STATISTICS IN TRANSITION new series, March 2025
Vol. 26, No. 1, pp. xx–xxx, https://doi.org/10.59139/stattrans-2025-xxx
Received – xx.xx.xxxx; accepted – xx.xx.xxxx

Optimal sample size in a triangular model for sensitive questions

Stanisław Jaworski[1]

Abstract

The estimation of the fraction of a population with a stigmatizing characteristic is the issue that this study attempts to address. In this paper the nonrandomized response model proposed by Tian et al. (2007) is considered. The exact confidence interval (CI) for this fraction is constructed. The optimal sample size for obtaining the CI of a given length is also derived. In order to estimate the proportion of the population with a stigmatizing characteristic, we explore the nonrandomized response model proposed by Tian et al. (2007). The prevalent approach to constructing a CI involves applying the Central Limit Theorem. Unfortunately, such CIs fail to consistently maintain the prescribed confidence level, contradicting the Neyman (1934) definition of CIs. In this paper, we present the construction of an exact CIs for this proportion, ensuring adherence to the designated confidence level. The length of the proposed CI depends on both the given probability of a positive response to a neutral question and the sample size. For these CIs, the probability of a positive response to a neutral question is established in relation to the provided limit on the privacy protection of the interviewee. Additionally, we derive the optimal sample size for obtaining a CI of a given length.

Key words: sensitive questions, nonrandomized response model, exact confidence interval.

1. Introduction

In surveys aiming to estimate the proportion of individuals with a stigmatizing characteristic, respondents often hesitate to provide truthful responses when directly questioned. To address this challenge, various methods of indirect questioning have been developed to safeguard privacy and encourage the disclosure of sensitive information. The initial approach to obscuring answers to sensitive questions was proposed by Warner (1965). This method involves the randomization of responses, with the interviewee determining the randomized answer, and the interviewer remaining unaware of the actual response to the sensitive question. Over time, Warner's model has been extended in different ways by researchers such as Horvitz et al. (1967), Greenberg et al. (1969), Raghavarao (1978), Franklin (1989), and Kuk (1990). Collectively, Warner's model and its extensions fall under the category of randomized response techniques, which necessitate the use of a randomization device.

Tian et al. (2007) and Yu et al. (2008) introduced two innovative techniques for addressing sensitive questions in population surveys: the triangular model and the crosswise model.

[1]The Institute of Economics and Finance, Warsaw University of Life Sciences, Nowoursynowska 159, 02-787 Warsaw, Poland. E-mail: stanislaw_jaworski@sggw.edu.pl. ORCID: https://orcid.org/0000-0002-6169-2886.

Both models involve asking two questions simultaneously – one sensitive and one neutral. A key advantage of these methods is that they do not require a randomization device, unlike earlier approaches. The triangular and crosswise models, along with the parallel model introduced by Tian (2014), belong to the same class of non-randomized models. The issue of determining the optimal sample size for these models has been examined by Liu and Tian (2014) and Yu et al. (2008).

An essential aspect of the sample survey design is determining the number of respondents. Tian et al. (2011) explored sample size determination for the non-randomized triangular model when dealing with sensitive questions in surveys. Their approach involved precision and power analyses for one-sided and two-sided tests, examining the hypothesis $H_0: \pi = \pi_0$, where π represents the population proportion with the sensitive characteristic, and π_0 is a pre-specified reference value. The sample size determination was guided by controlling the type I and II error rates of the tests. However, the resulting solution depends on both the pre-specified reference value π_0 and the true unknown value of π, making it challenging to apply directly in practical situations.

Qiu et al. (2014) also examined sample size determination for the triangular model, deriving formulas for estimating the parameter π. Unlike Tian et al. (2011), they explicitly incorporated an assurance probability of achieving the pre-specified precision into the formulas. However, these formulas still depend on the unknown value of π and are based on asymptotic confidence intervals, which do not maintain the nominal confidence level.

In this study, we present an alternative approach to determining the optimal sample size for the non-randomized triangular model. This approach was originally introduced by Jaworski and Zieliński (2023) for the non-randomized crosswise model. Their method simultaneously considers both the confidence interval length and the protection of respondent privacy.

In Section 2, we revisit the construction of asymptotic confidence intervals for π and elucidate the process of constructing an exact confidence interval for this parameter. Section 3 introduces the methodology for sample size selection, taking into account the privacy of the interviewee. Section 4 delves into various aspects of the numerical determination of the optimal sample size. Concluding remarks are provided in Section 5.

