



UWL REPOSITORY

repository.uwl.ac.uk

An efficient approximation method for calculating confidence level of negative survey

Liu, Ran, Peng, Jinghui and Tang, Shanyu ORCID logo ORCID: <https://orcid.org/0000-0002-2447-8135> (2015) An efficient approximation method for calculating confidence level of negative survey. *Mathematical Problems in Engineering*, 2015. p. 958154. ISSN 1024-123X

<http://dx.doi.org/10.1155/2015/958154>

This is the Draft Version of the final output.

UWL repository link: <https://repository.uwl.ac.uk/id/eprint/3947/>

Alternative formats: If you require this document in an alternative format, please contact: open.research@uwl.ac.uk

Copyright: Creative Commons: Attribution 4.0

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy: If you believe that this document breaches copyright, please contact us at open.research@uwl.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.

Rights Retention Statement:

Research Article

An Efficient Approximation Method for Calculating Confidence Level of Negative Survey

Ran Liu, Jinghui Peng, and Shanyu Tang

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

Correspondence should be addressed to Shanyu Tang; shanyu.tang@gmail.com

Received 27 April 2015; Revised 23 July 2015; Accepted 2 September 2015

Academic Editor: Ofer Hadar

Copyright © 2015 Ran Liu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The confidence level of negative survey is one of the key scientific problems. The 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 of 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.

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].

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 from. In contrast to traditional surveys, the participants are required to select a category that does *not* agree with the fact [6, 8]; that is, 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 it defines *negative category* as the other $c - 1$ categories that do *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 and tablets [9].

The positive survey can be reconstructed from a result of negative survey. For a survey consisting of a question and c ($c \geq 3$) categories for n interviewees to select from, a negative survey result is $R = (r_1, r_2, \dots, r_c)$, where r_i is the results of category i in negative survey. Meanwhile, the original positive survey is $T = (t_1, t_2, \dots, t_c)$, 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 :

$$V = \begin{bmatrix} 0 & v_{1,2} & \cdots & v_{1,c} \\ v_{2,1} & 0 & \cdots & v_{2,c} \\ \vdots & \vdots & \ddots & \vdots \\ v_{c,1} & v_{c,2} & \cdots & 0 \end{bmatrix}. \quad (1)$$

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 to the reconstruction of positive survey with negative values. Based on the problem, two methods [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 of low efficiency and complex and could not 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 reinforces 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 the following list:

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 ($1 \leq i \leq c$); that is, $q_i = r_i/n$.

t_i : the original number of interviewees in positive category i .

\hat{t}_i : the estimated number of t_i .

R : the participant vector; that is, $R = (r_1, r_2, \dots, r_c)$.

T : the participant vector; that is, $T = (t_1, t_2, \dots, t_c)$.

p_i : the proportion of positive category i ($1 \leq i \leq c$); that is, $p_i = t_i/n$.

\hat{p}_i : the estimated number of p_i ; that is, $\hat{p}_i = \hat{t}_i/n$.

Define n as the number of interviewees participating in the negative survey and c as the number of categories. The results of the negative survey are $R = (r_1, r_2, \dots, r_c)$, 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 $T = (t_1, t_2, \dots, t_c)$, 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 which is p_i , is written as $PS(n, c, p_i)$ for simplicity. And the corresponding negative category is written as $NS(n, c, q_i)$:

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

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 are in the reconstructed positive survey; that is, $\hat{p}_i > 0$ ($1 \leq i \leq c$). But this method only uses an implicit function to reconstruct the positive survey approximately. And the accuracy of this method lacks 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 efficient 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

Efficiency 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 has nonexact 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 nonexact values reconstructed from negative survey. More importantly, it is so complicated to exactly calculate the confidence level that a greedy algorithm uses [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.

