (7) is equivalent to the convolution of $m$ independent samples. In other words, the Fisher information is proportional to $m$. We need to consider dependent samples for further reduction of dummies.

## 3.2 Sampling with Bell polynomial distributions

Considering simple random sampling in the previous subsection clarifies two facts: (i) Sampling with replacement reduces necessary dummies. (ii) Unequal probability sampling is beneficial, but insufficient for reducing dummies. Observing these facts, we should consider cluster sampling with replacement that allows unequal inclusion probabilities.

Sampling with replacement is equivalent to a distribution over the same support as that of the multinomial distribution. The Bell polynomial distribution (BPD) is a family of such distributions with parameters to control inclusion probabilities. It also possesses a prameter to increase marginal variances, which is the effect of clustering. In fact, the BPD is an overdispersed multinomial distribution (Neerchal and Morel, 2005).

We define BPD sampling below so that samples are subject to the BPD. Hoshino (2021) shows that if $\boldsymbol{w} \in \mathbb{R}_+ \times {\mathbb{R}_{0+}}^{\infty}$ and $n_j + \gamma_j > 0, j \in [J]$, then $\mathrm{P}(\vec{m}) > 0$ for all $\vec{m} \in \mathcal{F}_{m,J}$ in (9). Hence there is a chance to be $\epsilon$-DP, for which $n_j + \gamma_j > 0$ must hold for any $n_j$. Thus we restrict $\gamma_j \in \mathbb{R}_+$.

**Definition 2** *BPD sampling under a dummy vector $\vec{\gamma} := (\gamma_1, \gamma_2, \ldots, \gamma_J) \in {\mathbb{R}_+}^J$ and a characteristic sequence $\boldsymbol{w} \in \mathbb{R}_+ \times {\mathbb{R}_{0+}}^{\infty}$ is defined by*

$$\mathrm{P}(\vec{m}; \vec{n}, \vec{\gamma}, \boldsymbol{w}) = \binom{m}{\vec{m}} \frac{1}{B_m(n + \gamma_\cdot, \boldsymbol{w})} \prod_{j=1}^{J} B_{m_j}(n_j + \gamma_j, \boldsymbol{w}). \tag{9}$$

Multinomial sampling is a special case of BPD sampling; putting $w_1 = 1, w_i = 0, i \geq 2$ into (9) results in (7). Also for any $\boldsymbol{w}$ where (9) depends on $n + \gamma_\cdot$, BPD sampling (9) converges in distribution to multinomial sampling as $(n + \gamma_\cdot) \to \infty$ when $(n_j + \gamma_j)/(n + \gamma_\cdot)$ is fixed; see Hoshino (2021).

For convenience, below we cite the moment properties of BPD samples. The mean vector and the correlation matrix of BPD sampling are the same as those of multinomial sampling, regardless of $\boldsymbol{w}$.

**Proposition 1 (Hoshino, 2021)** *Suppose that $\vec{m}$ is generated by (9). Write $\pi_j = (n_j + \gamma_j)/(n + \gamma_\cdot)$. Then for $m \geq 2$,*

$$\mathrm{E}(m_j) = m\pi_j, \quad j \in [J]. \tag{10}$$

$$\mathrm{V}(m_j) = m\pi_j(1 - \pi_j)\phi(m, n + \gamma_\cdot, \boldsymbol{w}), \quad j \in [J], \tag{11}$$

*where*

$$\phi(m, n + \gamma_\cdot, \boldsymbol{w}) = 1 + \frac{(n + \gamma_\cdot)(m-1)!}{B_m(n + \gamma_\cdot, \boldsymbol{w})} \sum_{i=0}^{m-2} \frac{B_i(n + \gamma_\cdot, \boldsymbol{w}) w_{m-i}}{i!(m - i - 2)!}.$$

$$\mathrm{Cov}(m_i, m_j) = -m\pi_i\pi_j\phi(m, n + \gamma_\cdot, \boldsymbol{w}), \quad i \in [J], j \in [J], i \neq j.$$

Data users are often interested in the $j$th cell probability $n_j/n$, and the bias of its usual estimator $m_j/m$ under BPD sampling is

$$\mathrm{E}\left(\frac{m_j}{m}\right) - \frac{n_j}{n} = \frac{n\gamma_j - n_j\gamma_\cdot}{n(n + \gamma_\cdot)} \tag{12}$$

from (10). This result provides intuitive understanding of the effect of dummies. If $\gamma_j$ is proportional to $n_j$ for all $j$ then $m_j/m$ is an unbiased estimator of the $j$th cell probability. This way generates useful data, but it is not $\epsilon$-DP since empty cells need positive dummies to exclude the case of $\mathrm{P}(\vec{m}) = 0$. Also differential privacy does not seem to suppose dummies depending on $n_j$.

Below we provide a sufficient condition for BPD sampling to be $\epsilon$-DP.

