Towards optimal noise distribution for privacy preserving in data aggregation

Hao Zhang a, Nenghai Yu a,*, Yonggang Wen b, Weiming Zhang a

a Key Laboratory of Electromagnetic Space Information, University of Science and Technology of China, Hefei 230027, China
b School of Computer Engineering, Nanyang Technological University, 639798 Singapore

Abstract

In aggregation applications, individual privacy is a crucial factor to determine the effectiveness, for which the noise-addition method (i.e., a random noise value is added to the true value) is a simple yet powerful approach. However, improper additive noise could result in bias for the aggregate result. It demands an optimal noise distribution to reduce the deviation. In this paper, we develop a mathematical framework to derive the optimal noise distribution that provides privacy protection under the constraint of a limited value deviation. Specifically, we first derive a generic system dynamic function that the optimal noise distribution must satisfy, and further investigate the special case that the original values obey Gaussian distribution. Then we detailedly investigate the general cases that the original values obey arbitrary continuous distribution, which can be expressed by Gaussian Mixture Model (GMM). Our theoretical analysis suggests that for the Gaussian input Gaussian distribution is the optimal solution, and for the general input, the optimal solution is composed of infinite number of Gaussian components. We further find the general term formula of the components, which reduces the number of unknowns from infinite to three, i.e. the parameters of the first component (variance, expectation, weight). Based on it, we investigate the properties and propose an algorithm in order to calculate the asymptotically optimal solution composed of finite Gaussian components. The numerical evaluation shows that the results has little deviation to the optimal solutions.

1. Introduction

With the advance of information age, data aggregation has been widely used in daily life and commercial applications. Even some companies such as Canalys make a living by providing all kinds of statistics. In aggregation applications the server wishes to distill valuable aggregate statistics from a mass of individual data. For example, CarTel (Hull et al., 2006) learns the traffic condition from the road information collected by using mobile phones. BikeNet (Eisenman et al., 2009) measures air and road condition to guide cyclists, where all the data is contributed by users’ devices.

However, the individual privacy may be violated during the aggregation. The server is able to obtain the individual data of participants from inputs. Nevertheless, much of this information is private for individuals, such as health condition,
income, etc., especially in the presence of curious server or data abuse. Actually the server only need to know the aggregate result without knowing the individual data. Thus in aggregation applications, calculate the aggregate statistics without compromising individual privacy is an important challenge.

Some methods are proposed to solve this problem. Secure Multiparty Computation (SMC) is a good choice to calculate the statistics. It uses cryptographic methods, doing operations on ciphertext domain. However, it also has its limitations. Firstly, because of the huge overhead, SMC is not suitable for large-scale systems. Secondly, both encryption/decryption and communication are high power consumption operations, which limit SMC deployed in energy-sensitive devices (e.g. sensor, phone). Therefore, SMC is not suitable for the large-scale energy-constraint environments such as large-scale mobile survey applications. On the contrary, noise addition, which prevents the adversary from getting the accurate individual data values, is a simple but effective method. Compared to SMC, it is much simpler and efficient, especially in this environment. Without collaboration with others, each participant only adds noise into his data independently before updating. However, in this method, how to choose the noise distribution is a headache. Improper additive noise could result in bias for the aggregate result. It demands an optimal noise distribution which provides the best protection to individual privacy while the aggregate result has tolerable bias. However, the optimal noise distribution is not evident. Usually the noise distribution (usually homogeneous noise or Gaussian noise) is proposed directly without any explanation.

In aggregation applications the accuracy of result and the privacy of individuals are two main concerned issues. Our goal is to find out the optimal noise distribution in noise addition method, where the individual privacy is protected best under the given accuracy requirement. In this paper, we measure the accuracy by expectation and variance of the noise distribution, and the privacy by conditional entropy. Based on the metrics, we develop a mathematical framework to derive the optimal noise distribution that provides privacy protection under the constraint of a limited value deviation, deriving a generic system dynamic function that the optimal noise distribution must satisfy, with the original data distribution as base, which needs the party who adds the noise to know the first component. In Section 6 we investigate the properties of the component sequence. In Section 7 we propose an algorithm to express the approximate optimal distribution, and reducing the algorithm complexity in two special cases. Section 8 shows some simulation results. At last we conclude the paper and the future work in Section 9.

1. Based on the accuracy and privacy metrics, we develop a mathematical framework to derive the optimal noise distribution.
2. We get the generic system dynamic function that the optimal noise distribution must satisfy, where the input is the distribution of original individual data. Based on the function, we find Gaussian distribution is the optimal noise distribution for the Gaussian input.
3. We deeply investigate the general case that the input is arbitrary continuous distribution. We propose a general method to calculate the optimal noise distribution.
4. We investigate the properties of the optimal noise distribution, which is composed of infinite Gaussian components, and find the periodic structure so that it can be approximated by finite components.
5. We propose an algorithm calculating the optimal distribution. For some special cases, the complexity of the algorithm is reduced greatly, and the approximation matches the theoretic analysis well.

The rest of the paper are organized as follows. Related work is introduced in Section 2. We formulate the problem in Section 3. In Section 4 we give the general solution and investigate the Gaussian input. Arbitrary continuous distribution input is explored in Section 5, and we find out that output consists of infinite Gaussian components, all of which are fixed by the first component. In Section 6 we investigate the properties of the component sequence. In Section 7 we propose an algorithm to express the approximate optimal distribution, and reducing the algorithm complexity in two special cases. Section 8 shows some simulation results. At last we conclude the paper and the future work in Section 9.

2. Related work

SMC enables parties to calculate the result by collaboration based on their own data without compromising others’ privacy. However it has lots of limitations. In Clifton et al. (2002) a secure sum protocol was depicted, where the summation is calculated serially which would spend too much time in large-scale systems. Another protocol (Shi et al., 2011) was proposed, which allows the untrusted server to calculate the summation. It requires that the sum of the keys of parties is 0. If one of the party leaves in the process, which is a common case in large-scale systems, the summation cannot be calculated. Jung et al. (2013) proposed a linear time protocol without secure channel, but it still needs lots of communications among parties. Meanwhile, in these methods each party has to communicate with others and do lots of mathematical operations, both of which are high power consumption operations. So SMC is not suitable for energy-constrained devices.

Noise addition has been studied for many years in secure data mining (Adam and Worthmann, 1989). It prevents the adversary from getting the accurate individual data values. Plenty of schemes are proposed to preserve the privacy of individual records, but they all do not solve the problem as ours. Most of them such as Oliveira and Zaiane (2003); Su et al. (2008) are not claimed whether their methods are optimal. Furthermore, they utilize the covariance of data in the database, which needs the party who adds the noise to know the global information of data. In some schemes the noise is added without concerning the covariance of the data, but the uniform distribution or Gaussian distribution is directly declared (Agrawal and Srikant, 2000; Domingo-Ferrer et al., 2004). In Zhu and Liu (2004) the authors considered the optimal randomization given the bias of results, but they did not solve it. Traub et al. (1984) analyzed the tradeoff of the magnitude of the perturbations and the error Chebyshev inequality, but the metrics are rough. Different from many other researches, Dwork et al. (2006) proposed a method that the bias of summation obeys specified distribution. It adds
noise into the summation directly, which is different from other works focussing on adding noises into individual values. Meanwhile, some researchers (Agrawal and Srikant, 2000; Agrawal and Aggarwal, 2001) found the original data distribution can be restructured by perturbed values, but the individual privacy is not violated yet.

To protect the privacy of individual data, the privacy metric is necessary and has greatly influence on the protection scheme. There are several different measures of privacy. In Agrawal and Srikant (2000) the privacy is measured by “confidence interval”. If the data concerned $x$ is in the interval $[l(x), u(x)]$ with at least certain probability $c\%$, the length of interval $|l(x)|$ is treated as a privacy measure. However, this measure is not accurate. Mutual information or differential entropy in Shannon’s information theory is another much more popular privacy metric (Agrawal and Aggarwal, 2001). It indicates the average privacy supporting by mathematical theory. Renyi entropy (an extension of Shannon entropy) is also used to measure privacy (Rachlin et al., 2009), but it is too complex and does not have obvious physical meaning. Besides, some privacy metrics for single record are proposed in Bezzi (2010), which are also based on Shannon entropy.

Recent years differential privacy (Dwork, 2006) is a hot noise addition technology protecting individual’s privacy in data mining. It guarantees the accuracy of statistical result while avoiding individual record disclosure. Ghosh et al. (2009, 2011) find out the optimized noise distribution that provides most accurate result under the given privacy requirement. However, differential privacy is against the adversary that obtains individual record from different statistic results. For example, if $A$ is the sum of 50 records, $B$ is the sum of 49 records from the former 50 ones. One record is disclosed by $A - B$. In our situation, the adversary can get the individual records directly, and we only focus on one aggregation process.

3. Problem formulation

In this section, firstly we introduce some aggregation applications where the violation of individual privacy potentially exists and noise addition method is appropriate. Then we quantify the accuracy and privacy requirements. Finally, based on the measurements the optimization problem is presented.

3.1. Applications

The individual privacy is potentially threatened in statistics aggregation applications. There are many examples, including:

- Sensor network aggregation. In sensor network applications, many energy-constrained sensors are widely deployed to monitor the surrounding environment and send data to the central server for aggregation. However, the data from individual sensor may contain privacy-sensitive information. So energy-efficient privacy protection in aggregation is an important issue.
- Mobile survey applications. In these applications, tens of thousands of participants exist and the phones are energy-constrained. The overall results are distilled from a large amount of individual information collected by mobile phones. However, the individual privacy may be violated during information collection.

In these large-scale energy-constrained applications, the server should know is the aggregate results, which are distilled from the information of individuals. However, the individual privacy may be violated during the collection. Noise addition technology is a simple but efficient method in these applications, which protects the individual privacy by adding noise into the individual data. To describe the problem more accurately, in the following we formulate the problem in the mathematical way.

3.2. Accuracy and privacy measurement

Suppose there are $n$ users with values $x_i$, $i = 1, 2, \ldots, n$, and a server calculating aggregate statistics. In this paper, we mainly focus on a simple but common statistic problem called summation. The server processes the aggregation function $\sum(y) = \sum_{i=1}^{n} x_i$. Of course there are several other aggregation types. Besides summation, Popa et al. (2011) list other classes: average, standard deviation, count and others. All of them can be constructed by summation.

To protect the individual privacy in the process of aggregating statistics, user $u_i$ adds random noise $z_i$ into his/her true value $x_i$. Instead of $x_i$, $u_i$ contributes the perturbed value $y_i = x_i + z_i$ to the server. The information that the attacker knows most is all the perturbed values and the scheme by which the noise is generated. So we suppose the attacker knows $y_i$ and the distributions of $y_i$ and $z_i$. He tries to get $x_i$ based on the information he knows. The aim of the noise is to prevent the attacker from getting the accurate true value.

