An Efficient Approximation Method for Calculating Confidence Level of Negative Survey

Ran Liu, Jinghui Peng, Shanyu Tang*

School of Computer Science, China University of Geosciences, Wuhan, China

Abstract

The confidence level of negative survey is one of the key scientific problems. Present work uses generation function to analyse the confidence level, and uses a greedy algorithm to calculate that, which is used to evaluate the dependable level of negative survey. However, the present method is low efficiency and complex. This study focuses on an efficient approximation method for calculating the confidence level of negative survey. This approximation method based on central limit theorem and Bayesian method can get the results efficiently.

Keywords: Privacy protection, Data collection, Negative survey, Confidence Level, Bayesian method

1. Introduction

Artificial Immune System simulates the mechanism of biology immune system to model and design effective algorithm for solving some complex issues. Negative selection principle [1] is one of the unique mechanisms of biology immune system, and the implication of negative selection principle is that the immaturity T cell dies if it *matches* with itself as it grows, and it survives if it *mismatches* with itself. Inspired by negative selection principle, the negative selection algorithm [2] is proposed and can be used for network security, virus detection [3, 4] and anomaly detection [5].

*Corresponding author: Tel: 86-27-67848563.

Email addresses: ranliu_cug@foxmail.com (Ran Liu), pjh826625501@163.com (Jinghui Peng), shanyu.tang@gmail.com (Shanyu Tang)

Similarly, the negative survey [6], which is inspired by negative selection principle, is a novel and promising indirect question method for information security and enhancing privacy in collecting sensitive data and individual privacy [7]. Negative surveys consist of a question and $c(c \geq 3)$ categories for the interviewees to select. In contrast to traditional surveys, the participants are required to select a category that does *not* agree with the fact [6, 8], i.e. randomly select a category from the other $c-1$ unreal categories. For convenience, it defines *positive category* as the category that agrees with the fact, while *negative category* as the other $c-1$ categories that does *not* agree with the fact [6].

The negative survey method can attain privacy protection with lower power and higher degree, and boost participants' confidence. The main calculation of collecting sensitive data with negative survey is reconstructing the corresponding positive survey in the central processor. The privacy preserving properties of negative survey do not rely on anonymity, cryptography or any legal contracts, but rather participants not revealing their own privacy information. And the negative survey method is applicable to collecting data at a high speed in low-powered mobile devices such as smart phones, tablets and so on [9].

The positive survey can be reconstructed from a result of negative survey. For a survey consist of a question and $c(c \geq 3)$ categories for n interviewees to select, a negative survey result is $R = (r_1, r_2, ..., r_c)$, where r_i is the results of category i in negative survey. Meanwhile, the original positive survey is $T = (t_1, t_2, ..., t_c)$, where t_i is the number of interviewees belonging to category i. Define $v_{i,j}$ as the probability that category i is chosen given that a respondent positively belongs to category j, where $\sum_{i=1}^{c} v_{i,j} = 1$ and $v_{i,i} = 0$. Define the probability matrix as V as Formula (1), and $R = TV$ and $T = RV^{-1}$. In consequence, the positive survey T can be reconstructed from a negative survey R.

$$V = \begin{bmatrix} 0 & v_{1,2} & \cdots & v_{1,c} \\ v_{2,1} & 0 & \cdots & v_{2,c} \\ \vdots & \vdots & \ddots & \vdots \\ v_{c,1} & v_{c,2} & \cdots & 0 \end{bmatrix} \tag{1}$$

Generally, $v_{i,j}|_{i \neq j} = 1/(c-1)$, which means the probability of selecting negative categories follows uniform distribution [6]. Following the work in [6], Xie et al. proposed Gaussian Negative Survey (GNS) [10], where the

probabilities of selecting negative categories (i.e. $v_{i,j}$) follow a Gaussian distribution centered at the corresponding positive category. The GNS could attain higher accuracy but lower ability of privacy protection.

The traditional reconstructing method in [6] may lead the reconstructed positive survey with negative values. Based on the problem, two method [11] were proposed for reconstructing positive survey which had no negative values. In [12], Bao et al. proposed a greedy algorithm for calculating the confidence level, which is analysed in generating function. But this method is low efficient and complex, and couldn't achieve the high efficiency of negative survey.

In this study, an efficient approximation method is proposed to calculate the confidence level of negative survey. This work reinforce the efficiency of negative survey.

In the remainder of this study, Section 2 introduces the related work of this study. Section 3 describes the problem in this study. Section 4 describes the efficient approximation method. Section 6 discusses some existing problems of this approximation method and Section 7 concludes the whole study.

