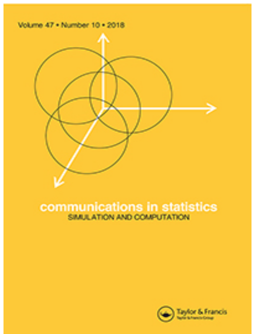


Communications in Statistics - Simulation and Computation

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/lssp20

# Logarithmic confidence intervals for the cross-product ratio of binomial proportions under different sampling schemes

Chanakan Sungboonchoo, Su-Fen Yang, Wararit Panichkitkosolkul & Andrei Volodin

**To cite this article:** Chanakan Sungboonchoo, Su-Fen Yang, Wararit Panichkitkosolkul & Andrei Volodin (2021): Logarithmic confidence intervals for the cross-product ratio of binomial proportions under different sampling schemes, Communications in Statistics - Simulation and Computation, DOI: 10.1080/03610918.2021.1914090

**To link to this article:** https://doi.org/10.1080/03610918.2021.1914090

Published online: 02 May 2021.

COMMUNICATIONS IN STATISTICS - SIMULATION AND COMPUTATION®
https://doi.org/10.1080/03610918.2021.1914090

Check for updates

# Logarithmic confidence intervals for the cross-product ratio of binomial proportions under different sampling schemes

Chanakan Sungboonchoo[a], Su-Fen Yang[b], Wararit Panichkitkosolkul[a], and Andrei Volodin[c]

[a]Department of Mathematics and Statistics, Thammasat University, Phathum Thani, Thailand; [b]Department of Statistics, National Chengchi University, Taipei City, Taiwan; [c]Department of Mathematics and Statistics, University of Regina, Regina, Canada

**ABSTRACT**

We consider the problem of logarithmic interval estimation for a cross-product ratio $\rho = \frac{p_1(1-p_2)}{p_2(1-p_1)}$ with data from two independent Bernoulli samples. Each sample may be obtained in the framework of direct or inverse Binomial sampling schemes. Asymptotic logarithmic confidence intervals are constructed under different types of sampling schemes, with parameter estimators demonstrating exponentially decreasing bias. Our goal is to investigate the cases when the relatively simple normal approximations for estimators of the cross-product ratio are reliable for constructing logarithmic confidence intervals. We use the closeness of the confidence coefficient to the nominal confidence level as our main evaluation criterion, and use the Monte-Carlo method to investigate the key probability characteristics of intervals corresponding to all possible combinations of sampling schemes. We present estimations of the coverage probability, expectation and standard deviation of interval widths in tables. Also, we provide some recommendations for applying each logarithmic interval obtained.

**ARTICLE HISTORY**
Received 10 June 2020
Accepted 4 April 2021

**KEYWORDS**
Cross-product ratio; Direct binomial sampling scheme; Inverse binomial sampling scheme; Logarithmic confidence interval; Normal approximation

## 1. Introduction

Comparing the success probabilities of Bernoulli trials is a problem that arises in biological and medical investigations. In this article, we investigate the accuracy properties for confidence estimation of the cross-product ratio of Binomial proportions under different sample schemes.

Ngamkham, Volodin, and Volodin (2016) and Ngamkham (2018) considered the problem of confidence estimation for the ratio of Binomial proportions. The cross-product ratio statistic is more frequently applied to real data, especially in medical and biological research, because the statistic is important in analyzing $2 \times 2$ contingency tables; see Lehmann (1997), Section 4.6.

For that reason, it is very interesting to investigate statistical inference for the cross-product ratio.

Goodman (1964) developed simple methods of obtaining confidence limits for the cross-product ratio in a $2 \times 2$ table, and extended these methods to obtain simultaneous confidence intervals for the $r(r-1)c(c-1)/4$ cross-product ratios in an $r \times c$ table and, likewise, for the relative differences between the corresponding cross-product ratios in $K$ different $r \times c$ tables. Goodman's methods improve on Gart's (1962) method for $2 \times 2$ tables, and can be applied to more general $r \times c$ tables. Moreover, these methods are easier to apply than those given by Cornfield (1956).

**CONTACT** Su-Fen Yang yang@mail2.nccu.tw Department of Statistics, National Chengchi University, No. 64, Sec. 2, Zhinan Rd., Wenshan Dist., Taipei City, Taiwan.

Anděl (1973) suggested a method based on logarithmic interactions for comparing the association in $k$ fourfold tables ($k$ independent samples).

Lee (1981) presented the empirical Bayes modification of the cross-product ratio for studying the trend and degree of relationship between two cross-classified factors in a $2 \times 2$ contingency table. The Independent Poisson, Product Multinomial, and Multinomial are the three sampling schemes used for determining the cell frequencies in contingency tables. These procedures were studied and compared with the classical procedures; the results indicated that the empirical Bayes estimation procedures had a lower average squared error than the classical procedures.

Albert and Gupta (1983) investigated the Bayesian approach to estimating cell probabilities for $2 \times 2$ and $I \times 2$ tables. For the $2 \times 2$ table, the prior information was declared in terms of the cross product ratio coefficient. For the $I \times 2$ table, estimators were based on a two-stage prior for the $I$ binomial probabilities, where the first stage was the conjugate beta distribution and the second the discrete uniform distribution.

Holland and Wang (1987) used the local dependence function that measures the margin-free dependence to order bivariate distributions. Wang (1987) applied the characterization of a bivariate normal distribution to generate a table of probability integrals via the iterative proportional fitting algorithm.

McCann and Tebbs (2009) constructed the simultaneous logit-based confidence intervals for odds ratios in the analysis of classification tables with a fixed reference level. They examined six procedures to control the familywise error rate, and consider the simultaneous coverage probability and mean interval width, which can be used to construct simultaneous confidence intervals.

Baxter and Marchant (2010) described how non-randomized trials can provide bias in the effectiveness of any intrusion. Their study showed a process to estimate the bias in such trials under the bivariate log-normal and gamma distributions, and the size of the bias under two different bivariate models.

Xu (2012) demonstrated that the odds ratio, or the cross-product ratio, is greater than or equal to one under the generalized proportional hazards model. The author used this property to improve a process of testing when the generalized proportional hazards model is not ideal to use for a data set.

Schaarschmidt, Gerhard, and Vogel (2017) proposed an asymptotic method for computing simultaneous confidence intervals for user-defined sets of pairwise, between-treatment comparisons and user-defined sets of odds ratios based on the assumption of several independent multinomial samples. They also give an improvement of this method by taking the correlation into account, and considered an application of Dirichlet posteriors with a vague Dirichlet prior.

Niebuhr and Trabs (2019) examined the impact of weighted data for estimating a discrete probability distribution for one-dimensional distributions. The weighting of observations usually increase estimation variances. In the two-dimensional discrete distribution, this research assumes that one marginal distribution is known. This additional information in one category of a contingency table allows for adjusting the estimation of another marginal if there is some degree of association between the two categories. For applications where the independence of the marginals is not reasonable, the authors suggested adjusted estimators.

Martín Andrés, Tapia García, and Gayá Moreno (2020) considered the two-tailed asymptotic inferences about the odds ratio in cross-sectional studies (under the multinomial sampling). The research investigated 15 different methods, 5 new and 10 classic. They compared the new methods with other procedures.

Nonparametric estimation procedures were considered in Lin and Yang (2008), Mahdizadeh and Arghami (2012), and Amini and Mahdizadeh (2017).

In our recent article, Sungboonchoo et al. (2021), we constructed linear asymptotic confidence intervals in accordance with different types of sampling schemes. Unfortunately, these linear intervals possess quite low precision and poor accuracy properties. A common practice in statistics is to take the log transformation of highly skewed data and construct confidence intervals for the population average on the basis of transformed data. In this article, we investigate logarithmic confidence intervals that show much better precision and accuracy properties. To support our theoretical findings, we apply the Monte-Carlo method to investigate the key probability characteristics of the logarithmic confidence intervals. In the Monte-Carlo simulations, we used the number of replications most commonly used in the literature $N = 10^5$.

A mathematical statement of the problem is as follows. Let $X_1, X_2, \ldots$ and $Y_1, Y_2, \ldots$ be two independent sequences of Bernoulli random variables with success probabilities $p_1$ and $p_2$, respectively. The observations are done according to the sequential sampling schemes with Markov stopping times $\nu_1$ and $\nu_2$. From the results of observations $X^{(\nu_1)} = (X_1, \ldots, X_{\nu_1})$ and $Y^{(\nu_2)} = (Y_1, \ldots, Y_{\nu_2})$, it is necessary to identify the most accurate method of interval estimating the cross-product ratio $\rho = \frac{p_1(1-p_2)}{p_2(1-p_1)}$. Each sample may be obtained in the framework of direct or inverse binomial sampling schemes.

*Direct binomial sampling*. In this scheme, a random vector $X^{(n)} = (X_1, \ldots, X_n)$ with Bernoulli components and a fixed number of observations $n$ is observed. Note that $T = \sum_{i=1}^{n} X_i$ has the Binomial distribution $B(n, p)$, which has two parameters $n$ and $p$, where $n$ is a natural number and $0 < p < 1$.

If the random variable $T$ has a Binomial distribution, then its probability mass function is

$$P\{T = t\} = \binom{n}{t} p^t (1-p)^{n-t}, t = 0, 1, \ldots, n.$$

The random variable $\overline{X}_n = \frac{T}{n}$ is asymptotically normal with a mean of $\mu_X = p$ and variance $\sigma_X^2 = p(1-p)/n$.