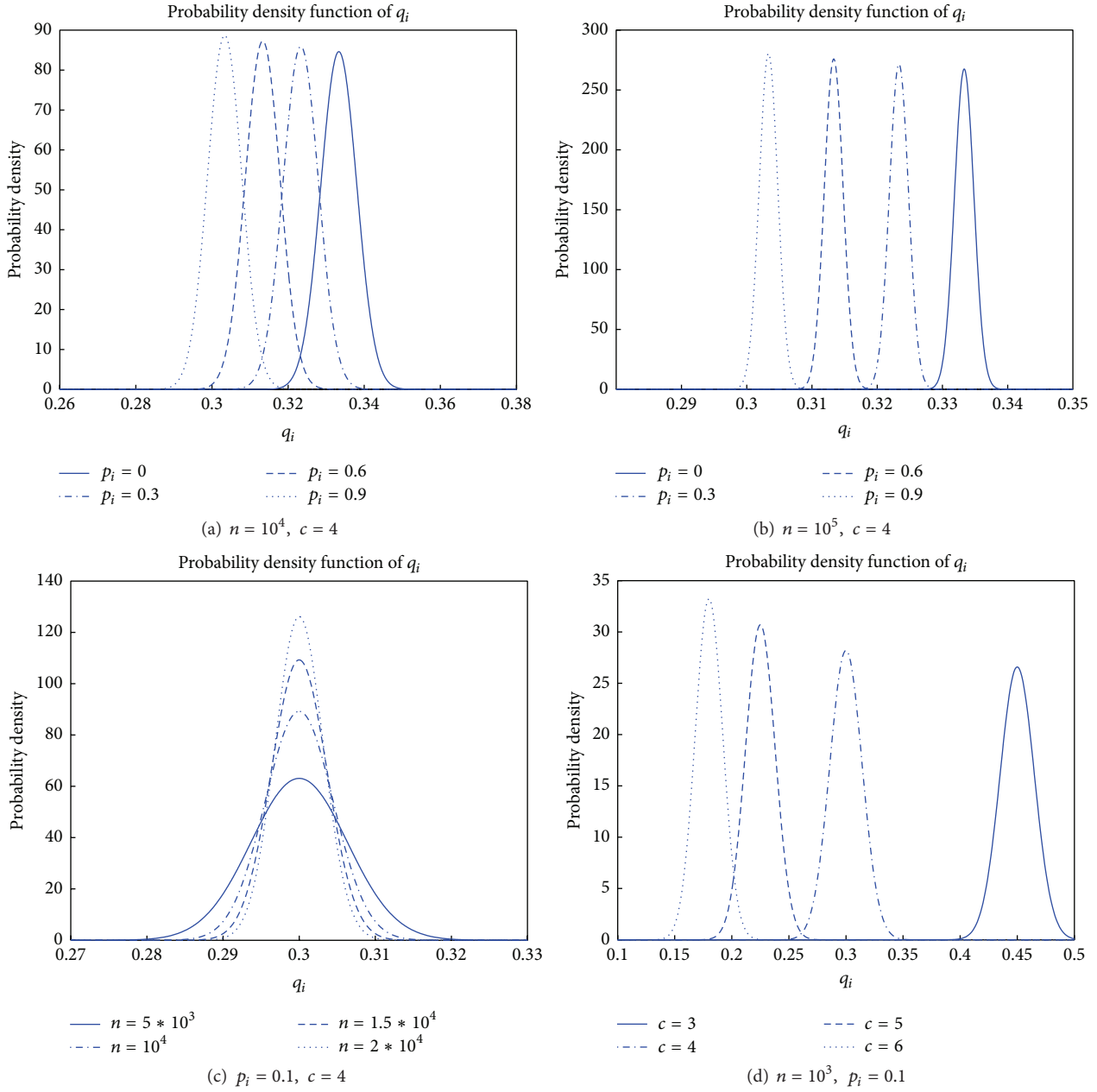


FIGURE 1: The function curve of $P(q_i | p_i)$ with different values of $p_i, c,$ or n .

4. The Efficient Method of Approximation

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

4.1. The Distribution of Negative Survey. Theorem 1 gives the distribution of category i in negative survey when that of positive survey is known.

Theorem 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:

$$\lim_{n \rightarrow \infty} P\left(\frac{q_i - \mu}{\sigma^2} \leq x\right) = \int_{-\infty}^x \frac{1}{2\pi} e^{-t^2/2} dt, \quad (3)$$

where $\mu = (1 - p_i)/(c - 1)$ and $\sigma^2 = (c - 2)(1 - p_i)/n(c - 1)^2$.