2. Related Work

In this study, the probability of selecting negative categories follows uniform distribution (i.e. $v_{i,j}|_{i\neq j} = 1/(c-1)$) as general negative survey in [6, 8, 11, 12]. So in this section, the related work of negative survey [6, 8, 11, 12] is introduced. For convenience, some definitions are given in Figure 1.

Define n as the number of interviewees participating the negative survey, and c as the number of categories. The results of the negative survey are $R = (r_1, r_2, \cdots, r_c)$, where $r_i (1 \leq i \leq c, c \geq 3)$ represents the total number of participants who select the i-th category in the negative survey. Similarly, the real positive survey is $T = (t_1, t_2, \cdots, t_c)$, and $n = \sum_{i=1}^{c} r_i = \sum_{i=1}^{c} t_i$. In [6, 8], the reconstructed positive survey can be calculated by Formula (2). In this study, a positive category i, which has n interviewees, c category, and the proportion of category i is p_i, is written as $PS(n, c, p_i)$ for simplicity. And the corresponding negative category is written as $NS(n, c, q_i)$.

$$\begin{cases} \hat{t}_j = n - (c-1)r_j \\ \hat{p}_j = 1 - (c-1)q_j \end{cases} \tag{2}$$

n : the number of interviewees for surveys
c : the number of categories in surveys
r_i : the number of interviewees selecting category i in negative survey
q_i : the proportion of negative category $i(1 \leq i \leq c)$, i.e. $q_i = r_i/n$
t_i : the original number of interviewees in positive category i
$\hat{t}_i$: the estimated number of t_i
R : the participant vector, *i.e.* $R = (r_1, r_2, \cdots, r_c)$
T : the participant vector, *i.e.* $T = (t_1, t_2, \cdots, t_c)$
p_i : the proportion of positive category $i(1 \leq i \leq c)$, i.e. $p_i = t_i/n$
$\hat{p}_i$: the estimated number of p_i, i.e. $\hat{p}_i = \hat{t}_i/n$

Figure 1: The definitions in this study

Although $\hat{p}_j = E(p_j)$, it can be observed that $\hat{p}_i < 0$ when $q_i > 1/(c-1)$. Therefore, this traditional method is not practical sometimes.Following the traditional method in [6, 8], two methods were proposed for reconstructing positive survey in [11]. Method I [11] uses an iteration method to reconstruct the positive survey. The advantage of Method I is that no negative values is in the reconstructed positive survey, i.e. $\hat{p}_i > 0(1 \leq i \leq c)$. But this method only use an implicit function to reconstruct the positive survey approximatively. And the accuracy of this method lacks of theoretical basis.

Method II [11] eliminates the negative values through adjusting the results of reconstructed positive survey. This method sets the negative value of the category in the reconstructed positive survey to 0, and then keeps the sum of the reconstructed positive survey unchanged by the proportion of the values in the other categories. This method is more efficiency than Method I, but there is no theoretical analysis of this method. In [12], the confidence level of negative survey is analysed in generation functions, and calculated in a greedy algorithm.

3. Problem Formulation

Efficient is one of the greatest advantages in collecting data by the negative survey method, because each participant only needs to send one of her or his negative categories (i.e. unreal information). The reconstructed positive survey from negative survey is non-exact values, so there are two important issues, which are the confidence level and the efficient, respectively. It is not

necessary and inefficient to use a generation function method to exactly calculate the confidence level [12] with the non-exact values reconstructed from negative survey. More importantly, it is so complicated to exactly calculate the confidence level that a greedy algorithm used [12].

This study proposes an efficient method, which is analysed by central limit theorem and Bayes method, to calculate the confidence level approximately, and this approximation method can reinforce the efficiency of negative survey. The core concept of this approximation method is using Normal Distribution to approximate the original distribution for fast calculation (more details in Section 4). The Bayes method is then used to calculate the confidence level of each category in negative survey, which is studied based on the analysis of the distribution of possible positive survey results.

4. The Efficient Method of Approximation

This section gives the proposed efficient approximation method for calculating the confidence level. In subsection 4.1, central limit theorem is used to calculate the approximated distribution of q_i. In subsection 4.2, the Bayes method is used to estimate the probability density function of p_i. In subsection 4.3, the confidence level is calculated based on Bayes method.

4.1. The distribution of negative survey

Theorem 4.1 gives the distribution of category i in negative survey when that of positive survey is known.