*Inverse binomial sampling*. In this scheme, a Bernoulli sequence $Y^{(\nu)} = (Y_1, \ldots, Y_\nu)$ is observed with a stopping time of $\nu = \min\{n : \sum_{k=1}^{n} Y_k \geq m\}$, where $m$ is a fixed number of successes. From Ngamkham (2018) and Ngamkham, Volodin, and Volodins (2016), we know that $\nu$ has the Pascal distribution $P(m, p)$, where parameter $m$ is a natural number, and $0 < p < 1$.

If the random variable $\nu$ has a Pascal distribution, then its probability mass function is

$$P\{\nu = k\} = \binom{k-1}{m-1} p^m (1-p)^{k-m}, k = m, m+1, m+2, \ldots$$

The random variable $\overline{Y}_m = \nu/m$ is asymptotically normal with a mean of $\mu_Y = 1/p$ and variance $\sigma_Y^2 = \frac{1-p}{mp^2}$.

In the following, we keep the notation $X_1, X_2, \ldots$ for a Bernoulli sequence obtained by the direct sampling scheme and $Y_1, Y_2, \ldots$ for a Bernoulli sequence obtained by the inverse sampling scheme.

## 2. Estimation of proportion $p$ and its reciprocal $p^{-1}$

First, we consider the problem of estimating parameter $p$ (success probability) and parametric function $\frac{1}{p}$ for the Bernoulli trials. It seems difficult to estimate $\frac{1}{q}$, where $q = 1 - p$, so we avoid this expression in our further derivations by expressing it in terms of $\frac{p}{q}$ and $\frac{1}{p}$; see Sec. 3. In this section, we discuss how to estimate $p$ and $\frac{1}{p}$. The following formulae are derived in Ngamkham (2018) and Ngamkham, Volodin, and Volodins (2016); we present them in Table 1.

In the case of direct binomial sampling, the estimate $\widehat{p^{-1}}_n$ is biased. Ngamkham (2018) and Ngamkham, Volodin, and Volodins (2016) proved that $\text{Bias}(\widehat{p^{-1}}_n) = \frac{1}{p} - E\widehat{p^{-1}}_n = \frac{1}{p}(1-p)^{n+1}$ is decreasing with an exponential rate as $n \to \infty$. In all other cases, $\hat{p}_n, \hat{p}_m$, and $\widehat{p^{-1}}_m$ provide unbiased estimators.

**Table 1.** Estimators for the proportion $p$ and the reciprocal $p^{-1}$ for direct and inverse sampling schemes.

| | Proportion $p$ | Reciprocal $p^{-1}$ |
|---|---|---|
| Direct Sampling Scheme | $\hat{p}_n = \overline{X}_n$ | $\widehat{p_n^{-1}} = \frac{n+1}{n\overline{X}_n+1} \approx \widetilde{p_n^{-1}} = \frac{1}{\overline{X}_n}$ |
| Inverse Sampling Scheme | $\hat{p}_m = \frac{m-1}{m\overline{Y}_m-1} \approx \widetilde{p}_m = \frac{1}{\overline{Y}_m}$ | $\widehat{p_m^{-1}} = \overline{Y}_m$ |

## 3. Estimating the parametric functions $\frac{p}{q}$ and $\frac{q}{p}$

To solve the problem stated in the Introduction, it is necessary to construct estimates, preferably with exponentially decreasing bias, for the parametric functions $\frac{p}{q}$ and $\frac{q}{p}$, where $q = 1 - p$ for two schemes of Bernoulli trials.

In terms of estimation, the simplest case is an estimation of the parametric function $\frac{q}{p}$ :

$$\frac{q}{p} = \frac{1-p}{p} = \frac{1}{p} - 1.$$

From Sec. 2, we already know how to estimate $\frac{1}{p}$ for both schemes of Bernoulli trials.

In the case of direct binomial sampling, we use statistics $\widehat{p_n^{-1}} = \frac{n+1}{n\overline{X}_n+1}$ to estimate $p^{-1}$ with an exponentially decreasing bias. In the case of inverse binomial sampling, we use statistics $\widehat{p_m^{-1}} = \overline{Y}_m = \nu/m$ as an unbiased estimator of $p^{-1}$.

We proceed with estimation of the parametric function $\frac{p}{q}$.

**Proposition 3.1.** *In the case of direct binomial sampling, the statistics*

$$\widehat{p/q}_n = \frac{n\overline{X}_n}{n+1-n\overline{X}_n}$$

*estimates the parametric functions $\frac{p}{q}$ with an exponentially decreasing bias.*

*Proof.* As we know, the statistic $T = n\overline{X}_n$ has the Binomial distribution $B(n, p)$. Therefore,

$$E\widehat{p/q}_n = E\frac{n\overline{X}_n}{n+1-n\overline{X}_n} = E\frac{T}{n+1-T} = \sum_{k=0}^{n} \frac{k}{n+1-k}\binom{n}{k} p^k q^{n-k}$$
$$= \sum_{k=1}^{n} \frac{k}{n+1-k}\frac{n!}{k!(n-k)!} p^k q^{n-k} \text{for } k=0 \text{ we have zero term}$$
$$= \sum_{k=1}^{n} \frac{n!}{(k-1)!(n+1-k)!} p^k q^{n-k}$$
$$= \sum_{j=0}^{n-1} \frac{n!}{j!(n-j)!} p^{j+1} q^{n-j-1} \text{make a substitution } j = k-1$$
$$= \frac{p}{q}\left[\sum_{j=0}^{n} \frac{n!}{j!(n-j)!} p^j q^{n-j} - \frac{n!}{n!0!} p^n q^0\right] = \frac{p}{q}\left[(p+q)^n - p^n\right] = \frac{p}{q}(1-p^n).$$

Therefore, $\text{Bias}(\widehat{p/q}_n) = \frac{p}{q} - E\widehat{p/q}_n = \frac{p^{n+1}}{q}$ is decreasing with an exponential rate as $n \to \infty$. □

**Table 2.** Estimators for the parametric functions $\frac{p}{q}$ and $\frac{q}{p}$ for direct and inverse sampling schemes.

| | Parametric function $\frac{p}{q}$ | Parametric function $\frac{q}{p}$ |
|---|---|---|
| Direct Sampling Scheme | $\widehat{p/q}_n = \frac{n\overline{X}_n}{n+1-n\overline{X}_n}$ | $\widehat{q/p}_n = \frac{n+1}{n\overline{X}_n+1} - 1$ |
| | $\approx \widetilde{p/q}_n = \frac{\overline{X}_n}{1-\overline{X}_n}$ | $\approx \widetilde{q/p}_n = \frac{1}{\overline{X}_n} - 1$ |
| | $E\widehat{p/q}_n = \frac{p}{q} - \frac{p^{n+1}}{q}$ | $E\widehat{q/p}_n = \frac{q}{p} - \frac{q^{n+1}}{p}$ |
| Inverse Sampling Scheme | $\widehat{p/q}_m = \frac{m-1}{m\overline{Y}_m-m+1}$ | $\widehat{q/p}_m = \overline{Y}_m - 1$ |
| | $\approx \widetilde{p/q}_m = \frac{1}{\overline{Y}_m-1}$ | |
| | $E\widehat{p/q}_m = \frac{p}{q} - \frac{p^m}{q}$ | $E\widehat{q/p}_m = \frac{q}{p}$ |

**Proposition 3.2.** *In the case of inverse binomial sampling, the statistics*

$$\widehat{p/q}_m = \frac{m-1}{m\overline{Y}_m - m + 1}$$

*estimates the parametric functions* $\frac{p}{q}$ *with an exponentially decreasing bias.*

*Proof.* As we know, the statistic $\nu = m\overline{Y}_m$ has the Pascal distribution $P(m, p)$. Therefore,

$$E\widehat{p/q}_m = E\frac{m-1}{m\overline{Y}_m - m+1} = E\frac{m-1}{\nu - m+1} = (m-1)\sum_{k=m}^{\infty}\frac{1}{k-m+1}\binom{k-1}{m-1}p^m q^{k-m}$$

$$= (m-1)p^m\sum_{i=0}^{\infty}\frac{1}{i+1}\binom{i+m-1}{m-1}q^i \text{make a substitution } i = k-m$$

$$= \frac{(m-1)p^m}{q}\sum_{i=0}^{\infty}\binom{i+m-1}{m-1}\frac{q^{i+1}}{i+1} = \frac{(m-1)p^m}{q}\sum_{i=0}^{\infty}\binom{i+m-1}{m-1}\int_0^q t^i dt$$

We can interchange the sum and integral signs by the Fubini-Tornelli theorem

$$= \frac{(m-1)p^m}{q}\int_0^q\sum_{i=0}^{\infty}\binom{i+m-1}{m-1}t^i dt = \frac{(m-1)p^m}{q}\int_0^q (1-t)^{-m}dt$$

because for any $|t| < 1$ and natural number $m \geq 1 : (1-t)^{-m} = \sum_{i=0}^{\infty}\binom{i+m-1}{m-1}t^i$;

see, for example Lemma 1, Ngamkham (2018)

$$= \frac{(m-1)p^m}{q}\left[\frac{p^{1-m}}{m-1} - \frac{1}{m-1}\right] = \frac{p}{q} - \frac{p^m}{q}$$

Therefore, $\text{Bias}(\widehat{p/q}_m) = \frac{p}{q} - E\widehat{p/q}_m = \frac{p^m}{q}$ is decreasing with an exponential rate as $m \to \infty$. □

To summarize, we present Table 2 for the estimation of $p/q$ and its reciprocal. In Table 2, $m$ and $n$ are fixed numbers, $\{X_1, X_2, ...\}$ and $\{Y_1, Y_2, ...\}$ are sequences of independent Bernoulli random variables with the parameter $p, T = \sum_{k=1}^n X_k, \overline{X}_n = T/n, \nu = \min\{n : \sum_{k=1}^n Y_k \geq m\}$, and $\overline{Y}_m = \nu/m$.