Obviously different noise distributions have different privacy protection capability. To protect the true value, how to choose a good noise distribution is the key issue. Noise $Z_i$ is a random variable with the probability density function (pdf) $f_{Z_i}$. To meet the requirements of accuracy and privacy, $f_{Z_i}$ should satisfy:

1. accuracy requirement: the difference of $\sum Y_i$ and $\sum X_i$ is small.
2. privacy requirement: the confusion of the true value is evident.

The first requirement guarantees that the aggregate result does not deviate from the true result too much. The second one guarantees the individual privacy is not violated. If the attacker gets the user’s value, he still doubts about it because of the existence of noise.

3.2.1. Accuracy measurement

For accuracy requirement, we define the difference

$$M_n = \sum_{i=1}^{n} Y_i - \sum_{i=1}^{n} X_i = \sum_{i=1}^{n} Z_i,$$

where $n$ is the number of participants. Due to $Z_1, \ldots, Z_n$ are random variables, $M_n$ also is a random variable, with the
expectation $E(M_n)$ and the variance $D(M_n)$. $Z_1, Z_2, ..., Z_n$ are independent, where $Z_i$ has the expectation $\mu_z$ and the variance $\sigma_z^2$ respectively. If they satisfy Lindenberg's condition (Lindeberg), $M_n$ obeys Gaussian distribution regardless of the distributions of individual noise. It is only decided by the expectation and the variance, i.e. $E(M_n) = \sum_{i=1}^{n}\mu_z$, $D(M_n) = \sum_{i=1}^{n}\sigma_z^2$. We try to keep $M_n$ small with high probability. It requires $E(M_n) = 0$ and $D(M_n)$ is small. Therefore, we quantify the accuracy requirement $U$ as

$$U = \frac{D(M_n)}{\sqrt{n}} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n}\sigma_z^2$$

with an additional condition $E(M_n) = 0$. It measures the average deviation tolerance of the perturbed result from the true result. Suppose $Z_1, ..., Z_n$ are independent and identically distributed random variables with the expectation $\mu_z$ and the variance $\sigma_z^2$. The accuracy requirement $U$ is simplified as

$$U = \sigma_z^2$$

with $\mu_z = 0$.

$U$ measures the average deviation tolerance of the perturbed result from the true result by the variance of the noise distribution. If two zero-mean noise distribution $f_{z_2}$ and $f_{z_1}$ satisfy $U_1 < U_2$, it means $f_{z_2}$ guarantees the accuracy of result better.

### 3.2. Privacy measurement

For privacy requirement, we should enlarge the uncertainty of the individual’s value. Suppose $Z_1, ..., Z_n$ are independent and identically distributed random variables, denoted as $Z$. Consider

$$Y = X + Z$$

where $Y$, $X$ and $Z$ are random variables which delegate individual’s perturbed value, true value and noise respectively, and $Z$ is independent from $X$. Suppose the adversary knows the distribution of $Z$. It is reasonable that any user knows it to generate noise, including malicious user compromised by the adversary. For user $u_x$ because of the perturbation of noise $z_{u_x}$, the adversary is uncertain about $x$ when he gets $y_{u_x}$. We use Shannon's information entropy to measure the uncertainty. Suppose the adversary gets $y_{u_x}$, the uncertainty of $X$ is measured by $H(X|Y = y_{u_x}) = -\sum_z P_{X|Y}(x|y)\log P_{X|Y}(x|y)$. The larger $H(X|Y = y_{u_x})$ is, the better the privacy protection is provided at $y_{u_x}$.

For different $y_{u_x}$, $H(X|Y = y_{u_x})$ is different. We use the average $H(X|Y = y)$ to quantify the privacy protection strength (denoted by $V$) of the noise, i.e.

$$V = \sum_{y} H(X|Y = y)P_Y(y) = H(X|Y).$$

$H(X|Y)$ denotes the average uncertainty of the true value when the perturbed value is captured. The larger $V$ is, the higher the average uncertainty is.

Generally speaking, for noise addition technology, the accuracy and the privacy are in contradiction. High accuracy leads to low privacy protection strength, and vice versa. However, for a given accuracy level, different $f_z$ usually has different privacy protection capability. Thus how to optimize the noise distribution that provides the best privacy protection under the accuracy constraint is the key problem.

### 3.3. Optimization problem formulation

For convenience, in the following we consider the continuous distributions. The discrete distribution can be regarded as the approximation of the corresponding continuous distribution. Consider the formulation $Y = X + Z$, where $X, Z$ are random variables with pdf $f_x(x)$ and $f_z(z)$ respectively. We will find the optimal $f_z$ providing the best privacy protection while guaranteeing that the result has an acceptable deviation, i.e.

$$\max_{f_z(z)} V = H(X|Y)$$

s.t. $U = \sigma_z^2 \leq \sigma_m^2$

$$E(Z) = 0$$

where $\sigma_m^2$ is the accuracy requirement bound required by applications.

Consider

$$H(X|Y) = H(X) - I(X;Y).$$

Since $f_x(x)$ is deterministic, $H(X)$ is a constant. Thus the optimization problem is translated to

$$\min_{f_z(z)} I(X;Y)$$

s.t. $\int f_z(z)z^2 \, dz \leq \sigma_m^2$

$$\int f_z(z)z \, dz = 0$$

$$\int f_z(z) \, dz = 1.$$
and investigate the corresponding optimal noise distribution $f_2$ (in the following we denote the optimal noise distribution by $f_2$).

5.1. Solution translation

From GMM, any continuous distribution can be approximated by the mixture of weighted Gaussian distributions, i.e. any continuous distribution $f$ can be represented as

$$f(x) = \sum_{i=0}^{M-1} p_i \phi_i(x),$$

where $p_i, i = 0, \ldots, M - 1 (M \in \mathbb{N}^+)$, are the nonnegative mixture weights, and $\phi_i(x), i = 0, \ldots, M - 1$, are the Gaussian densities with expectation $\mu_i$ and expectation $\sigma_i^2$. The mixture weights satisfy $\sum_{i=0}^{M-1} p_i = 1$. Here we call $p_i \phi_i(x)$ Gaussian component, which is the product of Gaussian pdf and the weight. So the component is decided by three parameters $(p_i, \mu_i, \sigma_i^2)$. $f$ is the merger of $M$ Gaussian components, expressed as $f = (p_0 \phi_0(x), p_1 \phi_1(x), \ldots, p_{M-1} \phi_{M-1}(x))$. Based on GMM, suppose $f_x$ and $f_z$ consist of $M$ Gaussian components and $N$ ones respectively. They can be expressed by the merger of multiple Gaussian components, i.e.

$$f_x(x) = \frac{1}{\sqrt{2\pi}} \sum_{i=0}^{N-1} p_{X_i} \frac{1}{\sigma_{X_i}} e^{-\frac{(x-x_{X_i})^2}{2\sigma_{X_i}^2}},$$

$$f_z(z) = \frac{1}{\sqrt{2\pi}} \sum_{i=0}^{N-1} p_{Z_i} \frac{1}{\sigma_{Z_i}} e^{-\frac{(z-z_{Z_i})^2}{2\sigma_{Z_i}^2}}.$$  

A simpler expression is $f_x = (x_0, x_1, \ldots, x_{M-1})$, where $x_i, i = 0, \ldots, M - 1$, is decided by the parameters $(p_{X_i}, \mu_{X_i}, \sigma_{X_i}^2)$ (denoted by $x_i = (p_{X_i}, \mu_{X_i}, \sigma_{X_i}^2)$), and $f_z = (z_0, z_1, \ldots, z_{N-1})$ where $z_j, j = 0, \ldots, N - 1$, is decided by $(p_{Z_i}, \mu_{Z_i}, \sigma_{Z_i}^2)$. In the following we consider the two operations of the system in Fig. 1.

5.1.1. Convolution

$f_f$ is the convolution of $f_x$ and $f_z$. According to the properties of convolution of Gaussian function,

$$f_f(x) = \frac{1}{\sqrt{2\pi}} \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} p_{X_i} p_{Z_j} \frac{1}{\sigma_{X_i}^2 + \sigma_{Z_j}^2} \frac{1}{\sigma_{X_i}^2 + \sigma_{Z_j}^2} e^{-\frac{(x-x_{X_i} - z_{Z_j})^2}{2(\sigma_{X_i}^2 + \sigma_{Z_j}^2)}},$$

or $f_f = (y_0, y_1, \ldots, y_{M-1}(N-1-M-1))$ where $y_i = z_i + x_i$ is a Gaussian component decided by $(p_{Y_i}, \mu_{Y_i}, \sigma_{Y_i}^2)$ ($y_i = (p_{Y_i}, \mu_{Y_i}, \sigma_{Y_i}^2)$), and $\mu_{Y_i} = \mu_{Z_i} + \mu_{X_i}, \sigma_{Y_i}^2 = \sigma_{X_i}^2 + \sigma_{Z_i}^2, p_{Y_i} = p_{Z_i} p_{X_i}$. Each component of $f_z$ generates $M$ component of $f_f$. If $f_f$ is composed of $N$ components, $f_f$ is composed of $M \times N$ components.

5.1.2. Multiplication

From the former operation, we get the expression of $f_f$ from $f_z$. Conversely, by the multiplication operation, $f_f$ is generated by
Fig. 2 - Multiplication: each component of $f_Y$ refers to one component of $f_Z$ in multiplication operation, where $k = iM + j$, $M$ is the number of Gaussian component of $f_X$.

$f_Y$ that multiplying all the Gaussian components of $f_Y$ by the same multiplier. Fig. 2 illustrates the operation, where all the components of $f_Y$ are multiplied by the same factor $C e^{-j2\pi z_j \omega}$. Each component of $f_Y$ is transformed into one of $f_Z$. We call the corresponding two components of $f_Y$ a pair or pair. For example, $z_0$ is the product of $y_{00}$ and the factor, so $z_0$ and $y_{00}$ are a pair. $z_0 (k = iM + j)$ is the product of $y_i$ and the factor, so $z_0$ and $y_i$ are a pair.

Suppose there are two pairs: $y_{ij}$, $z_{ij}$, and $y_{ik}$, $z_{ik}$. Since $y_{ij}$, and $y_{ik}$ are multiplied by the same factor, if $y_{ij}$ and $y_{ik}$ are different, $z_{ij}$, $z_{ik}$ are different too. So it is a one-to-one mapping between components of $f_Z$ and $f_Y$. $f_Z$ and $f_Y$ have the same number of components, i.e. $N = M$ in Section 4.2. $M$ and $N$ are different too. Thus in the same pair the co-efficients of $\xi_j$ (the exponent of $e^{j2\pi z_j \omega}$) in $y_{ij}$ are a pair.