Theorem 4.1. *For a given positive category $PS(n, c, p_i)$ and the corresponding negative category $NS(n, c, q_i)$, So q_i approximately follows Normal Distribution when n goes to infinity.*

$$\lim_{n\to\infty} P\left(\frac{q_i-\mu}{\sigma^2} \le x\right) = \int_{\infty}^{x} \frac{1}{2\pi} e^{-\frac{t^2}{2}} \mathrm{d}t \tag{3}$$

where $\mu = \frac{1-p_i}{c-1}$, and $\sigma^2 = \frac{(c-2)(1-p_i)}{n(c-1)^2}$.

Proof. Consider the negative category i and calculate the probability distribution of r_i. In the negative survey, $n(1-p_i)$ interviewees are likely to select the i-th category. Define the random variable $X_j (j = 1, 2, \cdots, n-t_i)$. If the j-th interviewee selects the i-th category, $X_j = 1$, or else $X_j = 0$. Obviously,

each X_j is independent and identically distributed, and follows the Binomial Distribution $B(n(1-p_i), 1/(c-1))$. Let $X = \sum_{j=1}^{n(1-p_i)} X_j$. So $r_i = X$, and

$$E(r_i) = \frac{n(1-p_i)}{c-1}, D(r_i) = \frac{n(c-2)(1-p_i)}{(c-1)^2}. \tag{4}$$

Owing to the De Moivre − Laplace central limit theorem, r_i follows Normal Distribution as n goes to infinity, *i.e.*

$$r_i \sim N\left(\frac{n(1-p_i)}{c-1}, \frac{n(c-2)(1-p_i)}{(c-1)^2}\right) \tag{5}$$

So

$$q_i \sim N(\mu, \sigma^2) = N\left(\frac{1-p_i}{c-1}, \frac{(c-2)(1-p_i)}{n(c-1)^2}\right) \tag{6}$$

In consequence, q_i follows the Normal Distribution when n goes to infinity and Theorem 4.1 and Formula 3 are both valid. □

Define $P(q_i|p_i)$ to be the conditional probability density function for q_i with given p_i, so

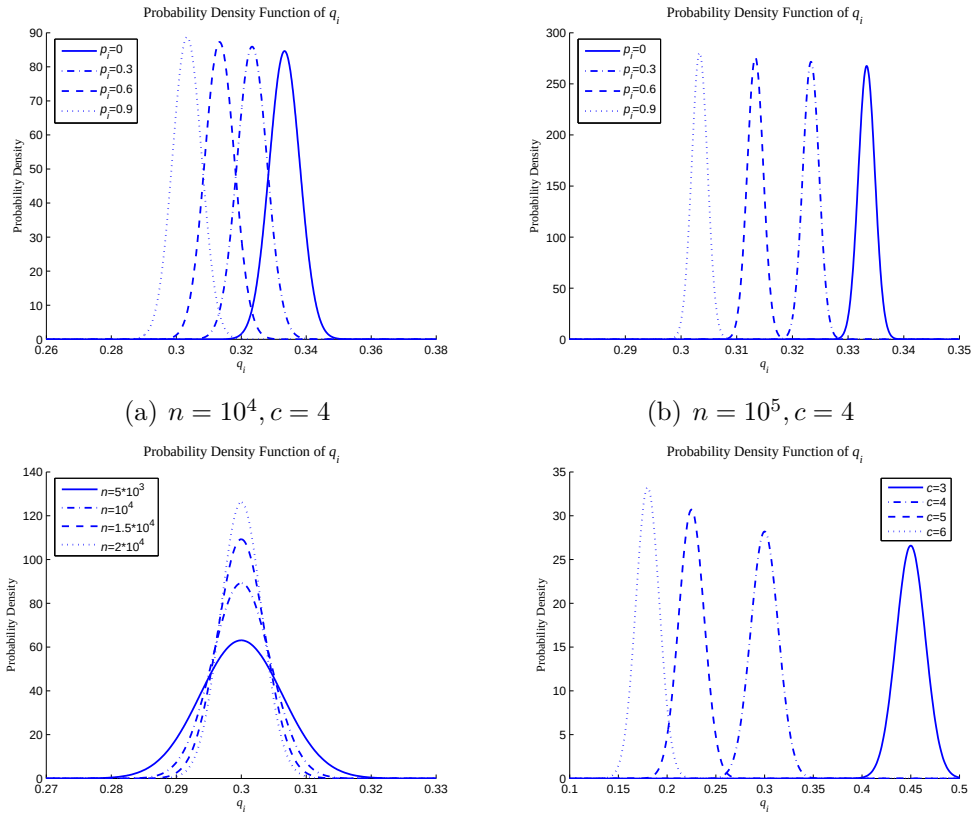
$$P(q_i|p_i) = \frac{(c-1)\sqrt{n}\exp\left(-\frac{n[q_i(c-1)-(1-p_i)]^2}{2(c-2)(1-p_i)}\right)}{\sqrt{2\pi(c-2)(1-p_i)}} \tag{7}$$

Figure 2 illustrates the function cure of Formula (7) varying with p_i, c, or n.