## 4. Point estimator for the cross-product ratio

In Table 3, we present all estimators of the cross-product ratio of two proportions $\rho = \frac{p_1(1-p_2)}{p_2(1-p_1)} = \frac{p_1}{q_1} \times \frac{q_2}{p_2}$ for these four possible sampling schemes.

**Table 3.** Estimators of the cross-product ratio of two proportions $\rho = \frac{p_1(1-p_2)}{p_2(1-p_1)}$ and their approximations for all possible combinations of direct and inverse sampling schemes.

| Sampling Schemes | Second Sample Direct | Second Sample Inverse |
|---|---|---|
| First Sample | $\hat{\rho}_{n_1,n_2} = \frac{n_1\overline{X}_{n_1}}{n_1+1-n_1\overline{X}_{n_1}}\left(\frac{n_2+1}{n_2\overline{X}_{n_2}+1}-1\right)$ | $\hat{\rho}_{n,m} = \frac{n\overline{X}_n}{n+1-n\overline{X}_n}(\overline{Y}_m-1)$ |
| Direct | $\approx \widetilde{\rho}_{n_1,n_2} = \frac{\overline{X}_{n_1}}{1-\overline{X}_{n_1}}\left(\frac{1}{\overline{X}_{n_2}}-1\right)$ | $\approx \widetilde{\rho}_{n,m} = \frac{\overline{X}_n(\overline{Y}_m-1)}{1-\overline{X}_n}$ |
| First Sample | $\hat{\rho}_{m,n} = \left(\frac{m-1}{m\overline{Y}_m-m+1}\right)\left(\frac{n+1}{n\overline{X}_n+1}-1\right)$ | $\hat{\rho}_{m_1,m_2} = \left(\frac{m_1-1}{m_1\overline{Y}_{m_1}-m_1+1}\right)(\overline{Y}_{m_2}-1)$ |
| Inverse | $\approx \widetilde{\rho}_{m,n} = \frac{1}{\overline{Y}_m-1}\left(\frac{1}{\overline{X}_n}-1\right)$ | $\approx \widetilde{\rho}_{m_1,m_2} = \frac{\overline{Y}_{m_2}-1}{\overline{Y}_{m_1}-1}$ |

Now we find the asymptotic of the mean and variance of logarithms of these estimates using the standard Delta method.

## 5. Delta-method

Let $g(v_1, v_2)$ be a differentiable scalar function of two variables. Consider an estimator $T = g(V_1, V_2)$, which is a function of two other basic statistics $V_1$ and $V_2$. Usually, statistics $V_1$ and $V_2$ have a simple form and are jointly asymptotically normal. The asymptotic distribution of an estimator $T$ is found with the help of Delta-method, which is a procedure of stochastic representation of $T$ with the accuracy $\mathcal{O}_P(1/\sqrt{n})$, where $n$ is the sample size.

By the Delta-method, we expand function $g$ into a Taylor series at the point $\mu_1 = \mathrm{E}V_1$ and $\mu_2 = \mathrm{E}V_2$ :

$$g(V_1, V_2) = g(\mu_1, \mu_2) + \sum_{i=1}^{2} \frac{\partial g(\mu_1, \mu_2)}{\partial v_i}(V_i - \mu_i) + \text{Remainder}.$$

It is possible to prove that the remainder term of the expansion converges in probability to zero with the rate $\mathcal{O}([\min\{m, n\}]^{-1/2})$ as sample sizes $m$ and $n$ tend to infinity. We have that $g(V_1, V_2) - g(\mu_1, \mu_2)$ is asymptotically normal with a mean of zero and variance

$$E\left[\sum_{i=1}^{2} \frac{\partial g(\mu_1, \mu_2)}{\partial v_i}(V_i - \mu_i)\right]^2.$$

Therefore, the test statistics $T$ is asymptotically normal with a mean of $g(\mu_1, \mu_2)$ and the variance of the form that is expressed through the elements of the covariance matrix of basic statistics $V_1$, $V_2$ and the coefficients $\frac{\partial g(\mu_1, \mu_2)}{\partial v_i}$.

For large values of $m$ and $n$, logarithms of all four estimators of the cross-product ratio $\rho$ are differentiable functions of statistics $\overline{X}_n$ and $\overline{Y}_m$ with finite second moments; therefore, the estimates are asymptotically normal. Our immediate task is to find the asymptotic of the mean and variance of these logarithmic estimates, for which we explore the standard Delta method described above. In our case, the method is based on a Taylor series expansion in the neighborhoods of the mean values of the statistics $\overline{X}_n$ and $\overline{Y}_m$. It is possible to calculate variances in all four cases because statistics $\overline{X}_n$ and $\overline{Y}_m$ are independent.

## 6. Asymptotic distribution of logarithms of estimates for the cross-product ratio

We consider the following four possible scenarios:

1. Direct-direct. Fix two natural numbers $n_1$ and $n_2$. Let $X^{(n_1)} = (X_{11}, ..., X_{1n_1})$ and $X^{(n_2)} = (X_{21}, ..., X_{2n_2})$ be two independent sequences of Bernoulli random variables. We know that

the sample means for both samples $V_1 = \overline{X}_{n_1}$ and $V_2 = \overline{X}_{n_2}$ are asymptotically normal and jointly approximately normal, because the samples are independent. The form of the function $g(V_1, V_2)$ is presented in Sec. 6.1. The accuracy of the Delta method, in this case, is $\mathcal{O}_P\left(1/\sqrt{\min\{n_1, n_2\}}\right)$.

2. Direct-inverse. Fix two natural numbers $n$ and $m$. Let $X_1, ..., X_n$ and $Y_1, ..., Y_\nu$ be two independent sequences of Bernoulli random variables, where $\nu = \min\{n : \sum_{k=1}^n Y_k \geq m\}$. We know that the sample mean for the first samples $V_1 = \overline{X}_n$ and statistic $V_2 = \overline{Y}_m = \nu/m$ are asymptotically normal and jointly approximately normal, because the samples are independent. The form of the function $g(V_1, V_2)$ is presented in Sec. 6.2. The accuracy of the Delta method, in this case, is $\mathcal{O}_P\left(1/\sqrt{\min\{n, m\}}\right)$.
3. Inverse-direct. Fix two natural numbers $n$ and $m$. Let $Y_1, ..., Y_\nu$ and $X_1, ..., X_n$ be two independent sequences of Bernoulli random variables, where $\nu = \min\{n : \sum_{k=1}^n Y_k \geq m\}$. We know that the statistic $V_1 = \overline{Y}_m = \nu/m$ and the sample mean for the second samples $V_2 = \overline{X}_n$ are asymptotically normal and jointly approximately normal, because the samples are independent. The form of the function $g(V_1, V_2)$ is presented in Sec. 6.3. The accuracy of the Delta method, in this case, is $\mathcal{O}_P\left(1/\sqrt{\min\{n, m\}}\right)$.
4. Inverse-inverse. Fix two natural numbers $m_1$ and $m_2$. Let $Y_{11}, ..., Y_{1\nu_1}$ and $Y_{21}, ..., Y_{2\nu_2}$ be two independent sequences of Bernoulli random variables, where $\nu_1 = \min\{n : \sum_{k=1}^n Y_{1k} \geq m_1\}$ and $\nu_2 = \min\{n : \sum_{k=1}^n Y_{2k} \geq m_2\}$. We know that the statistics $V_1 = \overline{Y}_{1m} = \nu_1/m_1$ and $V_2 = \overline{Y}_{2m} = \nu_2/m_2$ are asymptotically normal and jointly approximately normal, because the samples are independent. The form of the function $g(V_1, V_2)$ is presented in Sec. 6.4. The accuracy of the Delta method, in this case, is $\mathcal{O}_P\left(1/\sqrt{\min\{m_1, m_2\}}\right)$.

In the following, we see that the normal approximation to estimators $\log(\widetilde{\rho})$ for all sampling schemes have the same structure of means and variances:

$$\text{Asymptotic Mean of } \log(\widetilde{\rho}) = \log(\rho) \text{ and Asymptotic Variance of} \log(\widetilde{\rho}) = s^2(p_1, p_2).$$

Remember that (see, for example Propositions 3 and 4, Ngamkham (2018)):
statistic $\overline{X}_n$ has a mean $p$ and variance $\frac{pq}{n}$, and is asymptotically normal with these parameters,
statistic $\overline{Y}_m$ has a mean $1/p$ and variance $\frac{q}{mp^2}$, and is asymptotically normal with these parameters.

If we use formulae for $\hat{\rho}$, then our expressions for asymptotic variance are cumbersome. Hence, we use the approximate estimators $\widetilde{\rho}$ in Delta method derivations.

### *6.1. Direct-direct sampling scheme*

From Table 3, the statistic of interest is

$$\begin{aligned}\log(\widetilde{\rho}_{n_1, n_2}) &= \log\left(\frac{\overline{X}_{n_1}}{1 - \overline{X}_{n_1}}\left(\frac{1}{\overline{X}_{n_2}} - 1\right)\right) = g_{dd}(V_1, V_2) \\ &= \log(V_1) - \log(1 - V_1) + \log(1 - V_2) - \log(V_2),\end{aligned}$$

where $V_1 = \overline{X}_{n_1}$ and $V_2 = \overline{X}_{n_2}$. In this particular case, the function $g_{dd}(v_1, v_2) = \log(v_1) - \log(1 - v_1) + \log(1 - v_2) - \log(v_2)$.