**Theorem 3** *Any BPD sampling (9) is $\epsilon$-DP if*

$$\min_j \gamma_j \geq \frac{1}{\exp(\epsilon/m) - 1}. \tag{13}$$

The sufficient condition (13) equals to the necessary and sufficient condition (8) of the multinomial distribution. This result corresponds to the fact that the multinomial distribution has the least variance among the BPD; it is straightforward to verify that $\phi(m, n + \gamma_\cdot, \boldsymbol{w}) \geq 1$ in (11), and $\phi(m, n + \gamma_\cdot, (1, 0, 0, \ldots)) = 1$.

Necessary dummies are at most (13) for BPD sampling. Therefore, hypergeometric sampling does not belong to BPD sampling because it needs more dummies than multinomial sampling. Selecting appropriate $\boldsymbol{w}$ results in less dummies for differential privacy.

To simplify the necessary and sufficient condition for BPD sampling to be $\epsilon$-DP, we restrict $\boldsymbol{w}$ so that the Bell polynomial is monotone in two senses:

$$\mathcal{W}_1 := \left\{ \boldsymbol{w} : \boldsymbol{w} \in \mathbb{R}_+ \times \mathbb{R}_{0+}{}^{\infty}, \frac{B_{n+1}(\lambda + 1, \boldsymbol{w})}{B_{n+1}(\lambda, \boldsymbol{w})} \geq \frac{B_n(\lambda + 1, \boldsymbol{w})}{B_n(\lambda, \boldsymbol{w})}, \forall n \in \mathbb{N}, \forall \lambda \in \mathbb{R}_+ \right\},$$

and

$$\mathcal{W}_2 := \left\{ \boldsymbol{w} : \boldsymbol{w} \in \mathbb{R}_+ \times \mathbb{R}_{0+}{}^{\infty}, \frac{d\{B_n(\lambda + 1, \boldsymbol{w})/B_n(\lambda, \boldsymbol{w})\}}{d\lambda} \leq 0, \forall n \in \mathbb{N}, \forall \lambda \in \mathbb{R}_+ \right\}.$$

**Theorem 4** *Suppose that $\boldsymbol{w} \in (\mathcal{W}_1 \cap \mathcal{W}_2)$. Then BPD sampling (9) is $\epsilon$-DP if and only if*

$$\frac{B_m(1 + \min_j \gamma_j, \boldsymbol{w})}{B_m(\min_j \gamma_j, \boldsymbol{w})} \leq \exp(\epsilon). \tag{14}$$

For example, let $w_i = (i-1)!$. Then it is widely known that $B_n(\lambda, (0!, 1!, 2!, \ldots)) = \lambda(\lambda+1)(\lambda+2)\cdots(\lambda+n-1)$; see Pitman (2006, eq. 1.54). Then $B_n(1+\lambda, \boldsymbol{w})/B_n(\lambda, \boldsymbol{w}) = (\lambda+n)/\lambda$, which is increasing with respect to $n$ for $\lambda \in \mathbb{R}_+$. Also $B_n(1+\lambda, \boldsymbol{w})/B_n(\lambda, \boldsymbol{w}) = (\lambda+n)/\lambda$ is decreasing with respect to $\lambda$ for $n \in \mathbb{N}$. Therefore, Theorem 4 reproduces the following known result.

**Corollary 1 (Machanavajjhala et al., 2008)** *Negative hypergeometric sampling defined by*

$$\mathrm{P}(\vec{m}; \vec{n}, \vec{\gamma}) = \binom{m}{\vec{m}} \frac{\Gamma(n+\gamma_\cdot)}{\Gamma(n+\gamma_\cdot+m)} \prod_{j=1}^{J} \frac{\Gamma(n_j+\gamma_j+m_j)}{\Gamma(n_j+\gamma_j)} \tag{15}$$

*is $\epsilon$-DP if and only if*

$$\min_j \gamma_j \geq \frac{m}{\exp(\epsilon)-1}. \tag{16}$$

In (13), $\min_j \gamma_j \approx m/\epsilon$. Hence necessary dummies for negative hypergeometric sampling are less than those of multinomial sampling as $m \to \infty$, yet the difference disappears as $\epsilon \to 0$. Also the necessary dummies of negative hypergeometric sampling are still $O(m)$, which is distortive for publishing big data.

We would like to find better $\boldsymbol{w}$ that requires less dummies. The next section presents such an example, for which we provide a sufficient condition for $\boldsymbol{w}$ to belong to $\mathcal{W}_1$.

**Theorem 5** *Let $\boldsymbol{w} \in \mathbb{R}_+ \times \mathbb{R}_{0+}{}^{\infty}$. If $B_{n+1,k}(\boldsymbol{w})/B_{n,k}(\boldsymbol{w})$ is nondecreasing with respect to $k \in [n]$ for $n \in \mathbb{N}$ then $\boldsymbol{w} \in \mathcal{W}_1$.*

Even if the total Bell polynomial is not simple enough to verify the monotone condition of $\mathcal{W}_1$, the partial Bell polynomial can be simple enough. For example, let $w_i = i$. Then $B_{n,k}(1, 2, \ldots) = \binom{n}{k} k^{n-k}$, which is called an idempotent number; see Harris and Schoenfeld (1967). The corresponding total Bell polynomial has not been expressed simpler than the weighted sum of the partial Bell polynomials, but $B_{n+1,k}(1, 2, \ldots)/B_{n,k}(1, 2, \ldots) = (n+1)k/(n-k+1)$, which is apparently increasing with respect to $k$. Hence $(1, 2, \ldots) \in \mathcal{W}_1$.