According to the character of Normal Distribution, $P(|q_i - \mu| < 3\sigma) \approx 0.9974$. So we can regard that ${}_{max}q_i = \mu + 3\sigma$ and ${}_{min}q_i = \mu - 3\sigma$. Figure 3 illustrates the range of q_i for different values of p_i when $n = 10^4$ and $c = 4$.

4.2. The distribution of reconstructed positive survey

There are some differences between reconstructing positive survey from a given negative survey and traditional method for parameter estimating. The reason is that the given result of negative survey is only one sample for its original positive survey. In consequence, we use Bayes method to reconstruct the positive survey. The distribution of the reconstructed p_i is given in the following Theorem 4.2.



(a) $n = 10^4, c = 4$ (b) $n = 10^5, c = 4$

(c) $p_i = 0.1, c = 4$ (d) $n = 10^3, p_i = 0.1$

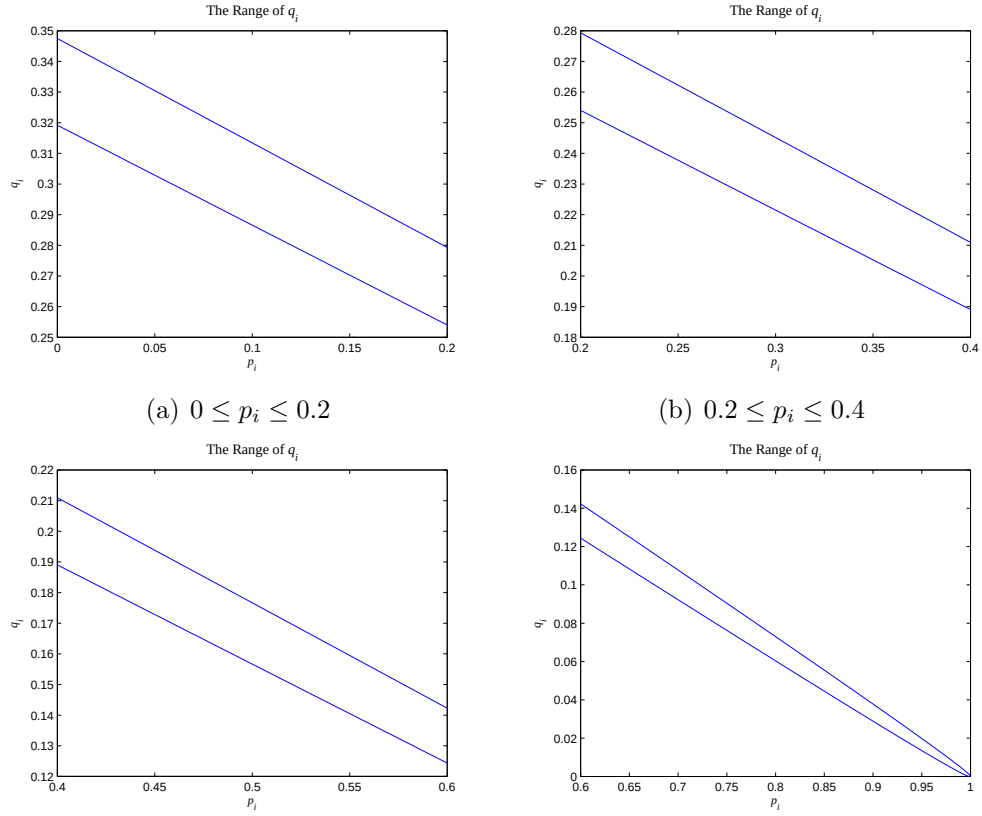
Figure 2: The Function Curve of $P(q_i|p_i)$ with different values of p_i, c, or n

Theorem 4.2. *If a negative category is $NS(n, c, q_i)$, the probability density function of corresponding $PS(n, c, p_i)$ is*

$$\pi(p_i|q_i) = \frac{e^{-\frac{n[(c-1)q_i-(1-p_i)]^2}{2(c-2)(1-p_i)}}/\sqrt{1-p_i}}{\int_0^1 e^{-\frac{n[(c-1)q_i-(1-p)]^2}{2(c-2)(1-p)}}/\sqrt{1-p}\mathrm{d}p} \tag{8}$$