2. Confidence interval in Triangular Model

Let Y be a binary variable, where $\{Y = 1\}$ indicates the occurrence of drawing a person with a stigmatizing trait, and $\{Y = 0\}$ is the complement to $\{Y = 1\}$. Our focus is on estimating the proportion (denoted by $\pi = P\{Y = 1\}$) of individuals with the stigmatizing trait and constructing a confidence interval for π. The challenge we encounter is that the random variable Y is not reliably observable. Therefore, we observe another variable Z, contingent on respondents' answers to two questions. The relationship between Z and the two questions is specified by the assumed model.

In the triangular model, respondents are simultaneously presented with two independent questions–one neutral and one sensitive. They are instructed to report 0 only if the answers to both questions are not positive (*NO*). Thus, the observable variable in this model is denoted as Z, where

$$Z = \begin{cases} 0, & \text{if both answers are } NO, \\ 1, & \text{otherwise.} \end{cases} \tag{1}$$

In the triangular model the probability q of answering *YES* to neutral question is assumed to be known. Therefore

$$Z = \begin{cases} 0, & \text{with probability } (1-\pi)(1-q), \\ 1, & \text{with probability } \pi + (1-\pi)q. \end{cases} \tag{2}$$

Let us denote the probability $\pi + (1-\pi)q$ by ρ. Hence, in the triangular model

$$\pi = \frac{\rho - q}{1-q}. \tag{3}$$

Let $Z_1, Z_2, \ldots, Z_n$ be a sample. Maximum likelihood estimator (MLE) of ρ is $\hat{\rho} = \frac{1}{n}\sum_{i=1}^{n} Z_i$. Therefore, $\hat{\pi}_q = \dfrac{\hat{\rho} - q}{1-q}$ is a natural estimator of π. However, the MLE of π has the form

$$\hat{\pi} = \max\left\{0, \pi_q\right\}. \tag{4}$$

Yu et al. (2008) proved that the estimators $\hat{\pi}$ and $\hat{\pi}_q$ are asymptotically equivalent. When $n \to \infty$, the central limit theorem implies that $\hat{\pi}_q$ is asymptotically normal. Hence, the following $\delta 100\%$ Wald confidence interval of π can be constructed:

$$\hat{\pi}_q \pm z_{\frac{1-\delta}{2}} \sqrt{v(\hat{\pi}_q)} \tag{5}$$

where z_ν denotes the upper $\nu-$th quantile of the standard normal variable and $v(\hat{\pi}_q) = \hat{\rho}(1-\hat{\rho})/[(n-1)(1-q)^2]$.

It is also possible to construct $\delta 100\%$ Wilson (score) confidence interval of π:

$$\left\{ \pi \in \langle 0, 1 \rangle : \; (\hat{\pi}_q - \pi)^2 \leq \frac{z^2_{\frac{1-\delta}{2}} Var(\hat{\rho})}{(1-q)^2} \right\}, \tag{6}$$

where $Var(\hat{\rho}) = \big(\pi + (1-\pi)q\big)(1-\pi)(1-q)/n$.

The Wald and Wilson confidence intervals are known to deviate from the prescribed confidence level, making them imprecise. In contrast, the Clopper-Pearson method (Clopper and Pearson (1934)) can be employed to construct an exact confidence interval for π. Notably, since π is a linear and increasing function of ρ, the resulting exact confidence interval for π is

$$(\pi_L(\hat{\pi}), \pi_R(\hat{\pi})) = \left(\max\left\{0, \frac{\rho_L(\hat{\rho}) - q}{1-q}\right\}, \frac{\rho_R(\hat{\rho}) - q}{1-q} \right) \tag{7}$$

where $(\rho_L(\hat{\rho}), \rho_R(\hat{\rho}))$ is a Clopper-Pearson exact confidence interval of ρ, that is

$$\rho_L(\hat{\rho}) = \begin{cases} 0 & \text{dla } \hat{\rho} = 0, \\ B^{-1}\left(n - n\hat{\rho} + 1, n\hat{\rho}; \frac{1+\delta}{2}\right) & \text{dla } \hat{\rho} > 0, \end{cases} \tag{8}$$

$$\rho_U(\hat{\rho}) = \begin{cases} 1 & \text{for } \hat{\rho} = 1, \\ B^{-1}\left(n - n\hat{\rho}, n\hat{\rho} + 1; \frac{1-\delta}{2}\right) & \text{for } \hat{\rho} < 1, \end{cases} \tag{9}$$

where $B^{-1}(a,b;\cdot)$ denotes the inverse of CDF of the Beta distribution with parameters (a,b). Note, that it is enough to use the $B^{-1}(\cdot,\cdot;\cdot)$ function for setting the exact confidence interval.