Hoshino (2021) discusses some properties of the BPD that are useful for drawing samples. For example, the computer generation of BPD samples enjoys the conditional distribution method (Devroye, 1986), where sequential sampling of each $m_j$ is justified. Based on these results, the author recommends Algorithm 1 to generate BPD samples. Hoshino (2021) also clarifies the stopping time $T := \min\{t : \sum_{j=1}^{t} m_j = m\}$ as $\mathrm{P}(T \leq t) = B_m(\sum_{j=1}^{t} \lambda_j + \gamma_j, \boldsymbol{w})/B_m(\lambda_\cdot + \gamma_\cdot, \boldsymbol{w})$ under Algorithm 1.

**Algorithm 1** *The following procedure generates $\vec{m}$ subject to (9).*

1. *Order the index of cells so that $n_1 + \gamma_1 \geq n_2 + \gamma_2 \geq n_3 + \gamma_3 \geq \cdots$.*
2. *Let $\vec{m} = (0, 0, \ldots)$ and $j = 1$.*
3. *Given $(m_1, m_2, \ldots, m_{j-1})$, draw $m_j$ subject to the following distribution:*

$$\mathrm{P}(m_j) = \binom{m - \sum_{i=1}^{j-1} m_i}{m_j} \frac{B_{m_j}(n_j+\gamma_j, \boldsymbol{w}) B_{m-\sum_{i=1}^{j} m_i}(\sum_{i=j+1}^{J} n_i + \gamma_i, \boldsymbol{w})}{B_{m-\sum_{i=1}^{j-1} m_i}(\sum_{i=j}^{J} n_i + \gamma_i, \boldsymbol{w})},$$

   *where $m_j = 0, 1, \ldots, m - \sum_{i=1}^{j-1} m_i$.*

4. *If $\sum_{i=1}^{j} m_i = m$ then output $\vec{m}$ and quit.*
5. *Increase $j$ by one; go to Step 3.*

Table 1: Minimum $\gamma$ given $(m, \epsilon)$

| | $\epsilon$ | | | |
|---|---|---|---|---|
| $m$ | 1 | 2 | 3 | 4 |
| 100 | 9.50 | .564 | .154 | .0516 |
| 1 000 | 31.1 | .580 | .156 | .0523 |
| 10 000 | 99.5 | .582 | .156 | .0524 |
| 100 000 | 316 | .582 | .157 | .0524 |
| 100 000 000 | 9999 | .582 | .157 | .0524 |
| 1 000 000 000 | 31574 | .582 | .157 | .0524 |
| $\infty$ | $\infty$ | .582 | .157 | .0524 |

| | $\epsilon$ | | | | | |
|---|---|---|---|---|---|---|
| $m$ | 1/2 | 1/3 | 1/4 | 1/5 | $\cdots$ | 1/10 |
| 100 | 102 | 201 | 301 | 401 | $\cdots$ | 901 |
| 1000 | 1002 | 2001 | 3001 | 4001 | $\cdots$ | 9001 |

# 4 Quasi-multinomial Sampling

In this section, we examine a promising instance of BPD sampling: Quasi-multinomial sampling. Let $w_i = i^{i-1}$. Then $B_{n,k}(1^0, 2^1, \ldots) = \binom{n-1}{k-1} n^{n-k}$, and $B_n(\lambda, (1^0, 2^1, \ldots)) = \lambda(\lambda + n)^{n-1}$; see Pitman (2006, eq. 1.56). The corresponding instance of the BPD is the quasi-multinomial distribution (type 2) proposed by Consul and Mittal (1977); see also Hoshino (2021).

We can confirm that $(1^0, 2^1, \ldots) \in \mathcal{W}_1$ by Theorem 5. Write $B_{n+1,k}(1^0, 2^1, \ldots)/B_{n,k}(1^0, 2^1, \ldots) = (n+1)^{n-k+1}/(n^{n-k-1}(n-k+1)) = f(k) > 0$. Then $f(k+1)/f(k) = (n^2-(k-1)n)/(n^2-(k-1)n-k) > 1$. Hence $f(k)$ is increasing with respect to $k \in [n]$, which suffices.

Also $B_n(\lambda+1, (1^0, 2^1, \ldots))/B_n(\lambda, (1^0, 2^1, \ldots)) = (\lambda+1)/\lambda\{(\lambda+1+n)/(\lambda+n)\}^{n-1}$, which is obviously decreasing with respect to $\lambda$. Consequently, the following result holds as a corollary to Theorem 4.