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 ($j = 1, 2, \dots, n - t_i$). If the j th

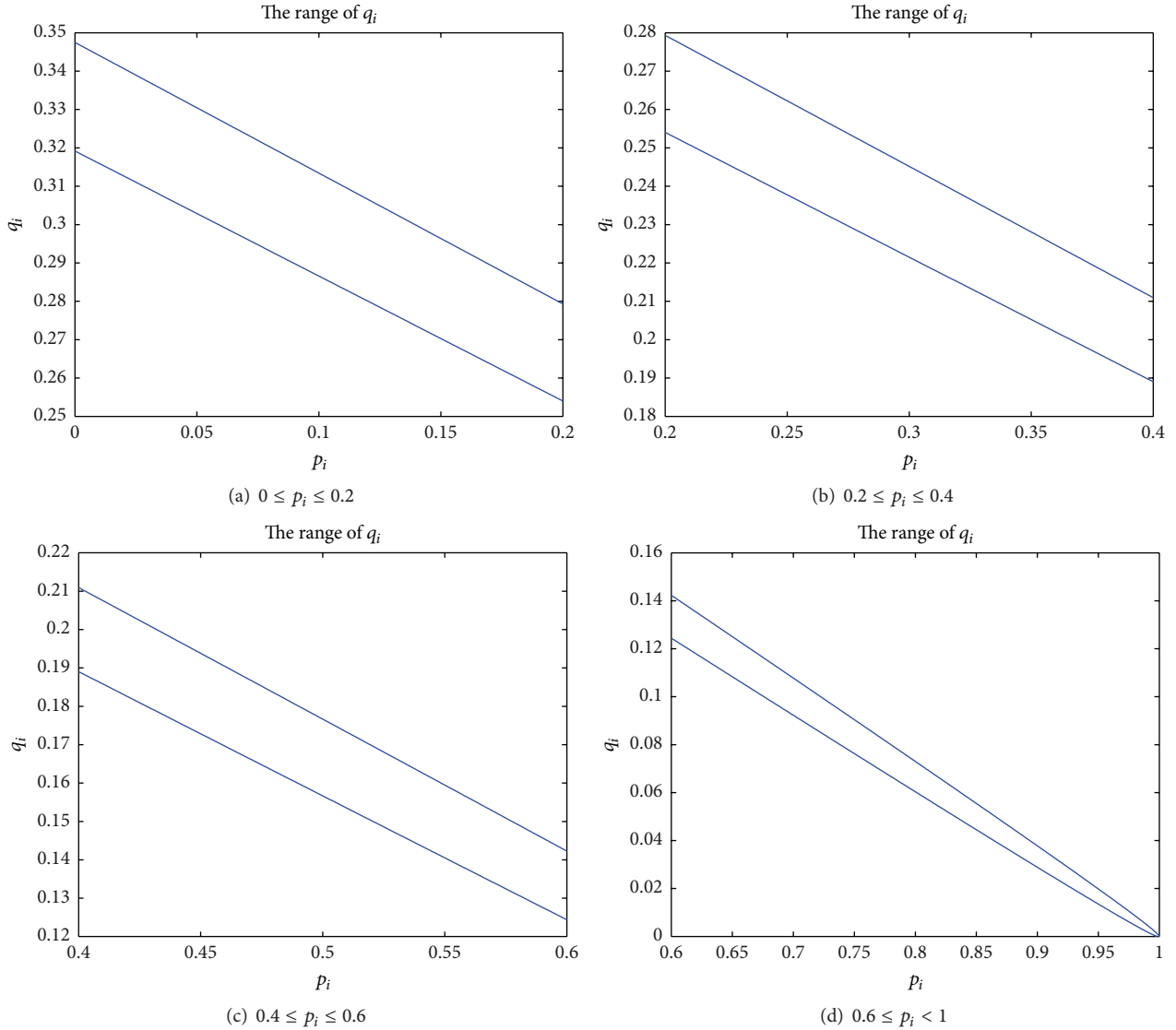


FIGURE 2: The range of q_i with probability 0.9974 ($n = 10^4, c = 4$).