Consider $y_i$ and $z_i$ which form a component pair, where $k = iM + j$. From the geometric point of view, $z_i$ stems from $y_i$ by narrowing ($e^{-j2\pi \omega}$), translating ($e^{j2\pi \omega}$), and normalizing ($C$) the curve of $y_i$. From the algebraic point of view, the three operations are visualized as Eqs. (17)–(19) respectively,

$$
\begin{align}
&\frac{1}{e^{(\sigma^2_X_i + \sigma^2_Z_j) i\omega/2}} e^{-\frac{x^2}{2\sigma^2_Z_j}} = e^{-\frac{x^2}{2\sigma^2_Z_j}} \\
&\frac{1}{\sigma^2_X_i + \sigma^2_Z_j} = e^{-\frac{\mu^2_Z_j}{\sigma^2_Z_j}} \\
&\frac{1}{\sigma^2_X_i + \sigma^2_Z_j} = e^{-\frac{\mu^2_Z_j}{\sigma^2_Z_j}} \\
&\frac{1}{\sigma^2_X_i + \sigma^2_Z_j} = e^{-\frac{\mu^2_Z_j}{\sigma^2_Z_j}}
\end{align}
$$

and the constraints for $f_Z$ are

$$
\begin{align}
&\sum_{i} P_Z (\sigma^2_Z_j + \mu^2_Z_j) = \sigma^2_m \\
&\sum_{i} P_Z \mu^2_Z_j = 0 \\
&\sum_{i} P_Z = 1.
\end{align}
$$

In these equations $\sigma^2_X, \mu^2_Z, \sigma^2_X, \mu^2_Z, \sigma^2_X, \mu^2_Z, \sigma^2_X, \mu^2_Z, \sigma^2_X, \mu^2_Z$ are known. $p_Z, \mu_X, \sigma_X, \lambda, \omega, C$ are unknown. Thus the solution is transformed to how to determine the parameters $(\mu_Z, \sigma^2_Z)$ of $z_j (j = 0, \ldots, N − 1)$ by Eqs. (17)–(19) and constraints (20)–(22). In the following subsections we try to calculate the three parameters of Gaussian components of $f_Z$ when $f_Z$ is composed of one, two and more than two Gaussian components.

5.2. $f_Z$ with one component

In this situation $M = 1$, $f_Z$ is a normal distribution. From Section 4.2 $f_Z$ is a normal distribution too, i.e. $N = 1, \sigma^2_Z = \sigma^2_Z, \mu_Z = 0$. Actually, if we choose $N > 1$, by the method proposed in Section 5.3, we would get $\sigma^2_Z = \sigma^2_{00} = \sigma^2_{01} = \sigma^2_{02} = \cdots = \sigma^2_{0N - 1}$ and $\mu_Z = \mu_Z = \cdots = \mu_{ZN - 1}$, i.e. $f_Z$ is a Gaussian distribution.

5.3. $f_Z$ with two components

In this situation $M = 2$, $f_Z = (\xi_0, \xi_1)$, $f_Z$ has infinite Gaussian components, i.e. $N \to \infty$. In the following we would calculate all the components of $f_Z$.

5.3.1. Calculate $\sigma^2_Z$

For all the component pairs of $f_Y$ and $f_Z$, the variances are controlled by the same $\lambda$. Thus in the same pair the coefficients of "$e^{j2\pi \omega}$" (the exponent of "$e^{j2\pi \omega}$") in $y_{ij}$ are equal. Suppose $\sigma^2_{Z_0} \geq \sigma^2_{Z_1}$, and $\sigma^2_Z = \max \{|\sigma^2_Z| j \in \mathbb{N} \}$. If $\sigma^2_Z$ is fixed, all the variances of $f_Z$ and $f_Y$ are determined by the following method.

Step 1 (Fix the first pair). Since $\sigma^2_Z$ is fixed, we get the variances of two components of $f_Y$ denoted by $y_{00}$ and $y_{01}$. With this $\sigma^2_{00}$ and $\sigma^2_{01}$ are different too. All the component pairs are generated by the same $\lambda, \omega, \sigma^2_Z$. $f_Z$ becomes $\sigma^2_{00}$. So we get

$$
\sigma^2_{00} = \frac{1}{\sigma^2_{Z_0}} - \frac{1}{\sigma^2_{Z_1}} = \frac{\sigma^2_{Z_0}}{\sigma^2_{Z_1}} > 0.
$$

Step 2 (Fix $\sigma^2_Z$). All the component pairs are generated by the same $\lambda, \omega, C$, so from any pair the first three parameters are fixed. $y_{00}$ and $z_0$ form a pair, i.e. by narrowing of $\lambda, \sigma^2_X \omega$ would become $\sigma^2_Z$. So we get

$$
\sigma^2_{Z_0} = \frac{1}{\sigma^2_{Z_0}} - \frac{1}{\sigma^2_{Z_1}} = \frac{\sigma^2_{Z_0}}{\sigma^2_{Z_1}} > 0.
$$

Step 3 (Fix second pair). From $\sigma^2_{Z_0}$ and $\lambda$, we get $\sigma^2_{Z_1}$ (Eq. (24)) shown in Fig. 3(b). So $\sigma^2_{Z_0}$ and $\sigma^2_{Z_1}$ are a pair.
Step 4 (Fix other pairs). When $\sigma_j^2$ is fixed, combing $\lambda$, the other two $\sigma^2_j$ (denoted by $\sigma_{y0}^2$ and $\sigma_{y1}^2$) are fixed (Fig. 3(c)). Then the corresponding pair members $\sigma_j^2$ and $\sigma_j^2$ can be calculated (Fig. 3(d))... In this way we can get the pairs $(x_{j0}, y_{j0})$, $j = 0, 1, 2, \ldots$ and $i = 0, 1$, and the variances (Fig. 3(e)). Eq. (24) shows the process of calculating all the variances. Thus given $\sigma_j^2$, all the $\sigma_j^2$ are fixed.

We represent $\sigma_j^2$ with one general term formula. Suppose the binary form of $j$ is $(j_0, j_1, \ldots, j_l)$, $\sigma_j^2$ has the general term formula as

$$\sigma_j^2 = \sum_{k=0}^{l} \sigma_j^2 \delta_{j_k}^2$$

where $\delta_{j_k}^2$ are coefficients of $\sigma_j^2$.

5.3.2. Calculate $\mu_2$

In the above we fix the coefficients of "x" of Eq. (17) and find out the component pairs $(x_{j0}, y_{j0}), j = 0, 1, 2, \ldots$ and $i = 0, 1$. The method of calculating $\mu_2$ is the same as the method of generating $\sigma_j^2$. Based on Eq. (18) we can fix the coefficients of "x". Since $(x_0, y_0)$ is a pair,

$$\omega = \frac{\mu_{x0} + \mu_{x0}}{\sigma_{x0}^2} - \frac{\mu_{x0}}{\sigma_{x0}^2} + \frac{\mu_{x0}}{\sigma_{x0}^2}.$$  

Then $\mu_{x0}$ is fixed by $\mu_{y0}$, and $\omega$ (all the components of $f_x$ translate the same distance under the normalized variances). From $\mu_{y0}$ two expectations of $f_y$, $\mu_{y1}$, and $\mu_{y1}$, are generated. Then $\mu_{y0}$ and $\mu_{y1}$ are fixed... By the iterative method, if $\mu_{y0}$ is fixed, all the $\mu_y^2$ are fixed. Eq. (27) shows the calculation process.

Suppose the binary form of $j$ is $(j_0, j_1, \ldots, j_l)$, $\mu_j$ has the general term formula as

$$\mu_j = \sum_{k=0}^{l} \mu_j \delta_{j_k} \delta_{j_k}.$$

5.3.3. Calculate $p_2$

The method of calculating $p_2$ is the same as the method of generating $\sigma_j^2$. From Eq. (19) the constant term of each pair have the same ratio $C$. Since $(x_0, y_0)$ is a pair, if $p_{x0}$ is fixed,

$$C = \frac{\mu_{x0} - \mu_{x0}}{\sigma_{x0}^2} e^{-\frac{\delta_{j_0}^2 \delta_{j_0}^2 + \frac{1}{\delta_{j_0}^2} \delta_{j_0}^2 + \frac{1}{\delta_{j_0}^2} \delta_{j_0}^2}}.$$  

Then $p_{x0}$ is fixed by $p_{y0}$ and $C$ (all the components of $f_y$ are normalized with the same $C$). From $p_{x0}$ two coefficients of $f_y$, $p_{y0}$ and $p_{y1}$, are generated. Then $p_{x0}$ and $p_{y0}$ are fixed... By the iterative method, if $p_{x0}$ is fixed, all the $p_{x0}$ are fixed. Eq. (30) shows the calculation process.
Suppose the binary form of \( j \) is \((j_0, j_1, \ldots, j_t)\), \( p_{Z_j} \) has the general term formula as

\[
p_{Z_{(i_h, \ldots, i_k)}} = C o_{Z_{(i_h, \ldots, i_k)}} \frac{P_{Z_{(i_h, \ldots, i_k)}} P_{X_{j_t}}}{\sqrt{\sigma_{Z_{(i_h, \ldots, i_k)}}^2 + \sigma_{X_{j_t}}^2} e^{\frac{\sigma_{Z_{(i_h, \ldots, i_k)}}^2}{\sigma_{X_{j_t}}^2} \left( \frac{\mu_{X_{j_t}} - \mu_{Z_{(i_h, \ldots, i_k)}}}{\sigma_{X_{j_t}}} \right)^2}}.
\]  

(31)

Now for constructing \( f_z \), the unknown numbers of Eq. (9) only are \( \sigma_{Z_{j_1}}^2, \mu_{Z_{j_1}} \) and \( p_{Z_{j_1}} \), which are determined by constraints Eqs. (20)–(22).

### 5.4. \( f_z \) with multiple components

When \( f_z \) is composed of more than two Gaussian components \((M > 2)\), the calculation process is similar as the situation \( M = 2 \). \( f_z \) consists of infinite number of Gaussian components \((N \to \infty)\). Given one Gaussian component with the largest variance, other components are calculated. A one to one mapping between the components of \( f_z \) and \( f_Y \) is constructed, i.e. \( y_{j_1} \) and \( z_{M, j_1} \) are a pair, where \( i = 0, 1, \ldots, M - 1 \) and \( j = 0, 1, \ldots, j_t \). Suppose the base-\( M \) form of \( j \) is \((j_0, j_1, \ldots, j_t)\), \( \sigma_{Z_{j_1}}^2, \mu_{Z_{j_1}} \) and \( p_{Z_{j_1}} \) have the general term formula Eqs. (25), (28), and (31) respectively. The formats are the same as the situation \( M = 2 \), but here \( t \in \{0, 1, \ldots, M - 1\} \). Furthermore, the equations fixing \( \sigma_{Z_{j_1}}^2, \mu_{Z_{j_1}} \) and \( p_{Z_{j_1}} \) are the same as Eqs. (20)–(22).

From above analysis, if one Gaussian component of \( f_z \) which has the largest variance among all the components of \( f_z \) is fixed, all other components of \( f_z \) can be calculated by Eqs. (25), (28), and (31). Theoretically speaking, we greatly reduce the number of unknown numbers in constructing \( f_z \) by Eq. (9) from infinite ones to only three ones \((\sigma_{Z_{j_1}}^2, \mu_{Z_{j_1}}, \ p_{Z_{j_1}})\).