**Corollary 2** *Quasi-multinomial sampling defined by*

$$\mathrm{P}(\vec{m}; \vec{n}, \vec{\gamma}) = \binom{m}{\vec{m}} \frac{\prod_{j=1}^{J} (n_j + \gamma_j)(n_j + \gamma_j + m_j)^{m_j - 1}}{(n + \gamma_\cdot)(n + \gamma_\cdot + m)^{m-1}} \tag{17}$$

*is $\epsilon$-DP if and only if*

$$\left(1 + \frac{1}{\min_j \gamma_j}\right) \left(1 + \frac{1}{\min_j \gamma_j + m}\right)^{m-1} \leq \exp(\epsilon). \tag{18}$$

The necessary dummy in (18) is implicit. Hence for some combinations of $m$ and $\epsilon$, the minimum $\min_j \gamma_j$ is numerically evaluated using Mathematica. Table 1 summarizes the result. These values are consistent with the following asymptotic formulae. The proof of Theorem 6 is easy and omitted.

**Theorem 6** *As $m \to \infty$, the equality in (18) holds when*

$$\min_j \gamma_j \approx \begin{cases} 1/(\exp(\epsilon) - 1) - 1, & \epsilon > 1, \\ \sqrt{m}, & \epsilon = 1, \\ (1/\epsilon - 1)m, & \epsilon < 1. \end{cases}$$

Theorem 6 shows that $\min_j \gamma_j = O(1)$ when $\epsilon > 1$. Hence quasi-multinomial sampling is very appropriate for publishing big data under a usual privacy budget.

To demonstrate the superiority of quasi-multinomial sampling, we compare sampling methods under the same privacy budget: $\epsilon = 7$, which is the choice of Machanavajjhala et al. (2008). Let $m = n = J$ be one million, and the dummy of each cell is set at the minimum. Assuming that $n_j$ is ten thousand, we evaluate the expectation of $m_j = \hat{n_j}$ under each sampling. Table 2 exemplifies a small magnitude of bias by quasi-multinomial sampling.

One should also be concerned about the variance of quasi-multinomial sampling. The variance formula (11) shows that quasi-multinomial sampling inflates the variance of multinomial sampling by $\phi(m, n + \gamma_\cdot, \boldsymbol{w})$. Hoshino (2021) shows that $1 \leq \phi(m, \lambda, \boldsymbol{w}) < m$ for $\lambda \in \mathbb{R}_+$. Although exceptions such as

Table 2: Bias of Differentially Private Sampling ($n_j = 10000$)

| | Hypergeometric | Multinomial | Negative Hypergeometric | Quasi-multinomial |
|---|---|---|---|---|
| Minimum $\gamma$ | 1000912 | 142857 | 914 | .00248 |
| $\mathrm{E}(\hat{n_j})$ | 1.01 | 1.07 | 11.9 | 9975.2 |

Table 3: Variance Inflation by Quasi-multinomial Sampling

Exact: $\phi(1000, J\gamma, \boldsymbol{w}) - 1$

| | $J$ | | | |
|---|---|---|---|---|
| $\gamma$ | 100 | 1000 | 10000 | 100000 |
| $\sqrt{10}$ | 15.7 | .731 | .0642 | .00633 |
| 10 | 2.98 | .210 | .0201 | .00200 |
| $\sqrt{1000}$ | .731 | .0642 | .00633 | .000632 |
| 100 | .210 | .0201 | .00200 | .000200 |
| 1000 | .0201 | .00200 | .000200 | .0000200 |

Approximate: $2(1000 - 1)/(J\gamma)$

| | $J$ | | | |
|---|---|---|---|---|
| $\gamma$ | 100 | 1000 | 10000 | 100000 |
| $\sqrt{10}$ | 6.32 | .632 | .0632 | .00632 |
| 10 | 2.00 | .200 | .0200 | .00200 |
| $\sqrt{1000}$ | .632 | .0632 | .00632 | .000632 |
| 100 | .200 | .0200 | .00200 | .000200 |
| 1000 | .0200 | .00200 | .000200 | .0000200 |

the multinomial distribution exist, $\phi(m, \lambda, \boldsymbol{w}) = 1 + O(\lambda^{-1})$ as $\lambda \to \infty$, and $\phi(m, \lambda, \boldsymbol{w}) = m - O(\lambda)$ as $\lambda \to 0$, In practice, $n + \gamma_\cdot$ is not close to zero because $n$ is a population size, and $\gamma_\cdot = O(J)$, where the number of cells $J$ is usually large. Hence we numerically evaluate an asymptotic formula $\phi(m, \lambda, \boldsymbol{w}) \approx 1 + (m - 1)w_2/(\lambda w_1^2)$ as $\lambda \to \infty$.

Variance formulae (11) specific to quasi-multinomial sampling are provided by Hoshino (2021). However, those are involved with the factorial of $m$, which is not convenient for exact computation. Hence we select rather small $m = 1000$. The dummy of each cell equals to $\gamma$. Then $\gamma_\cdot$ amounts to $J\gamma$, where $J$ is the number of cells. Typically, $n$ is greater than $m$, but we know that a large value of $n$ decreases $\phi(m, n + J\gamma, \boldsymbol{w})$. Hence we let $n = 0$. In the case of quasi-multinomial sampling, $w_1 = 1$ and $w_2 = 2$. Therefore, we compare $\phi(1000, J\gamma, (1^0, 2^1, \ldots)) - 1 \approx 2(m - 1)/(J\gamma)$ with the exact value.