Note that $EV_i = p_i, \text{Var}V_i = p_i q_i/n_i, i = 1, 2$ and $g_{dd}(p_1, p_2) = \log(p_1) - \log(1 - p_1) + \log(1 - p_2) - \log(p_2) = \log\rho$.

Partial derivatives are:

$$\frac{\partial g_{dd}(v_1,v_2)}{\partial v_1}=\frac{1}{v_1}+\frac{1}{1-v_1}\quad\text{and}\quad\frac{\partial g_{dd}(v_1,v_2)}{\partial v_2}=-\frac{1}{v_2}-\frac{1}{1-v_2};$$

thus,

$$\frac{\partial g_{dd}(p_1,p_2)}{\partial v_1}=\frac{1}{p_1}+\frac{1}{q_1}=\frac{1}{p_1q_1}\quad\text{and}\quad\frac{\partial g_{dd}(p_1,p_2)}{\partial v_2}=-\frac{1}{p_2}-\frac{1}{q_2}=-\frac{1}{p_2q_2}.$$

Linear term Taylor expansion in the neighborhoods of the mean values of the statistics takes the form:

$$\log\widetilde{\rho}_{n_1,n_2}=g_{dd}(V_1,V_2)\approx\log\rho+\frac{1}{p_1q_1}(\overline{X}_{n_1}-p_1)-\frac{1}{p_2q_2}(\overline{X}_{n_2}-p_2).$$

From this, the estimator $\log(\widetilde{\rho}_{n_1,n_2})$ is approximately normal with

$$\text{Mean}=\log(\rho)$$

and (remember that $\overline{X}_{n_1}$ and $\overline{X}_{n_2}$ are independent)

$$\text{Variance}=s^2=\frac{1}{p_1^2q_1^2}\frac{p_1q_1}{n_1}+\frac{1}{p_2^2q_2^2}\frac{p_2q_2}{n_2}=\frac{p_1}{q_1}(p_1^{-1})^2/n_1+\frac{p_2}{q_2}(p_2^{-1})^2/n_2.$$

To obtain the plug-in estimator of the variance, we substitute estimations for $\frac{p}{q}$ and $p^{-1}$ (see Tables 1 and 2), namely $\widetilde{p_1/q_1}=\frac{\overline{X}_{n_1}}{1-\overline{X}_{n_1}}, \widetilde{p_2/q_2}=\frac{\overline{X}_{n_2}}{1-\overline{X}_{n_2}}, \widetilde{p_1^{-1}}=\frac{1}{\overline{X}_{n_1}}$, and $\widetilde{p_2^{-1}}=\frac{1}{\overline{X}_{n_2}}$, and obtain that

$$\begin{aligned}\hat{s}^2&=\frac{\overline{X}_{n_1}}{1-\overline{X}_{n_1}}\left(\frac{1}{\overline{X}_{n_1}}\right)^2/n_1+\frac{\overline{X}_{n_2}}{1-\overline{X}_{n_2}}\left(\frac{1}{\overline{X}_{n_2}}\right)^2/n_2\\&=\frac{1}{n_1\overline{X}_{n_1}(1-\overline{X}_{n_1})}+\frac{1}{n_2\overline{X}_{n_2}(1-\overline{X}_{n_2})}.\end{aligned}$$

### 6.2. Direct-inverse sampling scheme

From Table 3, the statistic of interest is

$$\log(\widetilde{\rho_{n,m}})=\log\left(\frac{\overline{X}_n(\overline{Y}_m-1)}{1-\overline{X}_n}\right)=g_{di}(V_1,V_2)=\log(V_1)-\log(1-V_1)+\log(V_2-1),$$

where $V_1=\overline{X}_n$ and $V_2=\overline{Y}_m$. In this particular case, the function $g_{di}(v_1,v_2)=\log(v_1)-\log(1-v_1)+\log(v_2-1)$.

Note that $EV_1=p_1$, $EV_2=\frac{1}{p_2}$, $\text{Var}V_1=p_1q_1/n$, $\text{Var}V_2=\frac{q_2}{mp_2^2}$ and $g_{di}\left(p_1,\frac{1}{p_2}\right)=\log(p_1)-\log(1-p_1)+\log\left(\frac{1}{p_2}-1\right)=\log(\rho)$.

Partial derivatives are:

$$\frac{\partial g_{di}(v_1,v_2)}{\partial v_1}=\frac{1}{v_1}+\frac{1}{1-v_1}\quad\text{and}\quad\frac{\partial g_{di}(v_1,v_2)}{\partial v_2}=\frac{1}{v_2-1};$$

thus,

$$\frac{\partial g_{di}\left(p_1,\frac{1}{p_2}\right)}{\partial v_1}=\frac{1}{p_1}+\frac{1}{q_1}=\frac{1}{p_1q_1}\quad\text{and}\quad\frac{\partial g_{di}\left(p_1,\frac{1}{p_2}\right)}{\partial v_2}=\frac{p_2}{q_2}.$$

Linear term Taylor expansion in the neighborhoods of the mean values of the statistics takes the form:

$$\log \widetilde{\rho}_{n,m} = g_{di}(V_1, V_2) \approx \log(\rho) + \frac{1}{p_1 q_1}(\overline{X}_n - p_1) + \frac{p_2}{q_2}\left(\overline{Y}_m - \frac{1}{p_2}\right).$$

From this, the estimator $\log(\widetilde{\rho}_{n,m})$ is approximately normal with

$$\text{Mean} = \log(\rho)$$

and (remember that $\overline{X}_n$ and $\overline{Y}_m$ are independent)

$$\text{Variance} = s^2 = \frac{1}{p_1^2 q_1^2}\frac{p_1 q_1}{n} + \frac{p_2^2}{q_2^2}\frac{q_2}{m p_2^2} = \frac{p_1}{q_1}(p_1^{-1})^2/n + \left(\frac{p_2}{q_2}\right)p_2^{-1}/m.$$

To obtain the plug-in estimator of the variance, we substitute estimations for $\frac{p}{q}$ and $p^{-1}$ (see Tables 1 and 2), namely $\widetilde{p_1/q_1} = \frac{\overline{X}_n}{1-\overline{X}_n}, \widetilde{p_2/q_2} = \frac{1}{\overline{Y}_m - 1}, \widetilde{p_1^{-1}} = \frac{1}{\overline{X}_n}$ and $\widetilde{p_2^{-1}} = \overline{Y}_m$, and obtain that

$$\begin{aligned}\hat{s}^2 &= \frac{\overline{X}_n}{1-\overline{X}_n}\left(\frac{1}{\overline{X}_n}\right)^2/n + \frac{1}{\overline{Y}_m - 1}\overline{Y}_m/m \\ &= \frac{1}{n\overline{X}_n(1-\overline{X}_n)} + \frac{\overline{Y}_m}{m(\overline{Y}_m - 1)}.\end{aligned}$$

### *6.3. Inverse-direct sampling scheme*

From Table 3, the statistic of interest is

$$\log(\widetilde{\rho_{m,n}}) = \log\left(\frac{1}{\overline{Y}_m - 1}\left(\frac{1}{\overline{X}_n} - 1\right)\right) = g_{id}(V_1, V_2) = -\log(V_1 - 1) + \log(1 - V_2) - \log(V_2),$$

where $V_1 = \overline{Y}_m$ and $V_2 = \overline{X}_n$. In this particular case, the function $g_{id}(v_1, v_2) = -\log(v_1 - 1) + \log(1 - v_2) - \log(v_2)$.

Note that $EV_1 = \frac{1}{p_1}$, $EV_2 = p_2$, $\text{Var}V_1 = \frac{q_1}{m p_1^2}$, $\text{Var}V_2 = p_2 q_2/n$ and $g_{id}\left(\frac{1}{p_1}, p_2\right) = -\log\left(\frac{1}{p_1} - 1\right) + \log(1 - p_2) - \log(p_2) = \log(\rho)$.