Moreover, we give the three Eqs. (20)–(22), which are the constraints of the three parameters.

### 6. Properties of Gaussian components

Based on the investigation in the above section, we only need to calculate the three parameters \( \sigma_{Z_{j_1}}^2, \mu_{Z_{j_1}}, \) and \( p_{Z_{j_1}} \). However, it is impossible to calculate infinite number of Gaussian components of \( f_z \). In this section we explore the properties of these components, by which \( f_z \) could be approximated by finite number of Gaussian components with high accuracy. We mainly consider the properties of \( M = 2 \). The situation \( M > 2 \) is similar, which can be divided into \( M - 1 \) independent situations.

By the generation order of Gaussian components, we construct a tree of the components of \( f_z \) illustrated in Fig. 4, which is another form of Fig. 3 ignoring the components of \( f_Y \). It is composed of the root \( z_0 \) and its right subtree, where the left child is generated by the parent and \( x_0 \), and the right child is generated by the parent and \( x_1 \), where \( \sigma_{X_0}^2 \geq \sigma_{X_1}^2 \). Specifically, the tree also delegate the generation order of \( \sigma_{Z_{j_1}}^2, \mu_{Z_{j_1}} \) and \( p_{Z_{j_1}} \). Based on the tree, in the following we investigate the properties of the three parameters respectively.

#### 6.1. Properties of variance parameter

In this subsection we focus on the variances of all the components. From this binary tree (Fig. 4), where \( z_0 \) has the largest variance \( \sigma_{Z_{j_1}}^2 \), we have the following properties.

**Lemma 1.**  For any \( j \not\in N, \sigma_{Z_{j_1}}^2 \geq \sigma_{Z_{j_1}}^2 \).

The proof is given in Appendix B.
From this lemma, the left child is larger than or equal to the right child.

**Lemma 2.** For any \( j \not\equiv N \), \( \sigma_{Z_k}^2 \geq \sigma_{Z_j}^2 \).

The proof is given in Appendix C.

From this lemma, the left child is larger than or equal to the parent. Consider the condition of equation holds, if \( \sigma_{Z_k}^2 = \sigma_{Z_j}^2 \) for any \( j, \sigma_{Z_k} = \sigma_{Z_l} \). Thus the equation of Lemma 2 holds under the condition \( \sigma_{Z_k}^2 = \sigma_{Z_j}^2 \) too.

**Theorem 3.** For any \( j \not\equiv N \),

\[
s_m \leq \sigma_{Z_k}^2 \leq \sigma_{Z_j}^2,
\]

where \( s_m = \sqrt{\lambda^2 \sigma_{Z_k}^4 + 2\lambda \sigma_{Z_k}^2 - \lambda \sigma_{Z_j}^2}/2\lambda \).

The proof is given in Appendix D.

This theorem shows the range of \( \sigma_{Z_k}^2 \). Suppose \( \sigma_{Z_k}^2/\sigma_{Z_j}^2 = t_0 \), \( \sigma_{Z_k}^2/\sigma_{Z_j}^2 = t_1 \), where \( t_1 \leq 1 \). The ratio of maximum of the interval to minimum is

\[
\frac{\sigma_{Z_k}^2}{s_m} = \frac{\sqrt{t_0^2 + 4t_0(t_0 + 1)t_1} + t_1}{2(t_0 + 1)t_1}.
\]

Shown in Fig. 5, when \( \sigma_{Z_k}^2 \) is not much larger than \( \sigma_{Z_j}^2 \), the interval is small (Z-axis delegates the ratio of maximum of the interval to minimum). When \( t_1 \rightarrow 1, \sigma_{Z_k}^2/s_m \rightarrow 1 \), which means when \( f_c \) is composed of two Gaussian components with almost the same variance, variances of all components of \( f_c \) are almost equal.

**Lemma 3.** For any \( j \not\equiv N \), \( \sigma_{Z_k}^2 \geq \sigma_{Z_{2t+1}}^2 \).

The proof is given in Appendix E.

From this lemma, the parent is larger than or equal to the right child.

When \( \sigma_{Z_k}^2 = s_m \), we have \( \sigma_{Z_k}^2 = \sigma_{Z_{2t}}^2 = \cdots = s_m \). From Theorem 3, \( \sigma_{Z_k}^2 = \sigma_{Z_{2t+1}}^2 \). Thus the equality in this lemma holds as the same as the above lemmas.

**Lemma 4.** For any \( j \in [2^t, 2^{t+1}) \), \( t > 0 \) If \( j \) is even, \( \sigma_{Z_j}^2 \geq \sigma_{Z_{j+1}}^2 \geq \sigma_{Z_{j+2}}^2 \). If \( j \) is odd, \( \sigma_{Z_{j+1}}^2 \geq \sigma_{Z_{j}}^2 \geq \sigma_{Z_{j+2}}^2 \).

The proof is given in Appendix F.

From this lemma, both for variances with odd indices and for variances with even indices, in the same level the largest variance is the leftmost one, the smallest variance is the rightmost one.

**Lemma 5.** For any \( j \not\equiv N \), if \( j \) is even, \( \sigma_{Z_k}^2 \geq \sigma_{Z_l}^2 \geq 1/2 \lambda + 1/\sigma_{Z_l}^2 + s_m \).

If \( j \) is odd, \( \sigma_{Z_k}^2 \geq \sigma_{Z_l}^2 \geq s_m \).

The proof is given in Appendix G.

This lemma shows the ranges of variances with odd indices and even indices respectively. Suppose \( \sigma_{Z_k}^2/\sigma_{Z_l}^2 = t_0 \), \( \sigma_{Z_k}^2/\sigma_{Z_l}^2 = t_1 \) as before. We have the following theorem.

**Theorem 4.** All the \( \sigma_{Z_k}^2 \) with even orders and odd orders are located in \([1/2\lambda + 1/\sigma_{Z_l}^2 + s_m, \sigma_{Z_k}^2]\) and \([s_m, \sigma_{Z_l}^2]\) respectively. If \( t_1 \rightarrow 1 \), the length of the two intervals converge to 0.

The proof is given in Appendix H.

This property is more precise than Theorem 3 (shown in Figs. 6 and 7). For example, under the condition that \( t_0 = 0.8 \), if \( t_1 = 0.5 \), all the \( \sigma_{Z_k}^2 \) lie in \([59.60, 74.76]\). All the \( \sigma_{Z_k}^2 \) with odd indices and even indices lie in \([59.60, 63.82]\) and \([71.84, 74.76]\) respectively. If \( t_1 = 0.8 \), all the \( \sigma_{Z_k}^2 \) lie in \([70.31, 75.34]\). All the \( \sigma_{Z_k}^2 \) with odd indices and even indices lie in \([70.31, 71.39]\) and \([74.39, 75.34]\) respectively. Thus from this theorem, when \( \sigma_{Z_k}^2 \) is not much larger than \( \sigma_{Z_l}^2 \), all the variances can be replaced by two variances \( \sigma_{Z_k}^2 \) (even indices) and \( \sigma_{Z_l}^2 \) (odd indices).

For \( M > 2 \), all the properties are similar to the situation \( M = 2 \). The binary tree is extended to an \( M \)-branches tree, which also can be regarded as the mixture of \( M - 1 \) binary trees \( (\sigma_{Z_k}, \sigma_{Z_{k+1}}, \sigma_{Z_{k+2}}, \ldots, \sigma_{Z_{k+M-1}}) \). The properties of the tree are the similar too. All the variances are divided into \( M \) sets, each of which has a small range.
This approximation of expectations of components. The general term formula of the variance $\sigma^2_{z_j}$ with odd indices to minimum. Consider the situation that $\sigma^2_{z_j}$ is not much smaller than $\sigma^2_{z_0}$ (e.g. $t_1 \geq 0.5$), where all the components with odd indices have the variance $\sigma^2_{z_0}$, and components with even indices have the variance $\sigma^2_{z_j}$, i.e. for $j \in \mathcal{N}$, $\sigma^2_{z_j} = \sigma^2_{z_0}$ and $\sigma^2_{z_{j+1}} = \sigma^2_{z_j}$. Based on this approximation of $\sigma^2_{z}$, we investigate the properties of the expectations of components. The general term formula of $\mu_Z$ Eq. (28) can be refined as

$$\mu_{z_{2l}} = \sigma^2_{z_0} \left( \frac{\mu_{z_{2l}} + \mu_{z_{1}}} {\sigma^2_{z_0} + \sigma^2_{z_{1}}} - \omega \right) = h_0 \mu_{z_{2l}} + H_0$$ (33)

$$\mu_{z_{2l+1}} = \sigma^2_{z_{2l+1}} \left( \frac{\mu_{z_{2l+1}} + \mu_{z_{1}}} {\sigma^2_{z_{2l+1}} + \sigma^2_{z_{1}}} - \omega \right) = h_1 \mu_{z_{2l+1}} + H_1$$ (34)

$$\mu_{z_{2l+2}} = \sigma^2_{z_{2l+2}} \left( \frac{\mu_{z_{2l+2}} + \mu_{z_{1}}} {\sigma^2_{z_{2l+2}} + \sigma^2_{z_{1}}} - \omega \right) = h_2 \mu_{z_{2l+2}} + H_2$$ (35)

where

$$h_0 = \frac{t_0}{t_0 + 1} \quad H_0 = (1 - h_0) \mu_{z_0};$$

$$h_1 = \frac{t_0(t_0 + 1)} {t_0^2 + 2t_0 + t_1} \quad H_1 = (1 - h_1) \mu_{z_0} - t_0(1 - h_1) \mu_{z_{1}} + h_1 \mu_{z_{1}};$$

$$h_2 = \frac{t_0(t_0 + 1)(t_0 + t_1)} {t_0^2 + 2t_0 + 1} \quad H_2 = (1 - h_2) \mu_{z_0} - t_0(1 - h_0) \mu_{z_{1}} + h_2 \mu_{z_{1}};$$

$$h_3 = \frac{t_0(t_0 + 1)(t_0 + t_1)} {t_0^2 + 2t_0 + 1} \quad H_3 = (1 - h_3) \mu_{z_0} - t_0(1 - h_0) \mu_{z_{1}} + h_3 \mu_{z_{1}}.$$

Consider the tree in Fig. 4, suppose $l$ is the 1-th level of the tree from the root. $z_j$ is at the 0-th level. $z_{2l-1}, z_{2l-1+1}, \ldots, z_{2l-1}$ are at the l-th level. We have the following properties.

**Lemma 6.** When the level $l \to \infty$,

$$\mu_{z_{2l}} = \mu_{z_{2l-1}} = \mu_{z_0};$$ (37)

$$\mu_{z_{2l+1}} = \mu_{z_{2l+2}} = \mu_{\infty};$$ (38)

where $\mu_{\infty} = H_3/(1 - h_3)$.

The proof is given in Appendix I.