Table 3 exhibits the result. Although tabulated values are supplied for each combination of $J$ and $\gamma$, the same value of $J\gamma$ leads to the same result. The approximation seems acceptable when $J\gamma > 1000$. The case of $\phi(1000, 1000, (1^0, 2^1, \ldots)) \doteq 3.98$ is arguable; when a sampling fraction is close to one, the standard deviation of a sample frequency is almost doubled under few dummies. Increasing dummies can reduce it, which implies a tradeoff between bias and variance. Nevertheless, to set $\gamma_j$ proportional to $n_j$ for positive $n_j$ can reduce both bias and variance. Decreasing a sampling fraction also reduces the standard deviation of a sample frequency. It seems reasonable to accept $\phi(1000, 1000\sqrt{10}, (1^0, 2^1, \ldots)) \doteq 1.73$ for a price of differential privacy.

The computer generation of quasi-multinomial samples can employ Algorithm 1. Hu and Hoshino (2018) describe it specifically to quasi-multinomial sampling, where Step 3 of Algorithm 1 exploits rejection sampling that proposal obeys the beta-binomial mixture distribution.

# 5 Concluding Discussion

Let us compare random sampling with the discrete Laplace distribution (Inusah and Kozubowski, 2006), which is a popular differentially private method used to protect count data. This method publishes $m_j = n_j + x_j$, where $x_j$ is an independent random noise identically distributed as

$$\mathrm{P}(x_j = x; \epsilon) = \frac{1 - e^{-\epsilon/2}}{1 + e^{-\epsilon/2}} e^{-\epsilon|x|/2}, \quad x \in \mathbb{Z}. \tag{19}$$

Then puslished $\vec{m}$ is $\epsilon$-DP. Also this method publishes unbiased data, i.e., $\mathrm{E}(m_j) = n_j$, since (19) is an even function. Inusah and Kozubowski (2006) show that $\mathrm{V}(x_j) = 2e^{-\epsilon/2}/(1 - e^{-\epsilon/2})^2$. Table 4 enumerates the variance and the probability of $x_j$ to be negative for some $\epsilon$.

Table 4: Discrete Laplace Noise

| $\epsilon$ | 1/2 | 1 | 2 | 3 |
|---|---|---|---|---|
| $\mathrm{V}(x_j)$ | 31.8 | 7.84 | 1.84 | .739 |
| $\mathrm{P}(x_j < 0)$ | .438 | .378 | .269 | .182 |

The major disadvantages of adding the discrete Laplace noise are (i) $m$ is random and restricted so that $\mathrm{E}(m) = n$, and (ii) $m_j$ can be negative. First, the randomness of $m$ may not be problematic when $n$ is unknown and to be protected. However, $n$ is often known in official statistics. For example, the U.S. census bureau must unprotect the population of a state, $n$, by law. Second, negative frequencies are unacceptable in publishing microdata. Hence some post-processing to obliterate negative frequencies is employed under adding noise. However, post-processing is often a type of deterministic optimization, which seems too untractable to assure that post-processed data are valid for statistical analyses.

Random sampling is free from these two issues. Obviously, $m$ is deterministic under sampling, where we do not have to equate $m$ to $n$. We can even let $m > n$ under random sampling with replacement; it may be worthy of note that bootstrapping is equivalent to simple random sampling with replacement. The second issue of negativity never exists in sampling. However, bias and variance may be worse. The variance of multinomial sampling equals $\mathrm{V}(m_j) = m\pi_j(1-\pi_j)$. The solution to $m\pi(1-\pi) \leq \mathrm{V}(x_j)$ does not exist when $m > 4\mathrm{V}(x_j)$. From Table 4, we understand that the discrete Laplace noise is usually smaller than multinomial sampling. Nevertheless, post-processed data do not necessarily behave well. If a population contains many cells of frequency 0 and 1 (i.e., sparse) then heavy post-processing inevitably follows from the discrete Laplase noise. Random sampling can be preferable because of no need for post-processing.

Sparsity arises from $J \gg n$. Hence the discrete Laplace noise is disadvantageous for large $J$. Moreover, adding independent noise to each cell accumulates proportionally to $J$, which is also the disadvantage of the discrete Laplace noise. By contrast, not $J$ but $m$ samples are independent in multinomial sampling. Hence noise accumulation is relaxed under multinomial sampling when $J \gg m$. Cluster sampling further reduces noise accumulation due to the dependence of samples.

In conclusion, quasi-multinomial sampling is theoretically more scalable than known methods to publish microdata of large $J$ and $m$ under rigorous, not approximate, protection of differential privacy.


# Acknowledgements

The author's research has been financially supported by JSPS KAKENHI grant.


# Appendix—Proofs

In the proofs of differential privacy, suppose that one individual of the $k$th cell moves to the $j$th cell, where $k \neq j$. The indicies of $J$ cells can be permutated, and thus the following argument fixes $(j, k)$ without loss of generality. It is worthy of note that $n_k \geq 1$ since no individual can move when $n_k = 0$.