Partial derivatives are:

$$\frac{\partial g_{id}(v_1, v_2)}{\partial v_1} = -\frac{1}{v_1 - 1} \quad \text{and} \quad \frac{\partial g_{id}(v_1, v_2)}{\partial v_2} = -\frac{1}{1 - v_2} - \frac{1}{v_2};$$

thus,

$$\frac{\partial g_{id}\left(\frac{1}{p_1}, p_2\right)}{\partial v_1} = -\frac{1}{\frac{1}{p_1} - 1} = -\frac{p_1}{q_1} \quad \text{and} \quad \frac{\partial g_{id}\left(\frac{1}{p_1}, p_2\right)}{\partial v_2} = -\frac{1}{q_2} - \frac{1}{p_2} = -\frac{1}{p_2 q_2}.$$

Linear term Taylor expansion in the neighborhoods of the mean values of the statistics takes the form:

$$\log \widetilde{\rho}_{m,n} = g_{id}(V_1, V_2) \approx \log(\rho) - \frac{p_1}{q_1}\left(\overline{Y}_m - \frac{1}{p_1}\right) - \frac{1}{p_2 q_2}(\overline{X}_n - p_2).$$

From this, the estimator $\log(\widetilde{\rho}_{m,n})$ is approximately normal with

$$\text{Mean} = \log(\rho)$$

and (remember that $\overline{Y}_m$ and $\overline{X}_n$ are independent)

$$\text{Variance} = s^2 = \left(\frac{p_1}{q_1}\right)^2 \frac{q_1}{mp_1^2} + \left(\frac{1}{p_2 q_2}\right)^2 \frac{p_2 q_2}{n}$$
$$= \left(\frac{p_1}{q_1}\right) p_1^{-1}/m + \frac{p_2}{q_2}\left(p_2^{-1}\right)^2/n.$$

To obtain the plug-in estimator of the variance, we substitute estimations for $\frac{p}{q}$ and $p^{-1}$ (see Tables 1 and 2), namely $\widetilde{p_1/q_1} = \frac{1}{\overline{Y}_m - 1}, \widetilde{p_1^{-1}} = \overline{Y}_m, \widetilde{p_2/q_2} = \frac{\overline{X}_n}{1 - \overline{X}_n}$, and $\widetilde{p_2^{-1}} = \frac{1}{\overline{X}_n}$, and obtain that

$$\hat{s}^2 = \frac{1}{(\overline{Y}_m - 1)} \overline{Y}_m/m + \frac{\overline{X}_n}{1 - \overline{X}_n}\left(\frac{1}{\overline{X}_n}\right)^2/n$$
$$= \frac{\overline{Y}_m}{m(\overline{Y}_m - 1)} + \frac{1}{n\overline{X}_n(1 - \overline{X}_n)}.$$

### 6.4. Inverse-inverse sampling scheme

From Table 3, the statistic of interest is

$$\log \widetilde{\rho}_{m_1,m_2} = \log\left(\frac{\overline{Y}_{m_2} - 1}{\overline{Y}_{m_1} - 1}\right) = g_{ii}(V_1, V_2) = \log(V_2 - 1) - \log(V_1 - 1),$$

where $V_1 = \overline{Y}_{m_1}$ and $V_2 = \overline{Y}_{m_2}$. In this particular case, the function $g_{ii}(v_1, v_2) = \log(v_2 - 1) - \log(v_1 - 1)$.

Note that $EV_i = \frac{1}{p_i}, \text{Var}V_i = \frac{q_i}{m_i p_i^2}, i = 1, 2$ and $g_{ii}\left(\frac{1}{p_1}, \frac{1}{p_2}\right) = \log\left(\frac{1}{p_2} - 1\right) - \log\left(\frac{1}{p_1} - 1\right) = \log \rho$.

Partial derivatives are:

$$\frac{\partial g_{ii}(v_1, v_2)}{\partial v_1} = -\frac{1}{v_1 - 1} \text{ and } \frac{\partial g_{ii}(v_1, v_2)}{\partial v_2} = \frac{1}{v_2 - 1};$$

thus,

$$\frac{\partial g_{ii}\left(\frac{1}{p_1}, \frac{1}{p_2}\right)}{\partial v_1} = -\frac{p_1}{q_1} \text{ and } \frac{\partial g_{ii}\left(\frac{1}{p_1}, \frac{1}{p_2}\right)}{\partial v_2} = \frac{p_2}{q_2}.$$

Linear term Taylor expansion in the neighborhoods of the mean values of the statistics takes the form:

$$\log \widetilde{\rho}_{m_1,m_2} = g_{ii}(V_1, V_2) \approx \log \rho - \left(\frac{p_1}{q_1}\right)\left(\overline{Y}_{m_1} - \frac{1}{p_1}\right) + \frac{p_2}{q_2}\left(\overline{Y}_{m_2} - \frac{1}{p_2}\right).$$

From this, the estimator $\log(\widetilde{\rho}_{m_1,m_2})$ is approximately normal with

$$\text{Mean} = \log(\rho)$$

and (remember that $\overline{Y}_{m_1}$ and $\overline{Y}_{m_2}$ are independent)

$$\text{Variance} = s^2 = \left(\frac{p_1}{q_1}\right)^2 \frac{q_1}{m_1 p_1^2} + \left(\frac{p_2}{q_2}\right)^2 \frac{q_2}{m_2 p_2^2} = \left(\frac{p_1}{q_1}\right)(p_1^{-1})/m_1 + \left(\frac{p_2}{q_2}\right)(p_2^{-1})/m_2.$$

To obtain the plug-in estimator of the variance, we substitute estimations for $\frac{p}{q}$ and $p^{-1}$ (see Tables 1 and 2), namely $\widetilde{p_1/q_1} = \frac{1}{\overline{Y}_{m_1} - 1}, \widetilde{p_2/q_2} = \frac{1}{\overline{Y}_{m_2} - 1}, \widetilde{p_1^{-1}} = \overline{Y}_{m_1}$, and $\widetilde{p_2^{-1}} = \overline{Y}_{m_2}$, and obtain that

**Table 4.** Plug-in estimators of the variance component of estimators $s^2(p_1, p_2)$ for all possible combinations of direct and inverse sampling schemes.

| Sampling Schemes | Second Sample Direct | Second Sample Inverse |
|---|---|---|
| First Sample Direct | $\frac{1}{n_1(1-\overline{X}_{n_1})\overline{X}_{n_1}} + \frac{1}{n_2(1-\overline{X}_{n_2})\overline{X}_{n_2}}$ | $\frac{1}{n\overline{X}_n(1-\overline{X}_n)} + \frac{\overline{Y}_m}{m(\overline{Y}_m-1)}$ |
| First Sample Inverse | $\frac{\overline{Y}_m}{m(\overline{Y}_m-1)} + \frac{1}{n\overline{X}_n(1-\overline{X}_n)}$ | $\frac{\overline{Y}_{m_1}}{m_1(\overline{Y}_{m_1}-1)} + \frac{\overline{Y}_{m_2}}{m_2(\overline{Y}_{m_2}-1)}$ |

$$\begin{aligned}\hat{s}^2 &= \frac{1}{\overline{Y}_{m_1} - 1}\overline{Y}_{m_1}/m_1 + \frac{1}{\overline{Y}_{m_2} - 1}\overline{Y}_{m_2}/m_2 \\ &= \frac{\overline{Y}_{m_1}}{m_1(\overline{Y}_{m_1} - 1)} + \frac{\overline{Y}_{m_2}}{m_2(\overline{Y}_{m_2} - 1)}.\end{aligned}$$

Plug-in estimators of the asymptotic variance $s^2(p_1, p_2)$ are presented in Table 4.

## 7. Confidence limits for logarithmic interval

As mentioned, for all sampling schemes, the normal approximations for estimators $\log(\widetilde{\rho})$ show that means and variances have the same structure: Mean $= \log(\rho)$ and Variance $= s^2(p_1, p_2)$.

If the sample sizes in both sampling schemes tend to infinity, then using the inequity

$$|\log\rho - \log\hat{\rho}| \leq z_{\alpha/2}s(p_1, p_2),$$

(where $z_{\alpha/2}$ is $(1-\alpha/2)$-quantile of the standard normal distribution) and replacing $s^2(p_1, p_2)$ by its estimators that correspond to sampling schemes presented in Table 4, gives us the following end points for an asymptotically $(1-\alpha)$-confidence interval for the cross-product ratio $\rho$:

$$\hat{\rho}\exp\{\mp z_{\alpha/2}\hat{s}\}. \tag{1}$$

### *7.1. Direct-direct sampling scheme*

When both samples are obtained by a direct sampling scheme with sample sizes $n_1$ and $n_2$, then, according to Tables 3 and 4:

$$\begin{aligned}\hat{\rho}_{n_1,n_2} &= \frac{n_1\overline{X}_{n_1}}{n_1 + 1 - n_1\overline{X}_{n_1}}\left(\frac{n_2+1}{n_2\overline{X}_{n_2}+1} - 1\right) \quad \text{and} \\ \hat{s}^2 &= \frac{1}{n_1\overline{X}_{n_1}(1-\overline{X}_{n_1})} + \frac{1}{n_2\overline{X}_{n_2}(1-\overline{X}_{n_2})}.\end{aligned}$$

Hence, the asymptotic $n_1, n_2 \to \infty$ confidence interval (1) based on the relative frequencies $\overline{X}_{n_1}$ and $\overline{X}_{n_2}$ of successes (sample means) in each sample can be written as

$$\frac{n_1\overline{X}_{n_1}}{n_1 + 1 - n_1\overline{X}_{n_1}}\left(\frac{n_2+1}{n_2\overline{X}_{n_2}+1} - 1\right)\exp\left\{\mp z_{\alpha/2}\sqrt{\frac{1}{n_1\overline{X}_{n_1}(1-\overline{X}_{n_1})} + \frac{1}{n_2\overline{X}_{n_2}(1-\overline{X}_{n_2})}}\right\}. \tag{2}$$

Table 5 contains some simulation results. For each pair ($n_1$, $n_2$) of sample sizes and values ($p_1$, $p_2$) of success probabilities, we present the Monte-Carlo estimations of the coverage probability, mean width, and standard deviation of the width for the confidence interval (2). The nominal level is assumed to be 0.95.

The logarithmic interval (Table 5) has good coverage probability with an error less than 0.01 in most of the cases.