This lemma shows that with the increase of level of the tree, the leftmost $\mu_Z$ converge to $\mu_{z_0}$, and the rightmost $\mu_Z$ converge to $\mu_{\infty}$.

**Lemma 7.** When the level $l \to \infty$, the $\mu_Z$ at level $l$ are not related to the root of the tree. Approximately, when $t_0$ is close to 0, a small $l$ can be chosen.

The proof is given in Appendix J.

From Eqs. (33)–(36), when $t_1$ is close to 1, we have the approximate relationship $h_0 = h_1 = h_2 = h_3$. When $t_1$ is fixed, $h_i$ are the increasing functions of $t_0$. Figs. 8 and 9 illustrate the relationship.

**Theorem 5.** $\mu_Z$ can be regarded as a periodic sequence with the period $2^l$, where $l$ is large enough.

The proof is given in Appendix K.

This theorem shows the sequence of $\mu_Z$ can be regarded as a periodic sequence with $2^l$. The larger $l$ is, the more accurate the approximation is. From Lemma 7 $l$ is determined by $t_0$. When $t_0$ is small, a small $l$ can be chosen.
6.3. Properties of weight parameter

6.3.1. General properties
Suppose \( 2^l \) is the period of \( \mu_p \). Consider the two sequences \( \{ p_z, p_{z+1}, \ldots, p_{2^l z+1} \} \)
and \( \{ p_{2^l z+1}, p_{2^l z+1+1}, \ldots, p_{2^l z+1+1} \} \), where \( l \geq 1 \). Refer to Fig. 4, the former sequence contains all the \( p_z \) at first \( l + 1 \) levels ( \( p_{z0} \) is at level 0). The latter one contains all the \( p_z \) at level \( l + 2 \). Obviously the number of elements is the same. We have the following theorem for \( p_z \):

**Theorem 6.** Consider the two sequences above, the ratio of the two elements with the same order is equal to the ratio of their parents, i.e.

\[
\frac{p_{2^l z+1+i}}{p_{z0}} = \frac{p_{2^l z+1+i}}{p_{z+1}} = \frac{p_{z+i}}{p_{z+1}}.
\tag{39}
\]

The proof is given in Appendix L.

From Theorem 6 the process can be iterated until all the parent nodes are located at the first \( l \) levels. So the infinite number of \( p_z \) can be replaced by the combination of the first \( 2^{l+1} p_z \).

Along the order of \( p_z \), the sequence \( p_z \) is divided into multiple subsequences, denoted by \( S_0, S_1, \ldots \), each of which contains \( 2^l \) elements. Define \( S_p = R_{ab} S_{ap} \) where \( R_{ab} \) is a sequence containing \( 2^l \) elements, where the operation is that the \( i \)-th elements of \( R_{ab} \) and \( S_p \) multiply each other. For example, \( S_0 = R_{00} S_0 \), where \( R_{00} = [1, 1, \ldots, 1] \). If all the elements of \( R_{ab} \) are equal to \( r_{ab} \), we simplify \( S_p = R_{ab} S_{ap} \). If all the other \( S_p \) are fixed by Eq. (39).

Fig. 10 shows an example of dividing \( p_z \), where the period of \( \mu_p \) is \( 2^l \). Each subsequence contains 2 elements. Suppose \( p_{z+i}/p_z = r, i = 0, 1, \ldots, 2^l - 1 \). The elements of \( R_{ab} \) are relevant to \( \{ r_i \} = \{ 0, 1, \ldots, 2^l - 1 \} \). From theorem 6 all the subsequences can be calculated, where the process is shown in Eq. (40).

\[
\begin{align*}
0^l : & \quad S_0 = [1, 1, \ldots, 1] \cdot S_0 \\
1^l : & \quad S_1 = [r_0 r_1 r_2 \ldots r_{2^l - 1}] \cdot S_0 \\
2^l : & \quad S_2 = [r_0 r_1 r_2 \ldots r_{2^l - 1}] \cdot S_0 \\
& \quad S_{2+1} = [r_{2^l+1} r_{2^l+1} \ldots r_{2^l+1}] \cdot S_1 \\
3^l : & \quad S_3 = [r_0 r_1 r_2 \ldots r_{2^l - 1}] \cdot S_0 \\
& \quad S_{3+1} = [r_{2^l+1} r_{2^l+1} \ldots r_{2^l+1}] \cdot S_1 \\
& \quad S_{3+2} = [r_{2^l+1} r_{2^l+1} \ldots r_{2^l+1}] \cdot S_1 \\
& \quad S_{3+3} = [r_{2^l+1} r_{2^l+1} \ldots r_{2^l+1}] \cdot S_1 \\
& \quad \ldots \\
& \quad \ldots \\
& \quad S_{2^l-1+i} = [r_{i} \ldots r_{i}] + \frac{j}{2^l-1} \cdot [r_{i} \ldots r_{i}] + \frac{2^l-1}{2^l-1} \cdot [r_{i} \ldots r_{i}] \cdot S_i \\
& \quad \ldots \\
& \quad \ldots 
\end{align*}
\tag{40}
\]

Given \( p_{z+i} \), the other \( p_z \) are fixed. If we choose two different \( p_z \) denoted by \( p_{za} \) and \( p_{zb} \) respectively, from Eq. (31) the corresponding \( j \)-th \( p_z \) \( p_{za} \) and \( p_{zb} \) have the relationship as \( p_{zb}/p_{za} = p_{za}/p_{zb} \). Thus the \( p_z \) can be calculated as following: at first \( p_{z0} \) is initialized with arbitrary non-zero value (e.g. \( p_{z0} = 1 \)), and all the other \( p_z \) are fixed by Eq. (31). Then by constraint Eq. (22), all the \( p_z \) are normalized by dividing the sum of \( p_z \), i.e.

\[
p_z = \frac{1}{\sum_{i=0}^n p_{zi}} p_{zi}.
\tag{41}
\]

6.3.2. Special property
Here we focus on a special situation that the period of \( \mu_p \) is 2, where \( \mu_{z0} = \mu_{z1} = \ldots \) and \( \mu_{z2} = \ldots \). Under this constraint, we have either of the following results:

![Fig. 8 – The relationship between h_i and t_0, where t_1 = 0.9.](image-url)
Fig. 9 – The relationship between $h_i$ and $t_0$, where $t_1 = 0.6$.

1. $t_1 \to 1$, $t_0 \to 0$.
2. $t_1 \to 1$, $|\mu_{X_1} - \mu_{X_0}| \to 0$.

In this situation, we have the following theorem.

**Theorem 7.** If the period of $\mu_Z$ is 2,

$$r_0 = r_1.$$  \hspace{1cm} (42)

The proof is given in Appendix M.

From this theorem, when the period of $\mu_Z$ is 2, the complexity of the optimization problem can be reduced greatly.

### 7. Approximation of optimal distribution

In this section we consider how to approximate $f_Z$ by finite Gaussian components. The approximate distribution is denoted by $f_{Zap}$. From the properties of the Gaussian components in the above section, $f_{Zap}$ can be split into $2^l$ weighted Gaussian functions. In the following at first we consider the general situation, and propose an algorithm for expressing $f_{Zap}$. Then we investigate two special cases. The first is the period of $\mu_Z$ is 2, where $f_{Zap}$ contains two Gaussian components. The second is $|\mu_{X_1} - \mu_{X_0}| \to 0$, where the two components are far away from each other.

#### 7.1. General expression

Suppose $f_{Zap}$ contains $2^l$ weighted Gaussian functions, the weight sequence is expressed by $S$, denoted by

$$S = \sum_{i=0}^{m} R_i * WS_i.$$  \hspace{1cm} (43)

Then by normalization, $S$ is fixed. According to Eq. (40) $R_i$ can be calculated. Since the optimal $f_Z$ exists, when $i$ is large enough, the normalized $S$ would be stable, the elements of which are the Gaussian components of the optimal $f_Z$.

Consider the constraints (20) and (21), which can be expressed by the functions of $\sigma_{Z_0}^2$ and $\mu_{Z_0}$, i.e.

$$f_1(\sigma_{Z_0}^2, \mu_{Z_0}) = \sum_{i=0}^{l-1} CP_{Z_i} \left( \sigma_{Z_i}^2 + \mu_{Z_i}^2 \right)$$  \hspace{1cm} (44)

$$f_2(\mu_{Z_0}) = \sum_{i=0}^{l-1} CP_{Z_i} \mu_{Z_i}.$$  \hspace{1cm} (45)

where $f_1(\mu_{Z_0}) = 0$ and $f_1(\sigma_{Z_0}^2, \mu_{Z_0}) = \sigma_m^2$.

From Fig. 11 shows the relationship between $\mu_{Z_0}$ and $f_Z$. $f_2(\mu_{Z_0})$ is a monotone increasing functions of $\mu_{Z_0}$. Similarly, shown in Fig. 12 when $\mu_0$ is fixed, $f_1$ is a monotone increasing functions of $\sigma_{Z_0}^2$. Thus $\mu_{Z_0}$ and $\sigma_{Z_0}^2$ can be calculated by Newton iteration method.

Based on the analysis above, we propose Algorithm 1 to calculate the Gaussian components of $f_{Zap}$ based on the accuracy requirement bound $\sigma_m^2$. In the algorithm $f_{Zap}$ consists of $2^l$ Gaussian components, $l$ is chosen to satisfy $L > l$ so that $S$ is stable. “maxNumber” and “minNumber” are two constants which are large and small enough respectively. At line 20 the function “sum” calculate the summation of the elements in the sequence. If $l$ and $L$ are infinite, $f_{Zap}$ becomes $f_Z$.

The complexity of the algorithm is $O(2^l)$, where $L > l$. When $l$ is large, it is hard to calculate the accurate $S$. In the next
subsections, we focus on some special cases, where the algorithm is much simplified (the complexity is $O(1)$).

Actually, this algorithm is equal to the algorithm using the first $2^L$ components to approximate the optimal distribution. Since only $2^L$ Gaussian distributions in $f_{Zap}$, the similar items can be merged. The code from line 16 to 20 calculates the weights of the first $2^L$ components and merges them.

7.2. Case 1: the period of $\mu_2$ is 2

Consider a special case that the period of $\mu_2$ is 2, i.e. $l = 1$, where Theorem 7 holds. Thus from Eq. (40)

$$S = rS_0$$

where $r$ is a constant related to $r_0$ (or $r_1$). Therefore, by normalization, $S = (p_z, p_{z_1})$. From this conclusion, Algorithm 1 is simplified that getting stable $S$ directly. So line 16 to 20 can be replaced by “$S = S_0$” directly.

When $t_1 \to 1, |\mu_{X_1} - \mu_{X_0}| \to 0$, $f_x$ converges to Gaussian distribution, so Gaussian distribution is a near-optimal solution. In the next section, we could find this approximation is even better than Gaussian distribution.