## Proof of Theorem 1

We rewrite the definition (1) of differential privacy under (5) as

$$\begin{aligned} \frac{\mathrm{P}(\vec{m};\vec{n})}{\mathrm{P}(\vec{m};\vec{n}')} &= \frac{\Gamma(n_j+\gamma_j+1)}{\Gamma(n_j+\gamma_j-m_j+1)}\frac{\Gamma(n_j+\gamma_j-m_j+2)}{\Gamma(n_j+\gamma_j+2)}\frac{\Gamma(n_k+\gamma_k+1)}{\Gamma(n_k+\gamma_k-m_k+1)}\frac{\Gamma(n_k+\gamma_k-m_k)}{\Gamma(n_k+\gamma_k)} \\ &= \frac{n_j+\gamma_j-m_j+1}{n_j+\gamma_j+1}\frac{n_k+\gamma_k}{n_k+\gamma_k-m_k} \\ &\leq \exp(\epsilon). \end{aligned} \tag{20}$$

This inequality must hold for all $(\vec{m}, \vec{n}, \vec{n}')$. Hence we maximize (20) within the support of $(\vec{m}, \vec{n}, \vec{n}')$. Regarding (20) as the product of two ratios, the first ratio is maximized to unity when $m_j = 0$ regardless of $n_j$. The second ratio is maximized when $m_k = m$ and $n_k = 1$ since $m_k \leq m$ and $n_k \geq 1$. We note that (20) increases as $\gamma_k$ decreases. Therefore, the definition of differential privacy (1) reduces to $(1 + \min_j \gamma_j)/(1 + \min_j \gamma_j - m) \leq \exp(\epsilon)$. This inequality is equivalent to (6). □

## Proof of Theorem 2

The proof of this theorem is similar to that of Theorem 1; just replace (20) with $\{(n_j + \gamma_j)/(n_j + 1 + \gamma_j)\}^{m_j}\{(n_k + \gamma_k)/(n_k - 1 + \gamma_k)\}^{m_k}$. Then (1) reduces to $(1 + 1/\min_j \gamma_j)^m \leq \exp(\epsilon)$, which is equivalent to (8). □

## Proof of Theorem 3

Similarly to the proof of Theorem 1, we can simplify the definition (1) of differential privacy under (9) as

$$\frac{B_{m_j}(n_j + \gamma_j, \boldsymbol{w})}{B_{m_j}(n_j + 1 + \gamma_j, \boldsymbol{w})} \frac{B_{m_k}(n_k + \gamma_k, \boldsymbol{w})}{B_{m_k}(n_k - 1 + \gamma_k, \boldsymbol{w})} \leq \exp(\epsilon). \tag{21}$$

This condition (21) must hold for all $(\vec{m}, \vec{n}, \vec{n}')$. The l.h.s. of (21) is regarded as the product of two ratios, and we evaluate the upper bound of each ratio.

The first ratio of the l.h.s. of (21) can be rewritten for $m_j \in \mathbb{N}$ as

$$\frac{B_{m_j}(n_j + \gamma_j, \boldsymbol{w})}{B_{m_j}(n_j + 1 + \gamma_j, \boldsymbol{w})} = \sum_{l=1}^{m_j} \left(\frac{n_j + \gamma_j}{n_j + \gamma_j + 1}\right)^l \frac{(n_j + \gamma_j + 1)^l B_{m_j,l}(\boldsymbol{w})}{\sum_{l=1}^{m_j}(n_j + \gamma_j + 1)^l B_{m_j,l}(\boldsymbol{w})}. \tag{22}$$

The r.h.s. of (22) is the weighted mean of $((n_j + \gamma_j)/(n_j + \gamma_j + 1))^l, l \in [m_j]$. All the weights are nonnegative for $\boldsymbol{w} \in \mathbb{R}_+ \times \mathbb{R}_{0+}{}^{\infty}$. Therefore, the weighted mean is bounded as

$$0 < \left(\frac{n_j + \gamma_j}{n_j + \gamma_j + 1}\right)^{m_j} \leq \frac{B_{m_j}(n_j + \gamma_j, \boldsymbol{w})}{B_{m_j}(n_j + 1 + \gamma_j, \boldsymbol{w})} \leq \frac{n_j + \gamma_j}{n_j + \gamma_j + 1} < 1$$

when $m_j \in \mathbb{N}$ since $n_j + \gamma_j > 0$. Hence (22) is maximized to unity when $m_j = 0$.

The second ratio of the l.h.s. of (21) is also the weighted mean of $((n_k + \gamma_k)/(n_k - 1 + \gamma_k))^l$. Similarly,

$$1 < \frac{n_k + \gamma_k}{n_k - 1 + \gamma_k} \leq \frac{B_{m_k}(n_k + \gamma_k, \boldsymbol{w})}{B_{m_k}(n_k - 1 + \gamma_k, \boldsymbol{w})} \leq \left(\frac{n_k + \gamma_k}{n_k - 1 + \gamma_k}\right)^{m_k} \tag{23}$$

when $m_k \in \mathbb{N}$. Because $m_k \leq m$, the upper bound of (23) is maximized when $m_k = m$. Hence

$$\frac{B_{m_j}(n_j + \gamma_j, \boldsymbol{w})}{B_{m_j}(n_j + 1 + \gamma_j, \boldsymbol{w})} \frac{B_{m_k}(n_k + \gamma_k, \boldsymbol{w})}{B_{m_k}(n_k - 1 + \gamma_k, \boldsymbol{w})} \leq \left(\frac{n_k + \gamma_k}{n_k - 1 + \gamma_k}\right)^m,$$

where the upper bound is maximized when $n_k + \gamma_k$ is minimized. Consequently, we obtain the result. □