3. Optimal sample size

Let us consider the length $l(\hat{\pi}; q, n)$ of the exact confidence interval. For the $n\hat{\rho}$ observed *YES* answers to the questionnaire we have

$$l(\hat{\pi}; q, n) = \pi_R(\hat{\pi}) - \pi_L(\hat{\pi}), \quad \text{where } \hat{\pi} = \max\left\{0, \frac{\hat{\rho} - q}{1 - q}\right\}. \tag{10}$$

The length of the confidence interval is a random variable concerning $\hat{\pi}$, contingent on q and n. We explore two approaches to minimize the length of the CI:

1. **Minimizing expected length**: Find minimal sample size n such that the expected length of the confidence interval does not exceed a predetermined value.

2. **Almost sure minimizing**: Find minimal sample size n such that there is a high probability that the length of the confidence interval does not exceed a predetermined value.

The solution of these approaches is influenced by the probability of a positive answer to the neutral question. Thus, a rational criterion for the optimal selection of this probability needs to be formulated. Denoting the optimally selected q, dependent on the sample size n, as $q_e(n)$ and $q_d(n)$ in the first and second approaches, respectively. Let Π and Q represent acceptable sets for π and q, respectively. In the absence of prior knowledge about π and reasonable restrictions for q, the sets are $\Pi = (0,1)$ and $Q = \langle 0, 1)$.

Optimal q in the first approach. Let $\mathscr{X}$ denote the sample space of $\hat{\pi}$. The problem may be written in the following way:

$$q_e(n) = \arg\min_{q \in Q} \sup_{\pi \in \Pi} E_{\pi}^{C(\pi)} l(\hat{\pi}; q, n), \tag{11}$$

where $E_{\pi}^{C(\pi)} l(\hat{\pi}; q, n) = \sum_{x \in C(\pi)} l(x; q, n) P_{\pi}\{\hat{\pi} = x\}$ represents the expected length of the CI covering estimated value of π. Here, the set $C(\pi) = \{x \in \mathscr{X} : \pi_L(x) < \pi < \pi_R(x)\}$ comprises the values of the variable $\hat{\pi}$ for which the CI covers π.

Optimal q in the second approach. The problem may be written in the following way:

$$q_d(n) = \arg\max_{q \in Q} \inf_{\pi \in \Pi} P_{\pi}^{C(\pi)} \{l(\hat{\pi}; q, n) \leq d\}, \tag{12}$$

where $\delta \cdot P_{\pi}^{C(\pi)} \{l(\hat{\pi}, q, n) \leq d\} = \sum_{x \in C(\pi)} P_{\pi}\{\hat{\pi} = x\} \mathbb{1}(l(x, q, n) \leq d)$ represents the probability that the length of the CI covering the estimated value of π does not exceed the given value d. The function $\mathbb{1}(p)$ is equal to one if the logical value of p is true and zero otherwise.

In the case of $Q = \langle 0, 1)$, the minimal length concerning q is achieved when $q = 0$, equivalent to not asking the neutral question. However, such a questionnaire (without a neutral question) fails to ensure the privacy of respondents. Therefore, it is reasonable to impose a constraint on the probability q, considering the desired level of protection.

Tan et al. (2009) introduced the concept of the degree of privacy protection through the probabilities

$$P_{\pi}\{Y = 1 | Z = 1\} \text{ and } P_{\pi}\{Y = 1 | Z = 0\}. \tag{13}$$

These probabilities are connected with the safety of the interviewee of non-discovering her/his positive answer to the sensitive question. These probabilities should be small enough so that they do not exceed the given value $\gamma \in (0, 1)$. The researcher can set this value according to the requirements of the conducted survey. In the triangular model, the aforementioned probabilities are as follows:

$$\begin{aligned} P_{\pi}\{Y = 1 | Z = 1\} &= \frac{\pi}{\pi + (1 - \pi) q}, \\ P_{\pi}\{Y = 1 | Z = 0\} &= 0. \end{aligned} \tag{14}$$

The relationship between the probability $P_{\pi}\{Y = 1 | Z = 1\}$ and q is illustrated in Figure 1.