7.3. Case 2: $|\mu_{X_0} - \mu_{X_1}| \gg 0$

In this case, the two components of $f_x$ are far away from each other. Consider the general solution (Eq. (9)), which is equivalent to

$$f_x(z) = C \left( f_{x_0}^* f_z + f_{x_1}^* f_z \right) e^{-\left( z + c \right)^2},$$

(46)

where $f_{x_0}$ and $f_{x_1}$ are the expression of $x_0$ and $x_1$ respectively. The factor $e^{-\left( z + c \right)^2}$ is like a filter, which makes all the values satisfying $|z + c| > 0$ close to 0. In another word, $e^{-\left( z + c \right)^2}$ makes $z \in [-L - \omega/\lambda, L - \omega/\lambda]$ be close to 0.

Algorithm 1: Calculating the Gaussian components of $f_{Zap}$

input: $\sigma^2 = (\mu_{X_1} - \mu_{X_0}, \sigma^2_{X_1}), s = (\mu_{X_1}, \sigma^2_{X_1}),$ where $\sigma^2_{X_1} \geq \sigma^2_{X_0}, \sigma^2_{X_2}, l, L$

output: $S = \{p_{x_0}, p_{x_1}, \cdots, p_{x_{2^L}}\}, \{\mu_{x_0}, \mu_{x_1}, \cdots, \mu_{x_{2^L}}\},$

1. $\sigma^2_{x_1} \leftarrow \sigma^2_{x_0}$
2. $vUpperVar0 \leftarrow \sigma^2_{x_0}, vLowerVar0 \leftarrow 0; vVar \leftarrow 0$
3. while $|vVar - \sigma^2_{x_0}| > \varepsilon_0$ do
4. $\mu_{x_2} \leftarrow 0; p_{x_2} \leftarrow 1$
5. $vUpperMu0 \leftarrow \maxNumber; vLowerMu0 \leftarrow \minNumber$
6. $Mu \leftarrow 1$
7. while $|Mu| > \varepsilon_0$ do
8. calculate $\sigma^2_{x_0}, \sigma^2_{x_1}$ and $2 \lambda$ by Eq. (24)/(23);
9. calculate $\mu_{x_0}, \mu_{x_1}, \omega$ by Eq. (27)/(26);
10. calculate $p_{x_0}, p_{x_1}, C$ by Eq. (30)/(29);
11. for $i = 0$ to $2^L - 1$ do
12. calculate $\sigma^2_{x_i}, \sigma^2_{x_{i+1}}$ by Eq. (24);
13. calculate $\mu_{x_i}, \mu_{x_{i+1}}$ by Eq. (27);
14. calculate $p_{x_i}, p_{x_{i+1}}$, by Eq. (30);
15. end
16. $S0 = (p_{x_0}, p_{x_1}, \cdots, p_{x_{2^L}})$
17. for $k = 1$ to $2^L - 1$ do
18. calculate $S_k$ by Eq. (40);
19. end
20. $S = \sum_{k=0}^{2^L-1} S_k / \sum_{k=0}^{2^L-1} S_k$;
21. $(p_{x_0}, p_{x_1}, \cdots, p_{x_{2^L}}) = S$;
22. $Mu \leftarrow \sum_{k=0}^{2^L-1} p_{x_i} \mu_{x_i}$
23. if $Mu > 0$ then
24. $vUpperMu0 \leftarrow \mu_{x_0}$;
25. else
26. $vLowerMu0 \leftarrow \mu_{x_1}$;
27. end
28. $vUpperVar0 \leftarrow \sum_{i=0}^{2^L-1} p_{x_i} (\sigma^2_{x_i} - \mu_{x_i}^2)$;
29. if $vVar > \sigma^2_{x_0}$ then
30. $vUpperVar0 \leftarrow \sigma^2_{x_0}$;
31. else
32. $vLowerVar0 \leftarrow \sigma^2_{x_0}$;
33. end
34. $\sigma^2_{x_0} \leftarrow \sum_{i=0}^{2^L-1} p_{x_i} (\sigma^2_{x_i} - \mu_{x_i}^2)$
35. end
36. $\sigma^2_{x_0} \leftarrow \sum_{i=0}^{2^L-1} p_{x_i} \sigma^2_{x_i}$
37. end

Since translation of $f_x$ does not have influence on the result, at first we transmit $f_x$ to satisfy $\mu_{X_0} = -\omega/\lambda$. Suppose $|\mu_{X_0} - \mu_{X_1}| > 3 \sigma^2_{X_1}$. Suppose $|\mu_{X_0} - \mu_{X_1}| > 3 \sigma^2_{X_1}$, then $f_x \approx f_z$ is close to 0, so Eq. (46) is approximated by

$$f_x(z) = C \left( f_{x_0}^* f_z \right) e^{-\left( z + c \right)^2},$$

(47)

which is transmitted to the one Gaussian component problem as Section 4.2. Thus Gaussian distribution converges to the optimal solution.

8. Simulation

In this section, we inspect the performance of our methods constructing $f_{Zap}$ to approximate the optimal noise addition distribution $f_z$, especially the two special cases. $\min I(X; X + Z)$ can be calculated directly by optimization problem (8). $I(X; X + Z)$ and $I(X; X + Z_2)$ denote the results that the data distribution is $f_x$, the noise distributions are $f_{Zap}$ and Gaussian distribution respectively.
8.1. Arbitrary cases

Algorithm 1 calculates arbitrary \( f_x \) consisting of two Gaussian components. Three examples are shown in Figs. 14–16, where we change \( \sigma_x^2, \mu_x, \) and \( p_x \), respectively (we choose \( l = 4 \) and \( L = 13 \)). The performance of approximation is good that only a little difference to the optimal result, regardless of the change of the parameters of \( f_x \). From Algorithm 1, the accuracy of algorithm are based on two aspects. One is \( l \), which is related to the period of \( \mu_z \). Strictly speaking, the period of \( \mu_z \) is infinite. The periodicity is an approximative property. The other is \( L \), which converges to infinite to calculate the stable weights of 2\(^2\) Gaussian functions contained in \( f_{zap} \). The deviation exists because of these two constraints. In Fig. 14, the maximum deviation is 0.00073 at \( \sigma_x^2 = 60 \). The minimum deviation is 0.00052 at \( \sigma_x^2 = 100 \). That is because with the increase of \( \sigma_x^2 \), the period \( 2^l \) is more accurate. In Fig. 15, the maximum deviation is 0.00062 at \( \mu_x = 40 \). The minimum deviation is 0.0000005 at \( \mu_x = 0 \). The reason is the approximation of periodicity is more accurate when \( \mu_z \) is not 0 (when \( \mu_z \) is not 0, the period of \( \mu_z \) can be regarded as \( 2 \), so \( l = 4 \) is more accurate). In Fig. 16, the maximum and minimum deviations do not have the apparent rules as the former two examples. The maximum and minimum deviation are 0.00039 at \( p_x = 0.5 \) and 0.0000014 at \( p_x = 0.9 \) in our sample points. \( p_x \) is related to \( r_i \), \( i = 0, \ldots, 2^l - 1 \), which is more accurate when \( L \) is larger. Next we focus on two special cases, investigating the accuracy of the approximation.

8.2. Case 1: the period of \( \mu_z \) is 2

In this case we can only use the first 2 components to construct \( f_{zap} \). Fig. 17 shows the performance of \( f_{zap} \) when \( t_1 \to 1 \) and \( t_0 \to 0 \). From the figure, \( t_1 \) is fixed (0.9), when \( t_0 \) is smaller and smaller, the deviation between \( f_{zap} \) and the real optimal distribution is decreasing.

Fig. 18 shows the performance of \( f_{zap} \) when \( t_1 \to 1 \) and \( |\mu_{x_1} - \mu_{x_2}| \to 0 \). Under these two constraints, \( f_{zap} \) can be expressed by 2 Gaussian components too. \( f_z \) performs better and better with the decrease of \( |\mu_{x_1} - \mu_{x_2}| \). Actually, in this case \( f_z \) converges to Gaussian distribution, so Gaussian distribution also performs good. Fig. 19 shows the comparison of Gaussian distribution and \( f_{zap} \). When \( |\mu_{x_1} - \mu_{x_2}| \leq 15 \), \( f_{zap} \) performs better than Gaussian distribution. Particularly, when \( |\mu_{x_1} - \mu_{x_2}| \leq 10 \), \( f_{zap} \) and the real optimal distribution is almost the same. When \( |\mu_{x_1} - \mu_{x_2}| > 15 \), the condition
8.3 Case 2: $|\mu_X - \mu_X| \rightarrow 0$

Fig. 20 shows the performance of $f_{zap}$ when $|\mu_X - \mu_X| \gg 0$. In this case, Gaussian distribution is very close to the optimal one. From the figure, when $|\mu_X - \mu_X| > 50$, $I(X; X + Z) - \min I(X; X + Z)$ decreases with the increase of $|\mu_X - \mu_X|$. It is because with the increase of $|\mu_X - \mu_X|$, the values of $x$ in the range $2L$ (determined by $\exp(-2/L \cdot \sigma^2_0)$, shown in Fig. 13) are smaller and smaller. From the figure, we also find out when $|\mu_X - \mu_X| < 50$, $I(X; X + Z) - \min I(X; X + Z)$ increases with the increase of $|\mu_X - \mu_X|$. It is because $f_x$ is close to Gaussian distribution when $|\mu_X - \mu_X| \rightarrow 0$, where Gaussian distribution is optimal noise addition distribution.

8.4 Comparison

In noise addition method, some noise distributions such as Homogeneous distribution (e.g. Agrawal and Srikant, 2000) and Laplace distribution (e.g. Dwork, 2006) have been used. Fig. 21 shows the privacy-preserving capabilities of these three noise distributions, where $f_{ap}$ is the approximation of optimal noise distribution from our method. From the figure, we could find that the mutual information of $f_{ap}$, which measures the privacy protection strength, is the smallest in these three distribution. It means that by adding the noise following the distribution function $f_{ap}$, the attacker gets the least information of the true value from the perturbed value.

Fig. 22 illustrates the privacy-preserving capabilities of $f_{ap}$, Homogeneous distribution and Laplace distribution when $f_x$ satisfies $t_1 \rightarrow 1$ and $|\mu_X - \mu_X| \rightarrow 0$ ($-30 \leq \mu_X \leq 30$), and $|\mu_X - \mu_X| > 0 (40 \leq \mu_X)$. In the former situation, $f_{ap}$ can be expressed by 2 Gaussian components and be calculated in $O(1)$ times. In the letter situation, $f_{ap}$ is Gaussian distribution. From the figure, in the two situations $f_{ap}$ is the best one to protect the individual privacy among the three distributions.
8.5. Performance

From Section 7.1, the complexity of Algorithm 1 is $O(2^L)$. Fig. 23 illustrates the time cost with different $L$. Here we calculate $f_{ap}$ by Algorithm 1, where $f_X = (0.3, 0, 100)(0.7, -30, 60)$. The algorithm is coded by MATLAB, which runs on the notebook with Intel®Core™i5 CPU (2 × 2.40 GHZ) and 4 GB RAM. From the figure, when the larger $L$ is chosen, the time cost grows up with exponential rate, which limits the application of the algorithm.