**Table 5.** Coverage probability, width, and standard deviation for logarithmic CI (2).

| | $n_2$ | 20 | | | 50 | | | 100 | | |
|---|---|---|---|---|---|---|---|---|---|---|
| $n_1$ | $p_1 \setminus p_2$ | 0.2 | 0.5 | 0.8 | 0.2 | 0.5 | 0.8 | 0.2 | 0.5 | 0.8 |
| | | 0.952 | 0.956 | 0.957 | 0.970 | 0.967 | 0.959 | 0.971 | 0.970 | 0.969 |
| | 0.2 | 5.26 | 0.972 | 0.277 | 3.475 | 0.773 | 0.208 | 3.006 | 0.710 | 0.186 |
| | | 6.040 | 0.691 | 0.178 | 2.412 | 0.388 | 0.108 | 1.492 | 0.309 | 0.085 |
| | | 0.941 | 0.952 | 0.961 | 0.955 | 0.952 | 0.955 | 0.956 | 0.960 | 0.959 |
| 20 | 0.5 | 18.624 | 3.346 | 0.968 | 11.838 | 2.556 | 0.706 | 9.916 | 2.322 | 0.614 |
| | | 22.755 | 2.621 | 0.686 | 9.810 | 1.613 | 0.461 | 6.402 | 1.431 | 0.379 |
| | | 0.903 | 0.941 | 0.950 | 0.939 | 0.934 | 0.939 | 0.935 | 0.937 | 0.940 |
| | 0.8 | 95.361 | 18.769 | 5.225 | 68.765 | 15.663 | 4.143 | 61.079 | 14.685 | 3.795 |
| | | 140.072 | 22.940 | 6.088 | 85.213 | 17.834 | 4.687 | 70.067 | 16.493 | 4.261 |
| | | 0.938 | 0.954 | 0.960 | 0.959 | 0.955 | 0.956 | 0.957 | 0.956 | 0.955 |
| | 0.2 | 4.140 | 0.705 | 0.208 | 2.422 | 0.504 | 0.142 | 1.933 | 0.439 | 0.119 |
| | | 4.682 | 0.463 | 0.108 | 1.628 | 0.208 | 0.059 | 0.817 | 0.147 | 0.042 |
| | | 0.934 | 0.953 | 0.968 | 0.952 | 0.954 | 0.956 | 0.951 | 0.949 | 0.953 |
| 50 | 0.5 | 15.521 | 2.568 | 0.771 | 8.709 | 1.758 | 0.505 | 6.645 | 1.481 | 0.407 |
| | | 17.617 | 1.631 | 0.383 | 6.316 | 0.749 | 0.208 | 2.942 | 0.537 | 0.150 |
| | | 0.939 | 0.956 | 0.970 | 0.946 | 0.951 | 0.959 | 0.949 | 0.953 | 0.955 |
| | 0.8 | 68.947 | 11.813 | 3.501 | 41.637 | 8.731 | 2.432 | 33.442 | 7.705 | 2.059 |
| | | 84.318 | 9.437 | 2.501 | 36.558 | 6.572 | 1.696 | 24.246 | 5.480 | 1.442 |
| | | 0.940 | 0.958 | 0.968 | 0.955 | 0.953 | 0.955 | 0.952 | 0.952 | 0.948 |
| | 0.2 | 3.805 | 0.613 | 0.186 | 2.066 | 0.407 | 0.119 | 1.532 | 0.335 | 0.094 |
| | | 4.269 | 0.383 | 0.085 | 1.457 | 0.150 | 0.042 | 0.601 | 0.097 | 0.028 |
| | | 0.936 | 0.959 | 0.970 | 0.953 | 0.947 | 0.956 | 0.949 | 0.953 | 0.952 |
| 100 | 0.5 | 14.737 | 2.320 | 0.711 | 7.717 | 1.477 | 0.441 | 5.511 | 1.175 | 0.335 |
| | | 16.584 | 1.392 | 0.308 | 5.734 | 0.539 | 0.148 | 2.198 | 0.341 | 0.097 |
| | | 0.937 | 0.956 | 0.970 | 0.950 | 0.951 | 0.958 | 0.949 | 0.949 | 0.953 |
| | 0.8 | 61.385 | 9.928 | 3.008 | 33.391 | 6.659 | 1.929 | 25.051 | 5.514 | 1.530 |
| | | 70.291 | 6.533 | 1.501 | 24.078 | 2.979 | 0.811 | 11.760 | 2.202 | 0.603 |

Table 6 shows that, for $n_1, n_2 \geq 50$ and $p_1, p_2 \geq 0.05$, except for $n_2 = 50, p_2 = 0.05$ and $n_1 = 50, p_1 = 0.05$, the logarithmic interval has good coverage probability even for small values of success probabilities, but its accuracy properties in this region are poor.

## 7.2. Direct-inverse sampling scheme

When the first sample is obtained by the direct sampling scheme with sample size $n$ and the second sample is obtained by the inverse sampling scheme with the number of successes $m$, then, according to Tables 3 and 4:

$$\hat{\rho}_{n,m} = \frac{n\overline{X}_n}{n+1-n\overline{X}_n}(\overline{Y}_m - 1) \quad \text{and}$$
$$\hat{s}^2 = \frac{1}{n\overline{X}_n(1-\overline{X}_n)} + \frac{\overline{Y}_m}{m(\overline{Y}_m - 1)}.$$

Hence the asymptotic $n, m \to \infty$ confidence interval (1) based on $\overline{X}_n$ and $\overline{Y}_m$ can be written as

$$\frac{n\overline{X}_n}{n+1-n\overline{X}_n}(\overline{Y}_m - 1)\exp\left\{\mp z_{\alpha/2}\sqrt{\frac{1}{n\overline{X}_n(1-\overline{X}_n)} + \frac{\overline{Y}_m}{m(\overline{Y}_m - 1)}}\right\}. \tag{3}$$

In Table 7, we provide the results of statistical modeling. For each pair $(n, m)$ of sample sizes and values $(p_1, p_2)$ of success probabilities, we present the Monte-Carlo estimations of the coverage probability, mean width, and standard deviation of the width for the confidence interval (3). The nominal level is assumed to be 0.95.

The logarithmic interval (Table 7) has high correspondence for coverage probability to the nominal, but its precision properties region are not satisfactory. This confidence interval can be

**Table 6.** Coverage probability, width, and standard deviation for logarithmic CI (2) for small probabilities.

| | $n_2$ | 20 | | | 50 | | | 100 | | |
|---|---|---|---|---|---|---|---|---|---|---|
| $n_1$ | $p_1 \setminus p_2$ | 0.05 | 0.1 | 0.15 | 0.05 | 0.1 | 0.15 | 0.05 | 0.1 | 0.15 |
| | | 0.411 | 0.563 | 0.614 | 0.589 | 0.629 | 0.626 | 0.629 | 0.624 | 0.615 |
| | 0.05 | 3.158 | 3.078 | 2.323 | 6.925 | 3.408 | 1.892 | 7.105 | 2.778 | 1.639 |
| | | 4.650 | 4.043 | 3.169 | 9.088 | 4.623 | 2.195 | 9.515 | 2.785 | 1.466 |
| | | 0.553 | 0.765 | 0.842 | 0.811 | 0.869 | 0.870 | 0.869 | 0.861 | 0.860 |
| 20 | 0.1 | 5.645 | 5.402 | 4.072 | 12.215 | 5.775 | 3.171 | 12.065 | 4.539 | 2.673 |
| | | 6.905 | 5.851 | 4.627 | 13.207 | 6.544 | 3.043 | 13.705 | 3.383 | 1.629 |
| | | 0.585 | 0.812 | 0.908 | 0.879 | 0.950 | 0.950 | 0.951 | 0.952 | 0.946 |
| | 0.2 | 8.102 | 7.746 | 5.743 | 17.391 | 7.948 | 4.276 | 16.703 | 6.113 | 3.527 |
| | | 9.426 | 8.005 | 6.313 | 18.140 | 8.820 | 3.982 | 18.566 | 4.524 | 1.946 |
| | | 0.568 | 0.797 | 0.882 | 0.848 | 0.916 | 0.914 | 0.913 | 0.908 | 0.904 |
| | 0.05 | 2.461 | 2.344 | 1.743 | 5.265 | 2.436 | 1.304 | 5.069 | 1.873 | 1.081 |
| | | 2.840 | 2.402 | 1.879 | 5.399 | 2.682 | 1.173 | 5.521 | 1.309 | 0.573 |
| | | 0.576 | 0.822 | 0.917 | 0.882 | 0.967 | 0.972 | 0.965 | 0.970 | 0.972 |
| 50 | 0.1 | 4.397 | 4.170 | 3.060 | 9.248 | 4.069 | 2.093 | 8.497 | 2.929 | 1.649 |
| | | 4.813 | 4.067 | 3.224 | 9.157 | 4.537 | 1.853 | 9.352 | 2.073 | 0.794 |
| | | 0.566 | 0.824 | 0.908 | 0.869 | 0.962 | 0.965 | 0.958 | 0.960 | 0.963 |
| | 0.2 | 6.615 | 6.159 | 4.487 | 13.625 | 5.885 | 2.989 | 12.228 | 4.031 | 2.241 |
| | | 7.045 | 5.917 | 4.716 | 13.445 | 6.570 | 2.843 | 13.786 | 2.821 | 1.082 |
| | | 0.579 | 0.836 | 0.922 | 0.886 | 0.965 | 0.974 | 0.964 | 0.975 | 0.973 |
| | 0.05 | 2.077 | 1.956 | 1.425 | 4.333 | 1.889 | 0.978 | 3.983 | 1.354 | 0.764 |
| | | 2.239 | 1.876 | 1.482 | 4.270 | 2.082 | 0.871 | 4.407 | 0.930 | 0.365 |
| | | 0.560 | 0.827 | 0.913 | 0.869 | 0.953 | 0.962 | 0.952 | 0.959 | 0.961 |
| 100 | 0.1 | 3.951 | 3.711 | 2.673 | 8.175 | 3.452 | 1.709 | 7.121 | 2.281 | 1.243 |
| | | 4.142 | 3.486 | 2.758 | 7.928 | 3.927 | 1.641 | 8.081 | 1.672 | 0.589 |
| | | 0.560 | 0.817 | 0.910 | 0.863 | 0.947 | 0.954 | 0.944 | 0.960 | 0.956 |
| | 0.2 | 6.093 | 5.697 | 4.073 | 12.424 | 5.159 | 2.511 | 10.739 | 3.296 | 1.772 |
| | | 6.263 | 5.274 | 4.176 | 11.941 | 5.860 | 2.380 | 12.237 | 2.337 | 0.838 |

recommended for the practical applications for values of $p_1 \geq 0.1$ (see also Table 8) with sample size $n \geq 100$ for all values of the parameters of the second sample. Also it can be recommended for all $p_1, p_2 \geq 0.05,$ if $n \geq 200$.