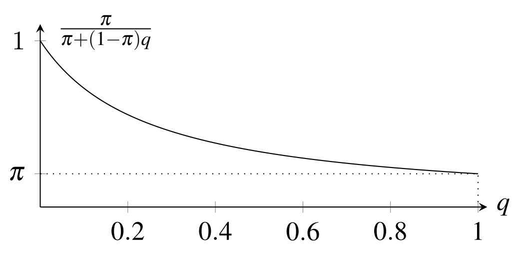

Figure 1: Privacy protection versus q.

We are interested in $q < 1$ such that

$$\frac{\pi}{\pi + (1 - \pi) q} \leq \gamma \quad \text{for } \pi \in \Pi. \tag{15}$$

Simple algebra yields the following condition for q:

$$q(\pi; \gamma) \leq q < 1 \quad \text{for } \pi \in \Pi, \tag{16}$$

where $q(\pi;\gamma) = \frac{\pi(1-\gamma)}{\gamma(1-\pi)}$ increases with respect to π. Since $q(\gamma,\gamma) = 1$, the condition (16) holds if and only if $\gamma > \pi$. This implies that the maximal privacy protection (i.e. the minimal γ to be chosen) is restricted by the percentage of the population that has committed socially stigmatizing characteristic. Consequently, the problem of minimizing the length, assuming $\pi \leq \pi_0$ for a given $\in (0,1)$, is well defined for $q \in \langle q(\pi_0;\gamma),1)$. In the following discussion, we assume that $\Pi = (0,\pi_0\rangle$ and $Q = \langle q(\pi_0;\gamma),1)$, where $\gamma > \pi_0$. The value π_0 reflects our prior knowledge about π, indicating that we know the percentage of people bearing a stigmatizing characteristic is less than π_0. The inequalities (15) and (16) lead us to the conclusion that without this knowledge, determining the appropriate value for γ is not feasible. Note that both Π and Q do not depend on the sample size n. Therefore, the length of the CI can be minimized by selecting an appropriate sample size. Let $d \in (0,1)$ be a given number. Our goal is to determine the sample size that yields a CI with a length not exceeding d. Specifically, we are interested in a CI covering the estimated value of π. We can define two approaches to address this problem.

Optimal sample size in the first approach. Identify minimal n such that

$$E_{\pi}^{C(\pi)} l(\pi, q_e(n), n) \leq d \quad \text{for all } \pi \in \Pi. \tag{17}$$

Optimal sample size in the second approach. Identify minimal n such that

$$P_{\pi}^{C(\pi)} \{l(\pi, q_d(n), n) \leq d\} \geq 1-\lambda \text{ for given } 1-\lambda \text{ and all } \pi \in \Pi. \tag{18}$$

In the first approach, our objective is to ensure that the average length of the CI covering the estimated value of π is less than a given d. In the second approach, our goal is to ensure that the length of at least $(1-\lambda)\%$ of the CIs covering the estimated π is less than the specified d. It is important to note that we have a minimum of $\delta\%$ of intervals covering the unknown parameter π, and for an infinitely large sample size n, the defined value $P_{\pi}^{C(\pi)} \{l(\hat{\pi}, q, n) \leq d\}$ is equal to one.

The approaches to determining the optimal sample size were initially introduced for the non-randomized crosswise model by Jaworski and Zieliński (2023).

4. Numerical consideration

Let us assume that $\pi < 0.5$ and the confidence level is set at $\delta = 0.95$. Moving on to the first approach, an analysis of $E_{\pi}^{C(\pi)} l(\hat{\pi}, q, n)$ reveals (refer to Figure 2) that for each $\pi < 0.5$ and n it increases with q. Consequently, it can be inferred that $q_e(n) = q(\pi_0,\gamma)$ for any $\gamma \in (0, 0.5\rangle$. In the triangular model, $q(\pi_0,\gamma)$ decreases with γ. Hence, our interest lies in identifying the smallest and acceptable value of γ. However, it is crucial to note that the measure of privacy protection revealed by the triangular model cannot be zero. Hence, opting for $\gamma = 0.5$ appears reasonable. In this scenario, the probability that the respondent belongs to a sensitive group is 50%, thereby mitigating the legal risks associated with the respondent's answers in the survey.