9. Conclusion and future work

In this paper, we quantify the accuracy of result and the privacy of individuals. Based on the metrics, we propose the optimization problem, and find out the optimal noise distribution that provides best privacy protection while maintaining the acceptable deviation from the accurate result. For the special cases that the original data of individuals follows Gaussian distribution, Gaussian distribution is the optimal result. For the general case that the original data follows arbitrary continuous distribution, the optimal solution consists of infinite number of Gaussian components. We give a way to calculate the components. By analyzing the properties of the components, the optimal distribution can be approximated by finite components. By simulation the approximation also performs good.

This paper is just the beginning. The complexity of Algorithm 1 is $O(2^L)$, which limits the algorithm when a large $L$ is needed. That is because Algorithm 1 is based on Eq. (40), which calculate $S$ in an iterative way. Next a non-iterative general formula is needed, which will break the shackle of $L$. Furthermore, the extended algorithm of Algorithm 1, with the $f_X$ composed of multiple Gaussian components as input and the optimal noise distribution as output, is needed, which could be applied for most of the aggregation applications in practice.

Acknowledgments

The work was supported partly by National Natural Science Foundation of China under Grant No. 61371192, National Natural Science Foundation of China under Grant No. 61271271, 100 Talents Program of Chinese Academy of Science, the Strategic Priority Research Program of the Chinese Academy of Sciences under Grant XDA06030601, and the Funding of Science and Technology on Information Assurance Laboratory under Grant KJ-13-003.

Appendix A. The proof of Theorem 1

Proof. Firstly we consider a more general problem...
\[ \begin{align*}
\min_{f_{XY}} (I(X;Y)) &= I(X;Y) \\
\int \int f_X(x)f_{XY}(y|x)(y-x)^2 \, dx \, dy &\leq \sigma_m^2 \\
\text{s.t.} \int f_{XY}(y|x)(y-x) \, dy &= 0 \\
\int f_{XY}(y|x) \, dy &= 1,
\end{align*} \]
(A.1
where \( Y = Z + X \). If \( Z \) is independent of \( X \), we have
\[ f_{Y|X}(y|x) = f_{Y|Z}(y|x) = f_{ZX}(z|x) = f_Z(z). \]
The constraints become
\[ \int f_X(x)f_{Y|X}(y|x)(y-x)^2 \, dx \, dy = \int f_Z(z) \, dz \]
(A.3
\[ \int f_X(x)f_{Y|X}(y|x)(y-x) \, dy = \int f_Z(z) \, dz \]
(A.4
\[ \int f_{XY}(y|x) \, dy = \int f_Z(z) \, dz. \]
(A.5
This problem is translated to the problem (8). In other words, the problem (8) is a special case of the problem (A.1).

The mutual information \( I(X;Y) \) is
\[ \begin{align*}
I(X;Y) &= \int \int f_{XY}(x,y) \log \frac{f_{XY}(x,y)}{f_X(x)f_Y(y)} \, dx \, dy \\
&= \int \int f_X(x)f_{Y|X}(y|x) \log \frac{f_{Y|X}(y|x)}{f_Y(y)} \, dx \, dy,
\end{align*} \]
where \( f_Y(y) = \int f_X(x)f_{Y|X}(y|x) \, dx. \)
We use the method of Lagrange multipliers to find the solution. Set up the functional
\[ J = \int \int f_X(x)f_{Y|X}(y|x) \log \frac{f_{Y|X}(y|x)}{f_Y(y)} \, dx \, dy + \lambda \int \int f_X(x)f_{Y|X}(y|x)(y-x)^2 \, dx \, dy + \mu \int u(x) \int f_{XY}(y|x) \, dy \, dx. \]
(A.7
Differentiating with respect to \( f_{Y|X}(y|x) \), we have
\[ \frac{\partial J}{\partial f_{Y|X}(y|x)} = f_X(x) \log \frac{f_{Y|X}(y|x)}{f_Y(y)} + f_X(x) - \int f_X(x)f_{Y|X}(y|x') \frac{1}{f_Y(y)} f_X(x') \, dx' + i_f(x)(y-x)^2 + u(x)(y-x) + u(x) \\
= f_X(x) \log \frac{f_{Y|X}(y|x)}{f_Y(y)} + i_f(x)(y-x)^2 + u(x)(y-x) + u(x) \\
= 0. \]
(A.8
Set \( \omega(x) = u(x)/f_x(x) \) and \( \tau(x) = v(x)/f_x(x) \),
\[ f_x(x) \left[ \log \frac{f_{Y|X}(y|x)}{f_Y(y)} + i_f(y-x)^2 + \omega(x)(y-x) + \tau(x) \right] = 0. \]
(A.9
Thus
\[ f_{Y|X}(y|x) = f_Y(y)e^{-i_f(y-x)^2 - \omega(x)(y-x) - \tau(x)}. \]
(A.10
Here we get the expression of \( f_{Y|X}(y|x) \). In problem (8), \( Y = X + Z \) and \( Z \) is independent of \( X \). From Eq. (A.10),
\[ f_Z(z) = f_X(x+z) e^{-i_f(z-x) - \omega(z-x) - \tau(z-x)}. \]
(A.11

Appendix B. The proof of Lemma 1

Proof. Since
\[ \sigma_{Z_1}^2 = \sigma_{Z_2}^2 = \sigma_{Z_3}^2 \leq \sigma_{Z_1}^2 + \sigma_{Z_2}^2 = \sigma_{Z_3}^2, \]
we have
\[ \frac{1}{\sigma_{Z_1}^2} = \frac{1}{\sigma_{Z_2}^2} + \frac{2\lambda}{\sigma_{Z_2}^2} \geq \frac{1}{\sigma_{Z_3}^2} + \frac{2\lambda}{\sigma_{Z_3}^2} = \frac{1}{\sigma_{Z_0}^2}. \]
(B.2
So \( \sigma_{Z_0}^2 \geq \sigma_{Z_1}^2 \). The equation holds when \( \sigma_{Z_0}^2 = \sigma_{Z_1}^2 \).

Appendix C. The proof of Lemma 2

Proof. For \( j = 0, \sigma_{Z_2}^2 = \sigma_{Z_1}^2 \).
For \( j \in \mathbb{N}^* \), from Eq. (24) we have
\[ \frac{\sigma_{Z_2}^2}{\sigma_{Z_1}^2} = \frac{\sigma_{Z_2}^2}{\sigma_{Z_2}^2 + 2\lambda \sigma_{Z_0}^2} = 1 - \frac{\sigma_{X_0}^2}{\sigma_{Z_2}^2 + \sigma_{X_0}^2} + \frac{\sigma_{X_0}^2 \sigma_{Z_2}^2}{(\sigma_{Z_2}^2 + \sigma_{X_0}^2)^2} = 1 + \frac{\sigma_{X_0}^2 (\sigma_{Z_2}^2 - \sigma_{Z_0}^2)}{(\sigma_{Z_2}^2 + \sigma_{X_0}^2) (\sigma_{Z_2}^2 + \sigma_{X_0}^2)} \leq 1 \]
(C.1
where (a) follows from \( \sigma_{Z_2}^2 \leq \sigma_{Z_0}^2 \). So \( \sigma_{Z_1}^2 \geq \sigma_{Z_2}^2 \). The equation holds when \( \sigma_{Z_2}^2 = \sigma_{Z_1}^2 \).

Appendix D. The proof of Theorem 3

Proof. Consider the series
\[ \left( \frac{1}{\lambda + \frac{1}{s_1}}, \frac{1}{\lambda + \frac{1}{s_2}}, \ldots \right) \]
denoted by \( (s_1, s_2, \ldots) \). Since
$$s_k = \frac{1}{2\lambda + \frac{1}{\sigma_{X_1}^2}}$$

and

$$s_{k+1} = \frac{1}{2\lambda + \frac{1}{\sigma_{X_1}^2 - \lambda}}$$

where $\gamma_k = 1/2 \lambda + 1/\sigma_{X_1}^2 > 0$, we have $s_{k+1} > s_k$. Since

$$s_{k+1} = \frac{1}{2\lambda + \frac{1}{\sigma_{X_1}^2 - \lambda}}$$

we have

$$\lim_{k \to \infty} s_k = \frac{\sqrt{\gamma_1^2 \sigma_{X_1}^2 + 2 \lambda \sigma_{X_1} - \lambda^2 \sigma_{X_1}^2}}{2\lambda} = \frac{1}{2} \left( \sigma_{X_1}^2 + \frac{2}{\lambda} \sigma_{X_1}^2 - \sigma_{X_1}^2 \right).$$

Appendix F. The proof of Lemma 4

**Proof.** For any $j \in \{2^i, 2^{i+1}, \ldots, 2^t\}, t > 0$ can be expressed as $(j_0, j_1, \ldots, j_t)$, where $j_0 = 1$.

$$\sigma_{Z_2}^2(j_0, j_1, \ldots, j_t) = \frac{1}{2\lambda + \frac{1}{\sigma_{X_1}^2}}$$

$$\leq \frac{1}{2\lambda + \frac{1}{\sigma_{X_1}^2}} = \sigma_{Z_2}^2(j_0, 0, \ldots, 0) = \sigma_{Z_2}^2.$$

If $j$ is even, $j_i = 0$, we have

$$\sigma_{Z_2}^2(j_0, 0, \ldots, 0) \geq \sigma_{Z_2}^2(j_0, 1, \ldots, 0).$$

Similarly

$$\sigma_{Z_2}^2(j_0, 0, \ldots, 0) \geq \sigma_{Z_2}^2(j_0, 1, \ldots, 1).$$

If $j$ is odd, $j_i = 1$, we have

$$\sigma_{Z_2}^2(j_0, 1, \ldots, 1) \leq \sigma_{Z_2}^2(j_0, 1, \ldots, 0).$$

Appendix G. The proof of Lemma 5

**Proof.** Suppose $j$ is even, and $j \in \mathbb{N}^*$. From Lemma 2 and Lemma 4 we know $(\sigma_{Z_2}^2, \sigma_{Z_3}^2, \ldots, \sigma_{Z_{2^t-1}}^2)$, each of which is the largest variance with even indices in its level, is a non-decreasing sequence. Thus the max $\sigma_{Z_1}^2 = \lim_{t \to \infty} \sigma_{Z_1}^2$, and

$$\max \sigma_{Z_1}^2 = \frac{1}{2\lambda + \frac{1}{\sigma_{X_1}^2}}.$$

Consider the sequence $(\sigma_{Z_2}^2, \sigma_{Z_2}^2, \ldots, \sigma_{Z_{2^t-1}}^2)$, each of which is the smallest variance with even indices in its level. Since

$$\sigma_{Z_2}^2(j_0, j_1, j_2, \ldots, j_t) = \sigma_{Z_2}^2(j_0, j_1, j_2, \ldots, j_{t-1}) + \sigma_{Z_2}^2(j_0, j_1, j_2, \ldots, j_{t-1}, j_t)$$

from which we have

$$\max \sigma_{Z_1}^2 = \sigma_{Z_1}^2.$$

Consider the sequence $(\sigma_{Z_2}^2, \sigma_{Z_2}^2, \ldots, \sigma_{Z_{2^t-1}}^2)$, each of which is the smallest variance with even indices in its level. Since

$$\sigma_{Z_2}^2(j_0, j_1, j_2, \ldots, j_t) = \sigma_{Z_2}^2(j_0, j_1, j_2, \ldots, j_{t-1}) + \sigma_{Z_2}^2(j_0, j_1, j_2, \ldots, j_{t-1}, j_t)$$

for any $j \in \mathbb{N}$. Thus the sequence is a non-increasing sequence, from which we have min $\sigma_{Z_1}^2 = \lim_{t \to \infty} \sigma_{Z_{2^t-1}}^2$. So

$$\min \sigma_{Z_1}^2 = \lim_{t \to \infty} \sigma_{Z_{2^t-1}}^2.$$

Similarly, for any $j \in \mathbb{N}^*$, if $j$ is odd, we have

$$\min \sigma_{Z_1}^2 = \lim_{t \to \infty} \sigma_{Z_{2^t-1}}^2 = \frac{1}{2\lambda + \frac{1}{\sigma_{X_1}^2}}.$$
From Eq. (33) we have
\[ m \] represented in the way as
\[ s \]
minimum is
\[ Z \] when
\[ l \]