## Proof of Theorem 4

Because $n_k - 1 + \gamma_k > 0$, the second ration of the l.h.s. of (21) is maximized when $m_k = m$ under the assumption that $\boldsymbol{w} \in \mathcal{W}_1$. Hence, observing the proof of Theorem 3, the definition (1) of differential privacy is simplified as

$$\frac{B_m(n_k + \gamma_k, \boldsymbol{w})}{B_m(n_k - 1 + \gamma_k, \boldsymbol{w})} \leq \exp(\epsilon). \tag{24}$$

Under the assumption that $\boldsymbol{w} \in \mathcal{W}_2$, (24) is further reduced to (14). □

## Proof of Theorem 5

The condition of $W$ requires that $B_{n+1}(\lambda+1,\boldsymbol{w})/B_{n+1}(\lambda,\boldsymbol{w}) \geq B_n(\lambda+1,\boldsymbol{w})/B_n(\lambda,\boldsymbol{w})$. We rewrite this condition as

$$\frac{B_{n+1}(\lambda+1,\boldsymbol{w})B_n(\lambda,\boldsymbol{w})}{B_{n+1}(\lambda,\boldsymbol{w})} - B_n(\lambda+1,\boldsymbol{w}) \geq 0. \tag{25}$$

The l.h.s. of (25) is further rewritten as

$$\sum_{k=1}^{n}(\lambda+1)^k\{B_{n+1,k}(\boldsymbol{w})\frac{B_n(\lambda,\boldsymbol{w})}{B_{n+1}(\lambda,\boldsymbol{w})} - B_{n,k}(\boldsymbol{w})\} + (\lambda+1)^{n+1}B_{n+1,n+1}(\boldsymbol{w})\frac{B_n(\lambda,\boldsymbol{w})}{B_{n+1}(\lambda,\boldsymbol{w})}. \tag{26}$$

We note that

$$\begin{aligned}&\sum_{k=1}^{n}\lambda^k\{B_{n+1,k}(\boldsymbol{w})\frac{B_n(\lambda,\boldsymbol{w})}{B_{n+1}(\lambda,\boldsymbol{w})} - B_{n,k}(\boldsymbol{w})\} + \lambda^{n+1}B_{n+1,n+1}(\boldsymbol{w})\frac{B_n(\lambda,\boldsymbol{w})}{B_{n+1}(\lambda,\boldsymbol{w})}\\ &= \quad B_{n+1}(\lambda,\boldsymbol{w})\frac{B_n(\lambda,\boldsymbol{w})}{B_{n+1}(\lambda,\boldsymbol{w})} - B_n(\lambda,\boldsymbol{w}) = 0.\end{aligned}$$

Because $\lambda^{n+1}B_{n+1,n+1}(\boldsymbol{w})\frac{B_n(\lambda,\boldsymbol{w})}{B_{n+1}(\lambda,\boldsymbol{w})}$ is strictly positive, $\sum_{k=1}^{n}\lambda^k\{B_{n+1,k}(\boldsymbol{w})\frac{B_n(\lambda,\boldsymbol{w})}{B_{n+1}(\lambda,\boldsymbol{w})} - B_{n,k}(\boldsymbol{w})\}$ is strictly negative. Then we use the condition of this theorem that

$$B_{n+1,k}(\boldsymbol{w})/B_{n,k}(\boldsymbol{w}) \leq B_{n+1,k+1}(\boldsymbol{w})/B_{n,k+1}(\boldsymbol{w}).$$

It implies that there exists an integer $C \in [n]$ such that

$$\{B_{n+1,i}(\boldsymbol{w})\frac{B_n(\lambda,\boldsymbol{w})}{B_{n+1}(\lambda,\boldsymbol{w})} - B_{n,i}(\boldsymbol{w})\} < 0, \quad i \in [C],$$

and

$$\{B_{n+1,i}(\boldsymbol{w})\frac{B_n(\lambda,\boldsymbol{w})}{B_{n+1}(\lambda,\boldsymbol{w})} - B_{n,i}(\boldsymbol{w})\} \geq 0, \quad i \in ([n] \setminus [C]).$$

For such $C$, define

$$T_1 := \sum_{i=1}^{C}\lambda^i\{B_{n+1,i}(\boldsymbol{w})\frac{B_n(\lambda,\boldsymbol{w})}{B_{n+1}(\lambda,\boldsymbol{w})} - B_{n,i}(\boldsymbol{w})\} < 0$$

and

$$T_2 := \sum_{i=C+1}^{n}\lambda^i\{B_{n+1,i}(\boldsymbol{w})\frac{B_n(\lambda,\boldsymbol{w})}{B_{n+1}(\lambda,\boldsymbol{w})} - B_{n,i}(\boldsymbol{w})\} + \lambda^{n+1}B_{n+1,n+1}(\boldsymbol{w})\frac{B_n(\lambda,\boldsymbol{w})}{B_{n+1}(\lambda,\boldsymbol{w})} > 0.$$