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

$$\begin{aligned} E(r_i) &= \frac{n(1 - p_i)}{c - 1}, \\ D(r_i) &= \frac{n(c - 2)(1 - p_i)}{(c - 1)^2}. \end{aligned} \quad (4)$$

Owing to the De Moivre-Laplace central limit theorem, r_i follows Normal Distribution as n goes to infinity; that is,

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

So

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

In consequence, q_i follows the Normal Distribution when n goes to infinity and Theorem 1 and Formula (3) are both valid. \square

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

$$\begin{aligned} P(q_i | p_i) &= \frac{(c - 1) \sqrt{n} \exp\left(-n[q_i(c - 1) - (1 - p_i)]^2 / 2(c - 2)(1 - p_i)\right)}{\sqrt{2\pi(c - 2)(1 - p_i)}}. \end{aligned} \quad (7)$$

Figure 1 illustrates the function curve 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 2 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 Theorem 2.

Theorem 2. *If a negative category is $NS(n, c, q_i)$, the probability density function of corresponding $PS(n, c, p_i)$ is*

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

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:

$$\pi(p_i | q_i) = \frac{P(q_i | p_i) \pi(p_i)}{\int_0^1 P(q_i | p) \pi(p) dp} \quad (9)$$

Suppose that we have no knowledge of p_i . Based on Bayesian assumption, the prior probability density function

$$1 - \alpha = \begin{cases} P\left(\widehat{p}_i - \frac{\delta}{2} \leq p_i \leq \widehat{p}_i + \frac{\delta}{2}\right) = \int_{\widehat{p}_i - \delta/2}^{\widehat{p}_i + \delta/2} \pi(p_i | q_i) dp_i, & q_i < \frac{1 - \delta/2}{c - 1}, \\ P(0 \leq p_i \leq \delta) = \int_0^\delta \pi(p_i | q_i) dp_i, & q_i \geq \frac{1 - \delta/2}{c - 1}, \end{cases} \quad (12)$$

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

Proof. According to Theorem 2, Theorem 3 is valid obviously. \square

Figure 4 illustrates the confidence level varying with q_i with different values of n and c . From this figure, obviously the confidence level has the following two characters: (1) when $q_i < (1 - \delta/2)/(c - 1)$, the confidence level increases with n (Figure 4(c)) and decreases with q_i (Figure 4(a)) or c (Figure 4(b)); (2) when $q_i \geq (1 - \delta/2)/(c - 1)$, the confidence level increases with q_i firstly (Figure 4(d)). Because in this case the p_i is 0 with a high probability, the confidence level decreases severely (Figure 4(d)). These values of q_i are nearly impossible because the prior probability to attain such a large

$\pi(p_i)$ 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). Consider the following:

$$\begin{aligned} \pi(p_i) &= 1, & 0 < p_i < 1, \\ \pi(p_i) &= 0, & \text{otherwise.} \end{aligned} \quad (10)$$

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

$$\pi(p_i | q_i) = \frac{P(q_i | p_i)}{\int_0^1 P(q_i | p_i) dp_i} \quad (11)$$

Combing Formula (7) and Formula (11), Formula (8) can be gotten and Theorem 2 is valid. \square

Figure 3 illustrates the function curve of $\pi(p_i | q_i)$ for different values of q_i, n , or c . Figures 3(a) and 3(b) show that less q_i makes p_i centred around $1 - (c - 1)q_i$ more closely, Figure 3(c) shows that greater n makes that, and Figure 3(d) shows that less c makes that, too. In addition, Figures 3(a) and 3(b) also show that greater q_i may lead to $1 - (c - 1)q_i < 0$, 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.