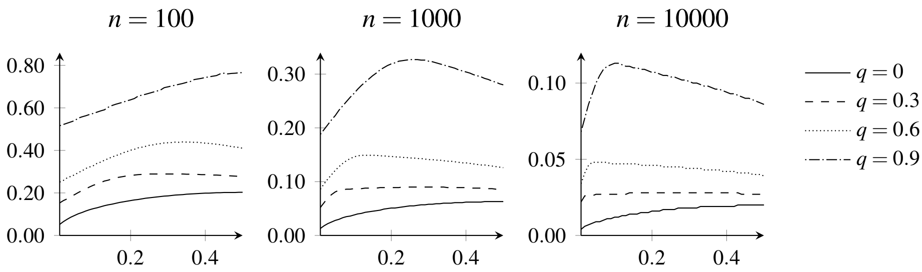

Figure 2: Expected length versus $\pi \in (0, \gamma = 0.5)$ with respect to q under the condition that π is covered by the CI.

The expected length $E_{\pi}^{C(\pi)} l(\hat{\pi}, q, n)$ is not monotonic with π for every q. Let us define

$$\pi_{max}(n; \pi_0) = \underset{\pi \in \Pi}{\operatorname{argsup}} E_{\pi}^{C(\pi)} l(\pi, q(\pi_0, 0.5), n). \tag{19}$$

It is depicted in Figure 3 that if $\pi_0 \leq 0.25$ then $\pi_{max}(n; \pi_0) = \pi_0$ otherwise it is a decreasing function of sample size n (with the accuracy implied by the discreteness of the distribution of the observed variable). This knowledge of π_{max} can be helpful in the optimal sample size numerical finding in the approach. In Table 1 some exemplary of optimal sample sizes are given for confidence level $\delta = 0.95$ and privacy protection level $\gamma = 0.5$. The optimal sample size is increasing with π_0. Larger values of π_0 correspond to higher uncertainty about parameter π. Therefore, the optimal sample sizes are smaller for smaller values of π_0.

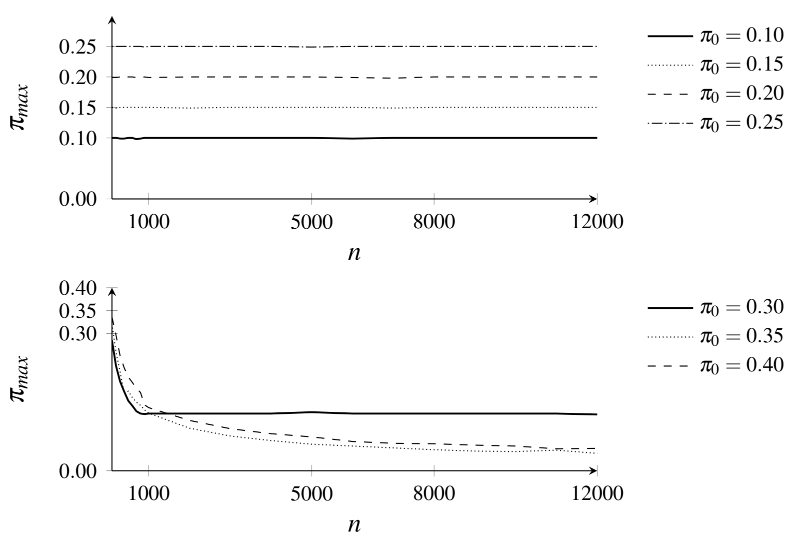

Figure 3: $\pi_{max}(n) = \operatorname{argsup}_{\pi \in \Pi} E_{\pi}^{C(\pi)} l(\pi, q(\pi_0, 0.5), n)$ versus sample size n

Consider the case for $d = 0.06$. The optimal sample size is equal to 822 when $\pi_0 = 0.1$ and is about 8 times greater for $\pi_0 = 0.4$. This means that the costs of conducting a survey are much higher for the latter case. Recall that when we conduct a survey by asking the sensitive question directly with no additional neutral question, the length of the CI for π is equal approximately to 0.06 when sample size $n = 1000$. This remark enable us to conclude that without additional knowledge about the scope of π we will incur much higher research costs with an appropriately secured level of privacy.

Table 1: The smallest n that $\sup_{\pi\in\Pi} E_\pi^{C(\pi)} l(\pi, q(\pi_0,\gamma), n) \le d$.