Now we rewrite (26) as

$$\begin{aligned}&\sum_{i=1}^{C}(\lambda+1)^i\{B_{n+1,i}(\boldsymbol{w})\frac{B_n(\lambda,\boldsymbol{w})}{B_{n+1}(\lambda,\boldsymbol{w})} - B_{n,i}(\boldsymbol{w})\}\\ &+ \quad \sum_{i=C+1}^{n}(\lambda+1)^i\{B_{n+1,i}(\boldsymbol{w})\frac{B_n(\lambda,\boldsymbol{w})}{B_{n+1}(\lambda,\boldsymbol{w})} - B_{n,i}(\boldsymbol{w})\} + (\lambda+1)^{1+n}B_{n+1,n+1}(\boldsymbol{w})\frac{B_n(\lambda,\boldsymbol{w})}{B_{n+1}(\lambda,\boldsymbol{w})}.\end{aligned}$$

The first line of this expression is not less than $\{(\lambda+1)/\lambda\}^C T_1$. The second line is not less than $\{(\lambda+1)/\lambda\}^{C+1}T_2$. By noting that $T_1 + T_2 = 0$, (26) is not less than $[\{(\lambda+1)/\lambda\}^{C+1} - \{(\lambda+1)/\lambda\}^C]T_2$, which is strictly positive. Consequently, (25) is satisfied. □

# References


[1] Bethlehem, J.G., Keller, W.J., and Pannekoek. J. (1990) Disclosure Control of Microdata, *J. Amer. Statist. Assoc.*, **85**, 38–45.

[2] Charalambides, Ch.A. (2002). *Enumerative Combinatorics*, Chapman and Hall/CRC, Florida.

[3] Cheng Ping (1964). Minimax estimates of parameters of distributions belonging to the exponential family. *Acta Mathematica Sinica*, **5**, 277–299.

[4] Comtet, L. (1974). *Advanced Combinatorics*, D. Reidel Pub. Co., Boston.

[5] Consul, P.C. and Mittal, S.P. (1977). Some discrete multinomial probability models with predetermined strategy. *Biom. J.*, **19**, 161–173.

[6] Devroye, L. (1986). *Non-Uniform Random Variate Generation*. Springer, New York.

[7] Dwork, C., Kenthapadi, K., McSherry, F., Mironov, I. and Naor, M. (2006a). Our data, ourselves: privacy via distributed noise generation. *Advances in Cryptology - EUROCRYPT 2006*, Vaudenay, S. (ed.), Lecture Notes in Computer Science, **4004**, 486–503, Springer, Berlin.

[8] Dwork, C., McSherry, F., Nissim, K. and Smith, A. (2006b). Calibrating noise to sensitivity in private data analysis. *TCC 2006-Theory of Cryptography Conference*, 265–284.

[9] Harris, B. and Schoenfeld, L. (1967). The number of idempotent elements in symmetric semigroups. *Journal of Combinatorial Theory*, **3**, 122-135.

[10] Hoshino, N. (2020). A firm foundation for statistical disclosure control. *Japanese Journal of Statistics and Data Science*, **3**, 721–746.

[11] Hoshino, N. (2021). Urn models closed under resizing. *ISM research memorandum*, **1215**, The Institute of Statistical Mathematics, Tokyo.

[12] Hu, J. and Hoshino, N. (2018). The quasi-multinomial synthesizer for categorical data. In: Domingo-Ferrer, J., Montes, F. (eds) *Privacy in Statistical Databases, PSD 2018*. Lecture Notes in Computer Science, **11126**. 75–91, Springer, Cham.

[13] Inusah, S. and Kozubowski, T.J. (2006). A discrete analogue of the Laplace distribution. *Journal of Statistical Planning and Inference*, **136**, 1090–1102.

[14] Machanavajjhala, A., Kifer, D., Abowd, J., Gehrke, J., and Vilhuber, L. (2008). Privacy: Theory meets practice on the map. *Proceedings of the 2008 IEEE 24th International Conference on Data Engineering, ICDE'08*, 277–286.

[15] Marsh, C., Skinner, C., Arber, S., Penhale, P., Openshaw, S., Hobcraft, J., Lievesley, D. and Walford, N. (1991). The Case for a Sample of Anonymized Records from the 1991 Census. *Journal of the Royal Statistical Society*, Series A, **154**, 305–340.

[16] Neerchal, N.K. and Morel, J.G. (2005). An improved method for the computation of maximum likelihood estimates for multinomial overdispersion models. *Computational Statistics and Data Analysis*, **49**, 33–43.

[17] Pitman, J. (2006). *Combinatorial Stochastic Processes*. Lecture notes in Mathematics, **1875**, Springer, New York.

[18] Riordan, J. (1968). *Combinatorial Identities*. Wiley, New York.

[19] Rocher, L., Hendrickx, J.M. and de Montjoye, Y. (2019). Estimating the success of re-identifications in incomplete datasets using generative models. *Nature Communications*, **10**, 3069.