### *7.3. Inverse-direct sampling scheme*

When the first sample is obtained by the inverse sampling scheme with the number of successes $m$, and the second sample is obtained by the direct sampling scheme with sample size $n$, then, according to Tables 3 and 4:

$$\hat{\rho}_{m,n} = \left(\frac{m-1}{m\overline{Y}_m - m + 1}\right)\left(\frac{n+1}{n\overline{X}_n + 1} - 1\right) \quad \text{and}$$
$$\hat{s}^2 = \frac{\overline{Y}_m}{m(\overline{Y}_m - 1)} + \frac{1}{n\overline{X}_n(1 - \overline{X}_n)}.$$

Hence, the asymptotic $m, n \to \infty$ confidence interval (1) based on $\overline{Y}_m$ and $\overline{X}_n$ can be written as

$$\left(\frac{m-1}{m\overline{Y}_m - m + 1}\right)\left(\frac{n+1}{n\overline{X}_n + 1} - 1\right) \exp\left\{\mp z_{\alpha/2}\sqrt{\frac{\overline{Y}_m}{m(\overline{Y}_m - 1)} + \frac{1}{n\overline{X}_n(1 - \overline{X}_n)}}\right\}. \tag{4}$$

In Table 9, we provide the results of statistical modeling. For each pair $(m, n)$ of number of successes and sample size, and values $(p_1, p_2)$ of success probabilities, we present the Monte-Carlo estimations of the coverage probability, mean width, and standard deviation of the width for the confidence interval (4). The nominal level is assumed to be 0.95.

Precision properties of the logarithmic interval (4) (Table 9) show that the coverage probability is still within the acceptable error 0.01. The accuracy and reliability properties of this interval for

Table 7. Coverage probability, width, and standard deviation for logarithmic CI (3).

| | *m* | 20 | | | 50 | | | 100 | | |
|---|---|---|---|---|---|---|---|---|---|---|
| *n* | $p_1 \setminus p_2$ | 0.2 | 0.5 | 0.8 | 0.2 | 0.5 | 0.8 | 0.2 | 0.5 | 0.8 |
| | | 0.967 | 0.962 | 0.954 | 0.969 | 0.968 | 0.962 | 0.970 | 0.969 | 0.968 |
| | 0.2 | 2.973 | 0.797 | 0.255 | 2.757 | 0.711 | 0.200 | 2.680 | 0.679 | 0.181 |
| | | 1.380 | 0.399 | 0.146 | 1.125 | 0.305 | 0.097 | 1.022 | 0.268 | 0.079 |
| | | 0.956 | 0.952 | 0.962 | 0.961 | 0.958 | 0.957 | 0.962 | 0.960 | 0.957 |
| 20 | 0.5 | 9.794 | 2.674 | 0.883 | 8.931 | 2.313 | 0.668 | 8.622 | 2.204 | 0.593 |
| | | 5.969 | 1.684 | 0.573 | 5.612 | 1.368 | 0.423 | 5.023 | 1.304 | 0.354 |
| | | 0.936 | 0.938 | 0.951 | 0.937 | 0.936 | 0.940 | 0.937 | 0.937 | 0.938 |
| | 0.8 | 60.524 | 15.995 | 4.883 | 57.285 | 14.647 | 3.991 | 55.967 | 14.265 | 3.729 |
| | | 68.535 | 18.009 | 5.502 | 64.081 | 16.383 | 4.470 | 62.263 | 16.045 | 4.191 |
| | | 0.953 | 0.952 | 0.956 | 0.958 | 0.955 | 0.952 | 0.960 | 0.958 | 0.955 |
| | 0.2 | 1.899 | 0.532 | 0.188 | 1.654 | 0.438 | 0.133 | 1.571 | 0.406 | 0.114 |
| | | 0.698 | 0.211 | 0.084 | 0.494 | 0.142 | 0.050 | 0.424 | 0.116 | 0.038 |
| | | 0.948 | 0.950 | 0.966 | 0.953 | 0.949 | 0.954 | 0.954 | 0.955 | 0.952 |
| 50 | 0.5 | 6.540 | 1.870 | 0.689 | 5.484 | 1.477 | 0.468 | 5.112 | 1.332 | 0.387 |
| | | 2.539 | 0.753 | 0.296 | 1.854 | 0.523 | 0.178 | 1.640 | 0.435 | 0.138 |
| | | 0.952 | 0.953 | 0.969 | 0.954 | 0.954 | 0.963 | 0.954 | 0.955 | 0.957 |
| | 0.8 | 33.012 | 9.166 | 3.167 | 29.250 | 7.706 | 2.291 | 27.824 | 7.189 | 1.990 |
| | | 23.503 | 6.633 | 2.119 | 20.743 | 5.767 | 1.558 | 19.857 | 5.047 | 1.447 |
| | | 0.948 | 0.952 | 0.963 | 0.952 | 0.951 | 0.954 | 0.953 | 0.952 | 0.952 |
| | 0.2 | 1.498 | 0.437 | 0.165 | 1.222 | 0.334 | 0.109 | 1.120 | 0.295 | 0.088 |
| | | 0.477 | 0.148 | 0.062 | 0.302 | 0.090 | 0.034 | 0.241 | 0.068 | 0.024 |
| | | 0.948 | 0.951 | 0.966 | 0.950 | 0.948 | 0.957 | 0.951 | 0.951 | 0.954 |
| 100 | 0.5 | 5.371 | 1.607 | 0.626 | 4.179 | 1.170 | 0.400 | 3.732 | 1 | 0.312 |
| | | 1.672 | 0.527 | 0.220 | 1.069 | 0.316 | 0.119 | 0.869 | 0.243 | 0.083 |
| | | 0.950 | 0.953 | 0.970 | 0.952 | 0.952 | 0.960 | 0.953 | 0.953 | 0.953 |
| | 0.8 | 24.508 | 7.133 | 2.673 | 20.204 | 5.504 | 1.778 | 18.630 | 4.904 | 1.445 |
| | | 9.921 | 2.951 | 1.130 | 7.755 | 2.139 | 0.700 | 7.001 | 1.846 | 0.553 |

Table 8. Coverage probability, width, and standard deviation for logarithmic CI (3) for small probabilities.

| | *m* | 20 | | | 50 | | | 100 | | |
|---|---|---|---|---|---|---|---|---|---|---|
| *n* | $p_1 \setminus p_2$ | 0.05 | 0.1 | 0.15 | 0.05 | 0.1 | 0.15 | 0.05 | 0.1 | 0.15 |
| | | 0.611 | 0.613 | 0.613 | 0.611 | 0.611 | 0.615 | 0.615 | 0.612 | 0.614 |
| | 0.05 | 5.274 | 2.514 | 1.589 | 5.078 | 2.412 | 1.527 | 5.005 | 2.368 | 1.502 |
| | | 4.402 | 2.098 | 1.330 | 4.049 | 1.927 | 1.213 | 3.926 | 1.868 | 1.178 |
| | | 0.858 | 0.855 | 0.856 | 0.854 | 0.855 | 0.853 | 0.850 | 0.849 | 0.851 |
| 20 | 0.1 | 8.491 | 4.042 | 2.552 | 8.077 | 3.839 | 2.411 | 7.938 | 3.754 | 2.377 |
| | | 4.472 | 2.160 | 1.370 | 3.900 | 1.859 | 1.180 | 3.726 | 1.767 | 1.116 |
| | | 0.942 | 0.941 | 0.941 | 0.942 | 0.941 | 0.941 | 0.939 | 0.939 | 0.939 |
| | 0.15 | 11.133 | 5.293 | 3.361 | 10.505 | 4.979 | 3.149 | 10.299 | 4.870 | 3.077 |
| | | 5.087 | 2.454 | 1.571 | 4.229 | 2.023 | 1.286 | 3.963 | 1.881 | 1.195 |
| | | 0.898 | 0.899 | 0.898 | 0.894 | 0.894 | 0.894 | 0.893 | 0.893 | 0.893 |
| | 0.05 | 3.415 | 1.630 | 1.033 | 3.223 | 1.530 | 0.966 | 3.159 | 1.500 | 0.945 |
| | | 1.501 | 0.722 | 0.463 | 1.248 | 0.595 | 0.379 | 1.159 | 0.551 | 0.348 |
| | | 0.967 | 0.966 | 0.967 | 0.969 | 0.969 | 0.968 | 0.969 | 0.9698 | 0.968 |
| 50 | 0.1 | 5.082 | 2.427 | 1.537 | 4.676 | 2.218 | 1.406 | 4.537 | 2.157 | 1.360 |
| | | 1.845 | 0.894 | 0.571 | 1.384 | 0.663 | 0.426 | 1.228 | 0.585 | 0.373 |
| | | 0.959 | 0.958 | 0.958 | 0.963 | 0.963 | 0.962 | 0.963 | 0.963 | 0.963 |
| | 0.15 | 6.793 | 3.250 | 2.062 | 6.141 | 2.910 | 1.846 | 5.900 | 2.805 | 1.769 |
| | | 2.429 | 1.175 | 0.754 | 1.796 | 0.852 | 0.550 | 1.574 | 0.750 | 0.475 |
| | | 0.966 | 0.965 | 0.966 | 0.966 | 0.967 | 0.968 | 0.966 | 0.965 | 0.965 |
| | 0.05 | 2.343 | 1.121 | 0.710 | 2.152 | 1.021 | 0.645 | 2.083 | 0.989 | 0.624 |
| | | 0.801 | 0.389 | 0.249 | 0.585 | 0.278 | 0.177 | 0.505 | 0.241 | 0.153 |
| | | 0.954 | 0.954 | 0.954 | 0.957 | 0.958 | 0.958 | 0.959 | 0.959 | 0.959 |
| 100 | 0.1 | 3.686 | 1.765 | 1.126 | 3.247 | 1.544 | 0.980 | 3.089 | 1.467 | 0.927 |
| | | 1.196 | 0.581 | 0.374 | 0.821 | 0.394 | 0.253 | 0.682 | 0.328 | 0.208 |
| | | 0.949 | 0.950 | 0.951 | 0.954 | 0.954 | 0.953 | 0.955 | 0.954 | 0.955 |
| | 0.15 | 5.145 | 2.470 | 1.580 | 4.395 | 2.093 | 1.328 | 4.123 | 1.960 | 1.239 |
| | | 1.624 | 0.786 | 0.508 | 1.074 | 0.515 | 0.332 | 0.878 | 0.422 | 0.268 |