Theorem 3. *If confidence interval length is δ , the confidence level $1 - \alpha$ is*

value of q_i is very low, and q_i may be the survey error (if $q_i > \mu + 3\sigma$ as described in Section 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 Tables 1 and 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 $(\widehat{p}_i - \delta/2, \widehat{p}_i + \delta/2)$ as $q_i < (1 - \delta/2)/(c - 1) = 0.475$. 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 $\widehat{p}_i < 0$, 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

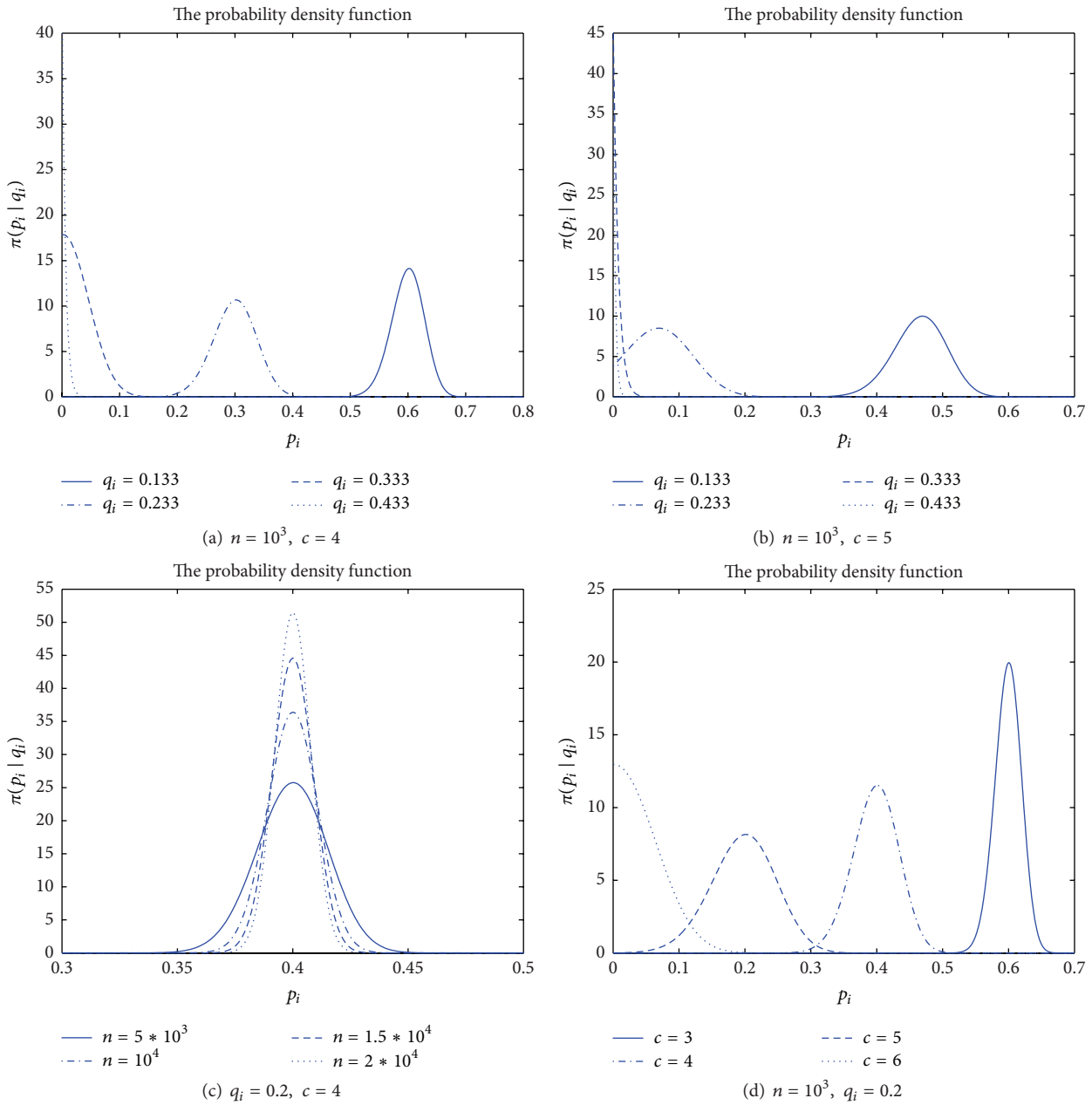


FIGURE 3: The function curve of $\pi(p_i | q_i)$ with different values of $q_i, c,$ or n .

prior probability to attain such a greater value of q_i is very low. In addition, when \hat{p}_i is a negative value, 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 works 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