π_0	$q(\pi_0,\gamma)$	$d = 0.05$	$d = 0.06$
0.1	0.11	1171	822
0.2	0.25	2422	1693
0.3	0.43	4146	2893
0.4	0.67	8120	5646

Note: $q(\pi_0,\gamma) = \frac{\pi_0}{1-\pi_0}$ for $\gamma = 0.5$

Now, let us consider the second approach. The probability $P_\pi^{C(\pi)}\{l(\pi,q,n) \le d\}$ decreases with q, and this dependency is illustrated in Figures 4, 5 and 6. When comparing Figures 4 and 5, we observe that the monotonicity of $P_\pi^{C(\pi)}\{l(\pi,q,n) \le d\}$ concerning π depends on the sample size n. It is noteworthy that the lines in Figures 4, 5 and 6 exhibit some lack of smoothness due to the discreteness of the observed variable, albeit small enough to explore the optimal sample size at $q = q(\pi_0,\gamma)$. In Table 2, we provide some exemples of optimal sample sizes.

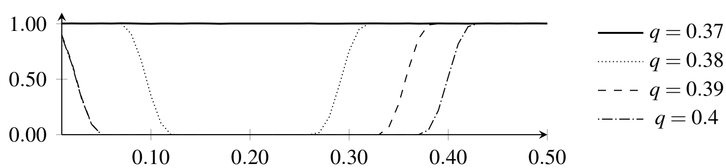

Figure 4: The probability as a function of π, with respect to q under the condition that π is covered by the CI. Here, $n = 4000$.

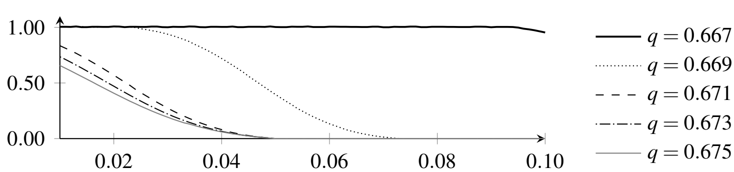

Figure 5: The probability as a function of π, with respect to q under the condition that π is covered by the CI. Here, $n = 1359$.

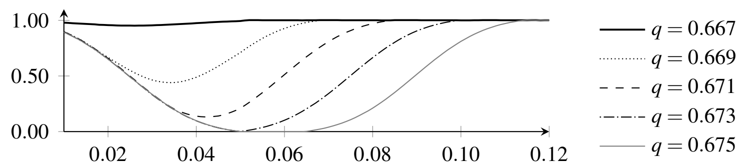

Figure 6: The probability as a function of π, with respect to q under the condition that π is covered by the CI. Here, $n = 12243$.

Please note that the optimal sample size exhibits a modest variation when λ is set to 0.01 compared to 0.05, with the maximum observed difference being 13 (refer to Table 2). However, noteworthy disparities arise in relation to π_0, contingent upon the prior knowledge of the true value of the π parameter. Increased uncertainty regarding π results in a higher optimal sample size requirement. For instance, when $\pi = 0.4$, the optimal sample size is approximately nine times greater than that for $\pi = 0.4$.

Table 2: The smallest n that $\inf_{\pi\in\Pi} P_{\pi}^{C(\pi)}\left\{l(\pi, q(\pi_0;\gamma), n) \le d\right\} \ge 1-\lambda$.

		$d = 0.05$		$d = 0.06$	
π_0	$q(\pi_0,\gamma)$	$\lambda = 0.01$	$\lambda = 0.05$	$\lambda = 0.01$	$\lambda = 0.05$
0.1	0.11	1376	1359	973	960
0.2	0.25	2708	2702	1891	1886
0.3	0.43	4774	4774	3324	3324
0.4	0.67	12250	12243	8499	8494

Note: $q(\pi_0,\gamma) = \frac{\pi_0}{1-\pi_0}$ for $\gamma = 0.5$

5. Conclusions

The paper introduces a novel CI for the fraction of sensitive questions in the triangular model. Unlike the widely used asymptotic CI, the new approach maintains the prescribed confidence level, which is consistent with Neyman's (1934) definition of CIs.

Addressing a crucial practical concern, we derived the minimum sample size satisfying two criteria: average length and almost sure length. To obtain these sample sizes, we impose restrictions on privacy protection, specifically the probability of discovering a YES answer to the sensitive question. This probability should be sufficiently small to ensure the interviewee's comfort in answering the questionnaire. Additionally, we limit our analysis to rare phenomena, focusing on sensitive questions with a small (predefined) probability of a positive answer.