Proof. Define $\pi(p_i)$ to be the prior probability density function of p_i, and $P(q_i|p_i)$ is the conditional probability density function for q_i. According to Bayes Function form of probability density function, the probability density function of p_i with given q_i is the following Formula (9).

$$\pi(p_i|q_i) = \frac{P(q_i|p_i)\pi(p_i)}{\int_0^1 P(q_i|p)\pi(p)\mathrm{d}p} \tag{9}$$



(a) $0 \leq p_i \leq 0.2$ (b) $0.2 \leq p_i \leq 0.4$

(c) $0.4 \leq p_i \leq 0.6$ (d) $0.6 \leq p_i < 1$

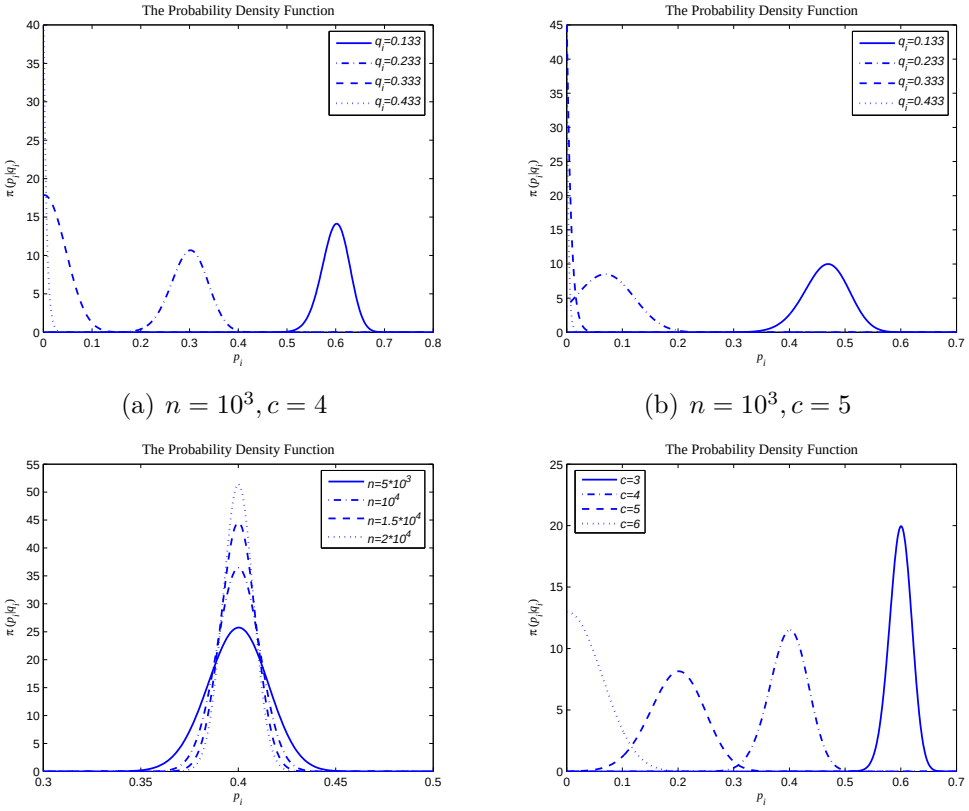
Figure 3: The Range of q_i with probability 0.9974 ($n = 10^4, c = 4$)

Suppose that we have no knowledge of p_i. Based on Bayesian assumption, the prior probability density function $\pi(p_i)$ can be considered as uniform distribution $U(0, 1)$. On this occasion, the density function $\pi(p_i)$ can be calculated in the following Formula (10). In addition, $P(q_i|p_i)$ can be calculated in Formula (7).

$$\begin{cases} \pi(p_i) = 1 & 0 < p_i < 1 \\ \pi(p_i) = 0 & \text{otherwise} \end{cases} \tag{10}$$