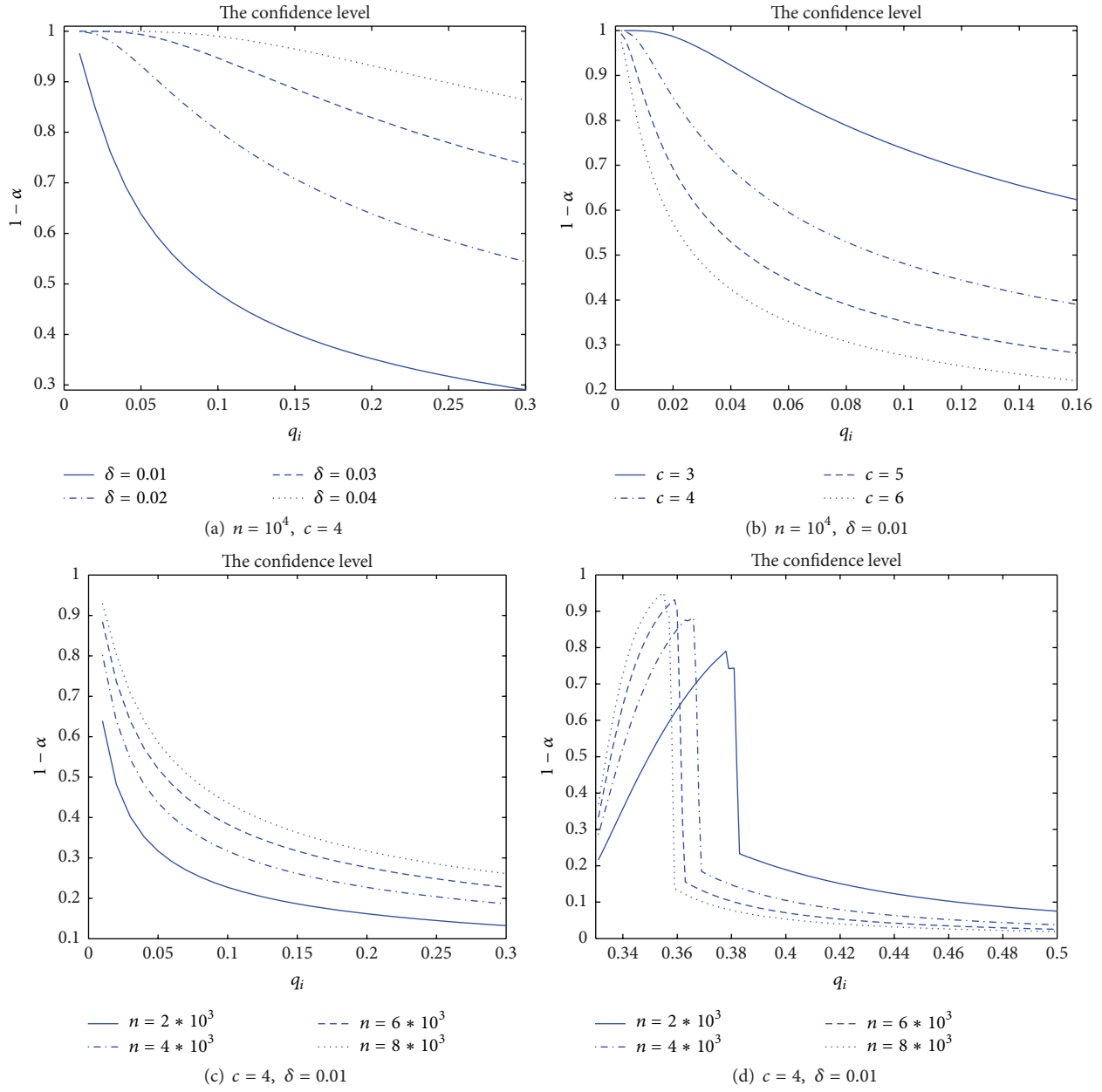


FIGURE 4: The confidence level of estimated p_i varying with q_i , c , or n .

TABLE 1: The confidence level as $q_i < (1 - \delta/2)/(c - 1)$ ($\delta = 0.1, c = 3$).