It is crucial to emphasize that we refrain from comparing the length of our CI with asymptotic versions. Asymptotic CIs are inherently shorter because they lack the capability to uphold a specified confidence level, leading to a real probability of coverage that is less than the designated confidence level. Consequently, the comparison of lengths is devoid of meaningful insights. Our CI is characterized by its ease of calculation; even a standard spreadsheet application can efficiently compute the quantiles of the Beta distribution. While asymptotic CIs based on normal approximation served a purpose in times when computers were not readily available, we advocate for the practical application of our CI in contemporary scenarios.

The provided numerical examples demonstrate that incorporating prior knowledge of the true value of π enables a reduction in the minimum sample size necessary to achieve the desired estimation precision. In the absence of this knowledge, the optimal sample size may inflate by more than eight times, posing an unfavorable scenario given the associated research costs.

References

Clopper, C. J., Pearson, E. S., (1934). The Use of Confidence or Fiducial Limits Illustrated in the Case of the Binomial. *Biometrika, 26(4)*, pp. 404–413.

Franklin, L. A., (1989). Randomized response sampling from dichotomous populations with continuous randomization. *Survey Methodology*, 15, pp. 225–235.

Greenberg, B. G., Abul-Ela, A.-L. A. and Horvitz, D. G., (1969). The Unrelated Question Randomized Response Model: Theoretical Framework. *Journal of the American Statistical Association*, 64, pp. 520–539.

Groenitz, H., (2014). A new privacy-protecting survey design for multichotomous sensitive variables. *Metrika*, 77, pp. 211–224.

Horvitz, D. G., Shah, B. V., Simmons and W. R., (1967). The Unrelated Question Randomized Response Model. *in Proceedings of the Social Statistics Section. American Statistical Association*, pp. 65–72.

Jaworski, S., Zieliński, W., (2023). The Optimal Sample Size in the Crosswise Model for Sensitive Questions. *Applicationes Mathematicae*, 50(1), pp. 2–34.

Kuk, A. Y. C., (1990). Asking Sensitive Question Indirectly. *Biometrika*, 77, pp. 436–438.

Liu, Y., Tian, G.-L., (2014). Sample size determination for the parallel model in a survey with sensitive questions. *Journal of the Korean Statistical Society*, 43(2), pp. 235–249. doi: 10.1016/j.jkss.2013.08.002.

Neyman, J., (1934). On the Two Different Aspects of the Representative Method: The Method of Stratified Sampling and the Method of Purposive Selection. *Journal of the Royal Statistical Society*, 97, pp. 558–625.

Qiu, S.-F., Zou, G. Y. and Tang, M.-L., (2014). Sample size determination for estimating prevalence and a difference between two prevalences of sensitive attributes using the non-randomized triangular design. *Computational Statistics & Data Analysis*, 77, pp. 157–169. doi: 10.1016/j.csda.2014.02.019.

Raghavarao, D., (1978). On an Estimation Problem in Warner's Randomized Response Technique. *Biometrics. [Wiley, International Biometric Society]*, 34(1), pp. 87–90. Available at: http://www.jstor.org/stable/2529591 (Accessed: 1 July 2022).

Tan, M., Tian, G. L. and Tang, M. L., (2009). Sample Surveys With Sensitive Questions: A Non-Randomized Response Approach. *The American Statistician*, 63, pp. 9–16.

Tian, G. L., Yu, J. W., Tang, M. L. and Geng, Z., (2007). A New Nonrandomized Model for Analyzing Sensitive Questions with Binary Outcomes. *Statistics in Medicine*, 26, pp. 4238–4252.

Tian, G. L., Tang, M. L., Liu, Z., Tan, M. and Tang, N. S., (2011). Sample Size Determination for the Non-Randomised Triangular Model for Sensitive Questions in a Survey. *Statistical Methods in Medical Research*, 20, pp. 159–173.

Tian, G. L., (2014). A New Non-Randomized Response Model: The Parallel Model. *Statistica Neerlandica*, 68(4), pp. 293–323.

Warner, S. L., (1965). Randomized Response: A Survey Technique for Eliminating Evasive Answer Bias. *Journal of the American Statistical Association*, 60, pp. 63–69.

Yu, J. W., Tian, G. L. and Tang, M. L., (2008). Two New Models for Survey Sampling With Sensitive Characteristic: Design and Analysis. *Metrika*, 67(3), pp. 251–263.