So the conditional probability density function of p_i with given q_i is

$$\pi(p_i|q_i) = \frac{P(q_i|p_i)}{\int_0^1 P(q_i|p_i)\mathrm{d}p_i} \tag{11}$$



(a) $n = 10^3, c = 4$ (b) $n = 10^3, c = 5$

(c) $q_i = 0.2, c = 4$ (d) $n = 10^3, q_i = 0.2$

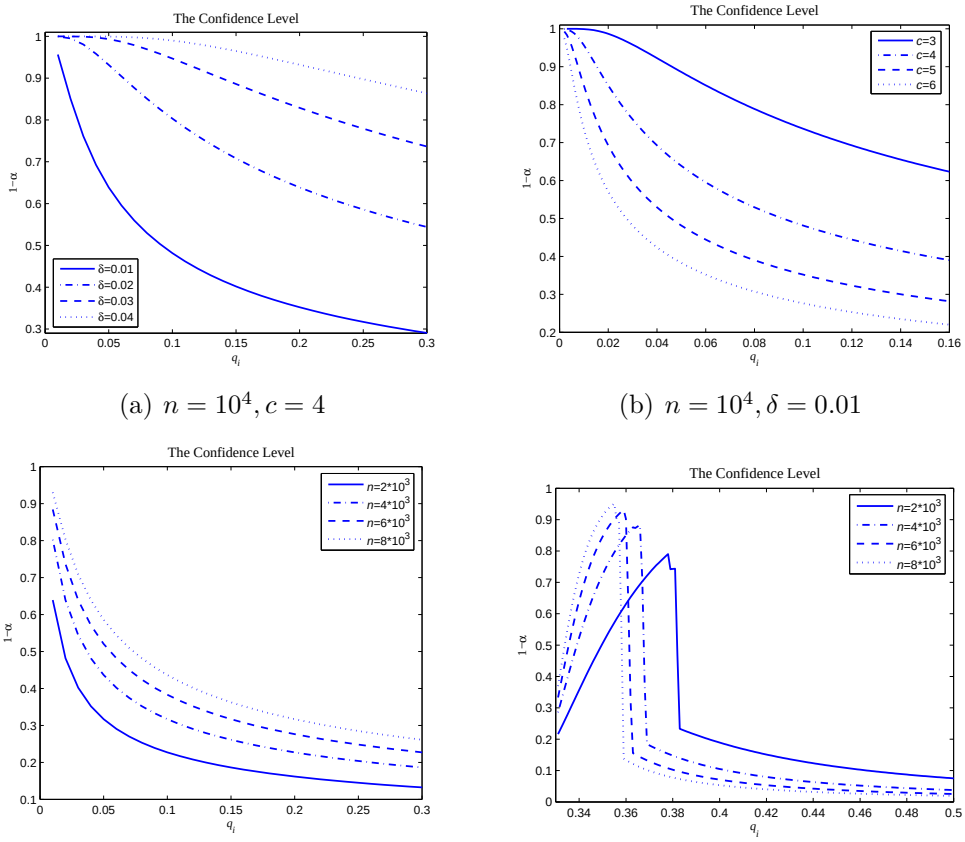
Figure 4: The Function Curve of $\pi(p_i|q_i)$ with different values of q_i, c, or n

Combing Formula (7) and Formula (11), Formula (8) can be get and Theorem 4.2 is valid. □

Figure 4 illustrates the function curve of $\pi(p_i|q_i)$ for different values of q_i, n or c. Figure 4(a) and Figure 4(b) show less q_i makes p_i centred around $1-(c-1)q_i$ more closely, Figure 4(c) show greater n makes that, and Figure 4(d) shows less c makes that, too. In addition, Figure 4(a) and Figure 4(b) also show greater q_i may lead $1-(c-1)q_i < 0$, and the corresponding p_i is 0 with a great probability.

4.3. The Confidence Level

In this subsection, an approximation method is used for calculating confidence level of reconstructed positive survey.



(a) $n = 10^4, c = 4$ (b) $n = 10^4, \delta = 0.01$

(c) $c = 4, \delta = 0.01$ (d) $c = 4, \delta = 0.01$

Figure 5: The confidence level of estimated p_i varying with q_i, c, or n

Theorem 4.3. *If confidence interval length is δ, the confidence level $1-\alpha$ is*

$$1-\alpha = \begin{cases} P(\hat{p}_i - \frac{\delta}{2} \le p_i \le \hat{p}_i + \frac{\delta}{2}) = \int_{\hat{p}_i - \frac{\delta}{2}}^{\hat{p}_i + \frac{\delta}{2}} \pi(p_i|q_i)\mathrm{d}p_i & q_i < \frac{1-\delta/2}{c-1} \\ P(0 \le p_i \le \delta) = \int_0^{\delta} \pi(p_i|q_i)\mathrm{d}p_i & q_i \ge \frac{1-\delta/2}{c-1} \end{cases} \tag{12}$$

where $\hat{p}_i = 1-(c-1)q_i$, and $\pi(p_i|q_i)$ is in Formula (8).