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

TABLE 2: The confidence level as $q_i \geq (1 - \delta/2)/(c - 1)$ ($\delta = 0.1, c = 3$).

q_i	\widehat{P}_i	Confidence interval (0, δ)	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)	$(\widehat{P}_1, \widehat{P}_2, \widehat{P}_3)$	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)

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 are 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 Distribution is used to approximate to the distribution of q_i at first; then Bayes method is used for approximately calculating the confidence level. Depending on the proposed efficient approximation method, the confidence level of negative survey can be approximately calculated efficiently.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

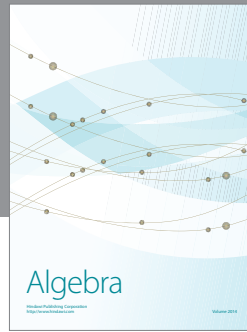
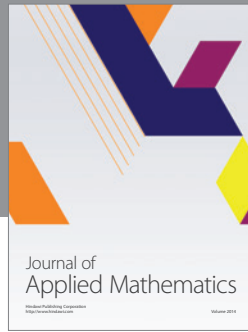
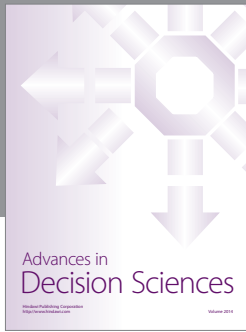
Acknowledgments

The project was supported by the National Natural Science Foundation of China under Grants 61502440, 61502439, and 61272469 and the Fundamental Research Funds for the Central Universities, China University of Geosciences (Wuhan), under Grant CUGL 140840.

References

- [1] S. A. Hofmeyr and S. Forrest, "Architecture for an artificial immune system," *Evolutionary Computation*, vol. 8, no. 4, pp. 443–473, 2000.
- [2] S. Forrest, A. S. Perelson, L. Allen, and R. Cherukuri, "Self-nonsel discrimination in a computer," in *Proceedings of the IEEE Symposium on Research in Security and Privacy*, pp. 202–212, May 1994.
- [3] J. Kim and P. J. Bentley, "Towards an artificial immune system for network intrusion detection: an investigation of clonal selection with a negative selection operator," in *Proceedings of the Congress on Evolutionary Computation (CEC '01)*, vol. 2, pp. 1244–1252, May 2001.
- [4] G. Du, T. Huang, B. Zhao, and L. Song, "Dynamic self-defined immunity model base on data mining for network intrusion detection," in *Proceedings of the 4th International Conference on Machine Learning and Cybernetics*, vol. 6, pp. 3866–3870, 2005.
- [5] Z. Ji and D. Dasgupta, "V-detector: an efficient negative selection algorithm with 'probably adequate' detector coverage," *Information Sciences*, vol. 179, no. 10, pp. 1390–1406, 2009.
- [6] F. Esponda, "Negative surveys," <http://arxiv.org/abs/math/0608176>.
- [7] J. Horey, M. M. Groat, S. Forrest, and F. Esponda, "Anonymous data collection in sensor networks," in *Proceedings of the 4th Annual International Conference on Mobile and Ubiquitous Systems: Computing, Networking and Services (MobiQuitous '07)*, pp. 1–8, August 2007.

- [8] F. Esponda and V. M. Guerrero, "Surveys with negative questions for sensitive items," *Statistics and Probability Letters*, vol. 79, no. 24, pp. 2456–2461, 2009.
- [9] J. L. Horey, S. Forrest, and M. Groat, "Reconstructing spatial distributions from anonymized locations," in *Proceedings of the 28th International Conference on Data Engineering Workshops*, pp. 243–250, ACM, 2012.
- [10] H. Xie, L. Kulik, and E. Tanin, "Privacy-aware collection of aggregate spatial data," *Data & Knowledge Engineering*, vol. 70, no. 6, pp. 575–595, 2011.
- [11] Y. Bao, W. Luo, and X. Zhang, "Estimating positive surveys from negative surveys," *Statistics and Probability Letters*, vol. 83, no. 2, pp. 551–558, 2013.
- [12] Y. Bao, W. Luo, and Y. Lu, "On the dependable level of the negative survey," *Statistics and Probability Letters*, vol. 89, pp. 31–40, 2014.



Hindawi

Submit your manuscripts at
<http://www.hindawi.com>