Appendix I. The proof of Theorem 4

Proof. From Lemma 5 \( m \) with odd indices are in one interval, and \( m \) with even indices are in another interval. The ratio of maximum of the interval of variances with odd indices to minimum is
\[ \frac{m}{s_{\infty}} \]

When \( l \to 1 \), the ratio converges to 1. Shown in Fig. 6, when \( m \) is not much larger than \( m \), \( s_{\infty}/s_{\infty} \) is close to 1, i.e. all the variances with odd indices can be regarded as the same. The ratio of maximum of the interval of variances with even indices to minimum is
\[ \frac{m}{s_{\infty}} = \frac{1}{l + 1} \left( \sqrt{l + 1} + 4l + 1 \right) \]

When \( l \to 1 \), the ratio converges to 1. The same result is shown by Fig. 7 for the variances with even indices. \( \square \)

Appendix J. The proof of Lemma 7

Proof. From Eqs. (33)–(36), any \( \mu_{Zj} \) at level \( l,j \in [2^{l-1}, 2] \), can be represented in the way as
\[ \mu_{Zj} = \mu_{Z0} + (1 - \mu_{Z0})\mu_{Z1} \]

where
\[ \mu_{Z0} = \frac{H_0}{1 - H_0} \]
and \( i, j, \ldots, h_1 \) is \( \{0, 1, 2, 3\} \). Since \( 0 < \mu_{Z0}, \mu_{Z2}, \mu_{Z3} < 1 \), we have
\[ \lim_{l \to \infty} \mu_{Z0}, \mu_{Z1}, \mu_{Z2}, \mu_{Z3} = 0, \]

which means when \( l \to \infty \), \( \mu_{Zi} \) is only decided by \( \mu_{Z0} \) and \( \mu_{Z2} \), \( i = 0, 1, 2, 3 \), not by \( \mu_{Zi} \) (there are two "\( \mu_{Zi} \)" in Eq. (J.1). This "\( \mu_{Zi} \)" is a variable in "\( \mu_{Z0} \), \( \mu_{Z2} \), \( \mu_{Z3} \)", which is the root of the tree. Another "\( \mu_{Zi} \)" is a constant in \( \mu_{Zj} \), which actually is a part of \( \omega \). If \( \omega \) is fixed, the "\( \mu_{Zi} \)" in \( \mu_{Zj} \) is fixed). Thus no matter what the root \( \mu_{Z0} \) is, \( \mu_{Zi} \) at level \( l \to \infty \) is the same.

Specially, when \( l \to 1 \), in Eq. (J.1) \( \mu_{Z0}, \mu_{Z1}, \mu_{Z2} \to \infty \) even when \( l \) is a small number. \( \square \)

Appendix K. The proof of Theorem 5

Proof. Suppose \( l = 2l \), where \( l \) is large enough to satisfy Lemma 6 and Lemma 7. We will prove the two sequences \( \{\mu_{Z0}, \mu_{Z2}, \ldots, \mu_{Zj}, \mu_{Zj+1}, \ldots, \mu_{Zp}\} \) and \( \{\mu_{Z0}, \mu_{Z2}, \ldots, \mu_{Zj+1}, \ldots, \mu_{Zp}\} \) are the same.

Each sequence has \( 2^l \) elements. We divide the elements into two parts, and consider each part respectively, where \( d_1 = 2l - 1 \).

1) \( \{\mu_{Z0}, \ldots, \mu_{Zj}\} \) and \( \{\mu_{Z0}, \ldots, \mu_{Zj+1}\} \).

From Lemma 6 we have \( \mu_{Z0} = \mu_{Z1} = \cdots = \mu_{Z2} = \mu_{Z2} = \mu_{Z2} \), and \( 2^{l-1} = \cdots = 2^{l-1} = 0 \) mod 4. Consider the right subtree of \( \mu_{Z0} \), \( \mu_{Z1} \) is the right child of \( \mu_{Z0} \), \( \mu_{Z2} \) is the right child of \( \mu_{Z1} \). So \( \mu_{Z0} = \mu_{Z2}, \mu_{Z1} = \mu_{Z2}, \mu_{Z2} = \mu_{Z2} \), and \( \mu_{Z3} \) are the second level right children of \( \mu_{Z2} \). \( \mu_{Z2} \) and \( \mu_{Z3} \) are the second level right children of \( \mu_{Z2} \). So \( \mu_{Z0} = \mu_{Z2}, \mu_{Z1} = \mu_{Z2}, \mu_{Z2} = \mu_{Z2} \). In this way we pair all the other nodes which are at the level less than \( l_1 \). So the two subsequences are equal.

2) \( \{\mu_{Z1}, \ldots, \mu_{Zj}\} \) and \( \{\mu_{Z1}, \ldots, \mu_{Zj+1}\} \).

\( \mu_{Z0}, \ldots, \mu_{Zj} \) are the right children at the level \( l_1 \) of \( \mu_{Z2} \). Consider any level \( l_2 \), where \( l_1 < l_2 \leq \), the nodes are \( \{\mu_{Zj}, \ldots, \mu_{Zj+1}\} \), which are the \( l_2 \)th level right children of \( \mu_{Z2} \). The corresponding sets \( \{\mu_{Zj}, \ldots, \mu_{Zj+1}\} \) are the \( l_2 \)th level right children of \( \mu_{Z2} \). Since \( l_1 \leq l_2 \), from Lemma 7 we know \( \{\mu_{Zj}, \ldots, \mu_{Zj+1}\} \) equals to \( \{\mu_{Zj}, \ldots, \mu_{Zj+1}\} \). So we can find the two subsequences are equal.

From above it shows the two sequences are the same. So \( 2^l \) is the period of \( \mu_{Z} \). \( \square \)
\[
p_{x_{i+1,3,3}} = C_{x_{i+1,3,3}} \frac{p_{x_{i+1}}p_{x_3}}{\sqrt{\sigma_{x_{i+1}}^2 + \sigma_{x_3}^2}} e^{-\frac{1}{2} \left( \frac{r_{x_{i+1}} - r_{x_3}}{\sigma_{x_{i+1}} + \sigma_{x_3}} \right)^2}. \]

(L2)

So

\[
p_{x_{i+1,3,3}} = \frac{p_{x_{i+1}}p_{x_3}}{p_{x_i}}. \]

(L3)

Similarly, for any \( l' \geq l \) and \( i = 0, 1, \ldots, 2^l - 1 \),

\[
p_{x_{i+1,3,3}} = \frac{p_{x_{i+1}}p_{x_3}}{p_{x_i}}. \]

(L4)

\[\Box\]

Appendix M. The proof of Theorem 7

Proof. From Eq. (L1), \( r_0 \) and \( r_1 \) are

\[
r_0 = \frac{p_{x_1}e}{p_{x_0}}, \quad \frac{p_{x_1}e}{p_{x_0}}. \]

(M.1)

\[
r_1 = \frac{p_{x_1}e}{p_{x_0}}, \quad \frac{p_{x_1}e}{p_{x_0}}. \]

(M.2)

From Theorem 3, \( \sigma_{x_k}^2 = \sigma_{x_k}^2 \) and \( \sigma_{x_k}^2 = \sigma_{x_k}^2 \), so

\[
\frac{1}{r_0} \left( r_{x_1} - r_{x_0} - r_{x_3} \right) \]

(M.3)

\[
\frac{1}{r_1} \left( r_{x_1} - r_{x_0} - r_{x_3} \right) \]

(M.4)

Thus

\[
\frac{r_0}{r_1} \left( \frac{1}{r_1} - \frac{1}{r_0} \right). \]

(M.5)

So if \( |\mu_{x_1} - \mu_{x_0}| \rightarrow 0 \) or \( |\mu_{x_1} - \mu_{x_3}| \rightarrow 0 \), \( r_0 = r_1 \).

\[\Box\]

REFERENCES


Eisenman SB, Miluzzo E, Lane ND, Peterson RA, Ahn GS, Campbell AT. The case for VM-based cloudlets in mobile computing. ACM Trans Sens Netw 2009;6.


Oliveira SRM, Zaiane OR. Privacy preserving clustering by data transformation. In: The 18th Brazilian symposium on databases; 2003.


Hao Zhang received his B.Eng. degree in information security from University of Science and Technology of China (USTC), Hefei, in 2005. He is currently a Ph.D. candidate at School of Information Science and Technology, University of Science and Technology of China. His research interest focuses on privacy protection, mobile security, and P2P network.

Nenghai Yu received the B.S. degree in 1987 from Nanjing University of Posts and Telecommunications, the M.E. degree in 1992 from Tsinghua University, and the Ph.D. degree in 2004 from University of Science and Technology of China, where he is currently a professor. His research interests include security and privacy in mobile networks, multimedia security, multimedia information retrieval, and information hiding.

Yonggang Wen received the Ph.D. degree in electrical engineering and computer science from the Massachusetts Institute of
Technology, Cambridge, MA, USA, in 2008. He is currently an Assistant Professor with the School of Computer Engineering, Nanyang Technological University, Singapore. Previously, he was with Cisco, San Jose, CA, USA, as a Senior Software Engineer and a System Architect for content networking products. His current research interests include cloud computing, mobile computing, cyber security, multimedia networks, and green ICT.

Weiming Zhang received the M.S. and Ph.D. degrees in 2002 and 2005, respectively, from the Zhengzhou Information Science and Technology Institute, China. Currently, he is an associate professor with the School of Information Science and Technology, University of Science and Technology of China. His research interests include information hiding, multimedia security, and privacy-preserving data searching and analysis.