Proof. According to Theorem 4.2, Theorem 4.3 is valid obviously. □

Figure 5 illustrates the confidence level varying with q_i with different values of n and c. From this figure, obviously the confidence level has the

Table 1: The Confidence Level as $q_i < \frac{1-\delta/2}{c-1}$ $(\delta = 0.1,\ c = 3)$

q_i	$\hat{p}_i$	Confidence Interval $(\hat{p}_i - \delta/2, \hat{p}_i + \delta/2)$	Confidence Level: $1-\alpha$			
			n=100	n=400	n=800	n=1000
0.05	0.9	(0.85, 0.95)	0.8675	0.994	0.9998	1
0.1	0.8	(0.75, 0.85)	0.7354	0.9701	0.9973	0.9991
0.15	0.7	(0.65, 0.75)	0.639	0.9299	0.9887	0.9952
0.2	0.6	(0.55, 0.65)	0.5711	0.8851	0.9735	0.9867
0.25	0.5	(0.45, 0.55)	0.5208	0.8422	0.9537	0.9739
0.3	0.4	(0.35, 0.45)	0.4816	0.803	0.9316	0.9582
0.35	0.3	(0.25, 0.35)	0.4508	0.7679	0.9086	0.9408
0.4	0.2	(0.15, 0.25)	0.4352	0.7364	0.8859	0.9226
0.45	0.1	(0.05, 0.15)	0.4851	0.7257	0.8659	0.9049

following two characters: (1) when $q_i < (1-\delta/2)/(c-1)$, the confidence level increases with n (Figure 5(c)), and decreases with q_i (Figure 5(a)) or c (Figure 5(b)). (2) when $q_i \geq (1-\delta/2)/(c-1)$, the confidence level increases with q_i firstly (Figure 5(d)). Because in this case, the p_i is 0 with a high probability, the confidence level decreases severely (Figure 5(d)). These values of q_i are nearly impossible because the prior probability to attain such a large value of q_i is very low, and q_i may be the survey error (if $q_i > \mu + 3\sigma$ as described in subsection 4.1).

5. Simulation Experiments

In this section, some examples of negative survey (similar with that in [12]) are specially designed to verify this approximation method. In Table 1 and Table 2, the confidence level is calculated independently by category when the confidence interval (abbreviated as CI) length is 0.1. As is indicated in Table 1, the confidence interval is $(\hat{p}_i - \delta/2, \hat{p}_i + \delta/2)$ as $q_i < \frac{1-\delta/2}{c-1} = 0.475$. In this case, the confidence level increases with n, and decreases with q_i. As shown in Table 2, the confidence level is diverse and complicated. If $\hat{p}_i < 0$, then the confidence level is very small as n is large. That means an excessive rise of q_i may even be a survey error because the prior probability to attain such a greater value of q_i is very low. In addition, when $\hat{p}_i$ is a negative value,

Table 2: The Confidence Level as $q_i \geq \frac{1-\delta/2}{c-1}$ $(\delta = 0.1,\ c = 3)$

q_i	$\hat{p}_i$	Confidence Interval $(0, \delta)$	Confidence Level: $1-\alpha$			
			n=100	n=400	n=800	n=1000
0.5	0	(0, 0.1)	0.7198	0.9665	0.9973	0.9992
0.55	-0.1	(0, 0.1)	0.8949	0.9995	1	1
0.6	-0.2	(0, 0.1)	0.9674	0.9933	*0.2612*	*0.2115*
0.7	-0.4	(0, 0.1)	0.9881	*0.2433*	*0.1336*	*0.1171*
0.8	-0.6	(0, 0.1)	*0.5796*	*0.1568*	*0.1064*	*0.1022*

Table 3: The Confidence Level as $\delta = 0.1$ and $c = 3$

(q_1, q_2, q_3)	$(\hat{p_1}, \hat{p_2}, \hat{p_3})$	Confidence Level: $1-\alpha$	
		$n = 100$	$n = 1000$
(0.35,0.35,0.3)	(0.3,0.3,0.4)	(0.451,0.451,0.482)	(0.941,0.941,0.958)
(0.4,0.3,0.3)	(0.2,0.4,0.4)	(0.435,0.482,0.482)	(0.923,0.958,0.958)
(0.45,0.3,0.25)	(0.1,0.4,0.5)	(0.485,0.482,0.521)	(0.905, 0.958, 0.974)
(0.5,0.3,0.2)	(0,0.4,0.6)	(0.720,0.482,0.571)	(0.999, 0.958, 0.987)
(0.6,0.3,0.1)	(*-0.2,0.4,0.8*)	(0.967,0.482,0.735)	(*0.212*,0.958,0.999)
(0.7,0.2,0.1)	(*-0.4,0.6,0.8*)	(0.988,0.571,0.735)	(*0.117*,0.987,0.999)
(0.8,0.15,0.05)	(*-0.6,0.7,0.9*)	(0.580,0.639,0.868)	(*0.102*,0.995,1)

the second method in [11] is needed to correct the reconstructed positive survey.

Table 3 shows the confidence levels of seven groups of negative survey. The confidence level includes three values, which is the confidence level of each category respectively. It is worth reminding that the confidence levels in the last three groups of negative survey are less when $n = 1000$. The reason that the probability to get such a large value of q_i is rather low if $n = 1000$. When $n = 1000$, the confidence levels of the last three groups of negative survey are low, and the survey results may be faulty.

6. Discussion

In this study, we propose an efficient approximation method for calculating the confidence level of negative survey, but there are some work for future study.

Firstly, this approximation method is based on central limit theorem, which is valid when n is sufficiently large. However, the degree of "sufficiently large" (of n) is diverse when p_i has various values. So the "sufficiently large" cannot only be measured in n, and should be measured in both p_i and n. If np_i or $n(1-p_i)$ is smaller in amount, the Poisson Distribution is the better approximation distribution rather than Normal Distribution. In addition, Normal Distribution, which is a symmetric distribution, is used to approximate the original distribution, but the original distribution is not perfectly symmetrical.

Secondly, this method in this study analyses each category independently. The correlation of different categories should be taken into account in future work. For example, the confidence level may be high when $q_i \geq 1/(c-1)$. Because the corresponding p_i has a high probability to be 0. But in this case, the sum of all the estimated p_i is greater than 1, and the results is needed for further revision.

Thirdly, the confidence interval is set to be $(\mu - \delta/2, \mu + \delta/2)$ when $q_i < 1/(c-1)$. But strictly, $\pi(p_i|q_i)$ is not a completely symmetrical function. So the confidence interval may not be the smallest one.

Finally, the confidence level calculated in this study is by category independently. How to compare the two close confidence levels (such as the first two examples in Table 3) still needs to be studied further.

7. Conclusions

This study proposes an efficient approximation method for calculating the confidence level of negative survey. Normal Distribtuion is used to approximate to the distribution of q_i at first, then Bayes method is used for approximated calculating the confidence level. Depending on the proposed efficient approximation method, the confidence level of negative survey can be approximated calculated efficiently.

Acknowledgment

The project was supported by the Fundamental Research Funds for the Central Universities, China University of Geosciences (Wuhan) under Grant CUGL 140840, and the National Natural Science Foundation of China under Grant 61272469. The authors declare that there is no conflict of interests regarding the publication of this manuscript.

References

[1] S. A. Hofmeyr, S. Forrest, Architecture for an artificial immune system, Evolutionary Computation 8 (4) (2000) 443–473.

[2] S. Forrest, A. S. Perelson, L. Allen, R. Cherukuri, Self-nonself discrimination in a computer, in: the IEEE Symposium on Research in Security and Privacy, 1994, pp. 202–212.

[3] J. Kim, P. J. Bentley, Towards an artificial immune system for network intrusion detection: an investigation of clonal selection with a negative selection operator, in: the 2011 Congress on Evolutionary Computation(CEC'01), Vol. 2, 2001, pp. 1244–1252.

[4] G. Du, T. Huang, B. Zhao, L. Song, Dynamic self-defined immunity model base on data mining for network intrusion detection, in: the 4th International Coference on Machine Learning and Cybernetics, Vol. 6, 2005, pp. 3866–3870.

[5] Z. Ji, D. Dasgupta, V-detector: An efficient negative selection algorithm with "probably adequate" detector coverage, Information Sciences 179 (10) (2009) 1390–1406.

[6] F. Esponda, Negative surveys, Arxiv: math/0608176.

[7] J. Horey, M. Groat, S. Forrest, F. Esponda, Anonymous data collection in sensor networks, in: the Fourth Annual International Conference on Mobile and Ubiquitous Systems: Computing, Networking and Services, 2007, pp. 1–8.

[8] F. Esponda, V. M., G. c, Surveys with negative questions for sensitive items, Statistics and Probability Letters 79 (2009) 2456 – 2461.

[9] J. L. Horey, S. Forrest, M. Groat, Reconstructing spatial distributions from anonymized locations, in: 28th International Conference on Data Engineering Workshops, 2012, pp. 243 – 250.

[10] H. Xie, L. Kulik, E. Tanin, Privacy-aware collection of aggregate spatial data, Data & Knowledge Engineering 70 (6) (2011) 576 – 595.

[11] Y. Bao, W. Luo, X. Zhang, Estimating positive surveys from negative surveys, Statistics and Probability Letters 83 (2) (2013) 551 – 558.

[12] Y. Bao, W. Luo, Y. Lu, On the dependable level of the negative survey, Statistics and Probability Letters 89 (2014) 31 – 40.