**Table 9.** Coverage probability, width, and standard deviation for CI (4).

| | $n$ | 20 | | | 50 | | | 100 | | |
|---|---|---|---|---|---|---|---|---|---|---|
| $m$ | $p_1 \setminus p_2$ | 0.2 | 0.5 | 0.8 | 0.2 | 0.5 | 0.8 | 0.2 | 0.5 | 0.8 |
| | | 0.936 | 0.956 | 0.969 | 0.950 | 0.949 | 0.954 | 0.947 | 0.948 | 0.949 |
| | 0.2 | 3.818 | 0.616 | 0.186 | 2.073 | 0.411 | 0.119 | 1.543 | 0.338 | 0.094 |
| | | 4.352 | 0.399 | 0.090 | 1.484 | 0.173 | 0.048 | 0.684 | 0.122 | 0.034 |
| | | 0.934 | 0.954 | 0.970 | 0.949 | 0.951 | 0.956 | 0.950 | 0.952 | 0.953 |
| 20 | 0.5 | 16.308 | 2.754 | 0.820 | 9.487 | 1.956 | 0.552 | 7.472 | 1.690 | 0.459 |
| | | 19.088 | 2.218 | 0.515 | 7.593 | 1.127 | 0.306 | 4.552 | 0.974 | 0.252 |
| | | 0.920 | 0.942 | 0.960 | 0.931 | 0.933 | 0.945 | 0.935 | 0.934 | 0.933 |
| | 0.8 | 89.014 | 17.067 | 4.83 | 62.722 | 13.964 | 3.756 | 54.760 | 13.132 | 3.383 |
| | | 129.609 | 21.044 | 5.538 | 79.098 | 16.021 | 4.310 | 63.799 | 15.194 | 3.880 |
| | | 0.938 | 0.960 | 0.970 | 0.955 | 0.953 | 0.957 | 0.950 | 0.950 | 0.952 |
| | 0.2 | 3.592 | 0.558 | 0.172 | 1.838 | 0.343 | 0.104 | 1.266 | 0.262 | 0.076 |
| | | 4.025 | 0.352 | 0.070 | 1.405 | 0.118 | 0.032 | 0.491 | 0.070 | 0.020 |
| | | 0.934 | 0.958 | 0.970 | 0.952 | 0.951 | 0.957 | 0.952 | 0.950 | 0.954 |
| 50 | 0.5 | 14.636 | 2.334 | 0.713 | 7.719 | 1.490 | 0.441 | 5.557 | 1.183 | 0.336 |
| | | 16.565 | 1.417 | 0.315 | 5.597 | 0.572 | 0.156 | 2.315 | 0.379 | 0.105 |
| | | 0.934 | 0.957 | 0.973 | 0.947 | 0.955 | 0.960 | 0.949 | 0.952 | 0.957 |
| | 0.8 | 66.121 | 11.239 | 3.319 | 38.796 | 8.046 | 2.271 | 30.822 | 6.985 | 1.889 |
| | | 79.966 | 9.138 | 2.233 | 33.627 | 5.766 | 1.546 | 22.1 | 4.814 | 1.285 |
| | | 0.935 | 0.962 | 0.969 | 0.954 | 0.955 | 0.959 | 0.953 | 0.950 | 0.954 |
| | 0.2 | 3.505 | 0.539 | 0.167 | 1.750 | 0.319 | 0.098 | 1.168 | 0.233 | 0.070 |
| | | 3.914 | 0.318 | 0.064 | 1.284 | 0.102 | 0.027 | 0.437 | 0.055 | 0.015 |
| | | 0.936 | 0.963 | 0.970 | 0.956 | 0.952 | 0.959 | 0.952 | 0.950 | 0.952 |
| 100 | 0.5 | 14.283 | 2.204 | 0.682 | 7.181 | 1.337 | 0.406 | 4.907 | 1.006 | 0.296 |
| | | 15.970 | 1.339 | 0.271 | 5.093 | 0.454 | 0.120 | 1.889 | 0.263 | 0.073 |
| | | 0.935 | 0.958 | 0.970 | 0.953 | 0.948 | 0.957 | 0.951 | 0.951 | 0.953 |
| | 0.8 | 59.998 | 9.604 | 2.926 | 32.146 | 6.315 | 1.851 | 23.641 | 5.132 | 1.443 |
| | | 68.037 | 6.076 | 1.389 | 23.452 | 2.712 | 0.734 | 10.701 | 1.951 | 0.537 |

small values of $p_1$ and $p_2$ are presented in Table 10. According to these results, for small values of success probabilities, it is possible to recommend the logarithmic interval for the sample sizes of the second sample $n = 50, p_2 \geq 0.1$ and $n \geq 100, p_2 \geq 0.05$.

### *7.4. Inverse-inverse sampling scheme*

When both samples are obtained by the inverse sampling scheme with the number of successes $m_1$ and $m_2$, then, according to Tables 3 and 4:

$$\hat{\rho}_{m_1,m_2} = \left(\frac{m_1 - 1}{m_1\overline{Y}_{m_1} - m_1 + 1}\right)(\overline{Y}_{m_2} - 1) \text{ and}$$

$$\hat{s}^2 = \frac{\overline{Y}_{m_1}}{m_1(\overline{Y}_{m_1} - 1)} + \frac{\overline{Y}_{m_2}}{m_2(\overline{Y}_{m_2} - 1)}.$$

Hence, the asymptotic $m_1, m_2 \to \infty$ confidence interval (1) based on $\overline{Y}_{m_1}$ and $\overline{Y}_{m_2}$ can be written as

$$\left(\frac{m_1 - 1}{m_1\overline{Y}_{m_1} - m_1 + 1}\right)(\overline{Y}_{m_2} - 1)\exp\left\{\mp z_{\alpha/2}\sqrt{\frac{\overline{Y}_{m_1}}{m_1(\overline{Y}_{m_1} - 1)} + \frac{\overline{Y}_{m_2}}{m_2(\overline{Y}_{m_2} - 1)}}\right\}. \tag{5}$$

In Table 11, we provide the results of statistical modeling. For each pair ($m_1$, $m_2$) of numbers of successes and values ($p_1$, $p_2$) of success probabilities, we present the Monte-Carlo estimations of the coverage probability, mean width, and standard deviation of the width for the confidence interval (5). The nominal level is assumed to be 0.95.

Table 11 shows the precise correspondence of the coverage probability to the nominal for the asymptotic confidence interval (5), except in the cases $m_2 = 20, p_2 = 0.8$ and $m_1 = 20, p_1 = 0.8$.

**Table 10.** Coverage probability, width, and standard deviation for CI (4) for small probabilities.

| | $n$ | 20 | | | 50 | | | 100 | | |
|---|---|---|---|---|---|---|---|---|---|---|
| $m$ | $p_1 \setminus p_2$ | 0.05 | 0.1 | 0.15 | 0.05 | 0.1 | 0.15 | 0.05 | 0.1 | 0.15 |
| | | 0.561 | 0.812 | 0.903 | 0.857 | 0.941 | 0.951 | 0.942 | 0.952 | 0.949 |
| | 0.05 | 1.762 | 1.643 | 1.183 | 3.572 | 1.471 | 0.702 | 3.052 | 0.913 | 0.480 |
| | | 1.801 | 1.520 | 1.220 | 3.438 | 1.697 | 0.684 | 3.516 | 0.685 | 0.238 |
| | | 0.567 | 0.811 | 0.904 | 0.860 | 0.943 | 0.952 | 0.940 | 0.952 | 0.949 |
| 20 | 0.1 | 3.760 | 3.460 | 2.494 | 7.604 | 3.114 | 1.508 | 6.467 | 1.961 | 1.026 |
| | | 3.824 | 3.211 | 2.573 | 7.321 | 3.581 | 1.498 | 7.498 | 1.511 | 0.528 |
| | | 0.565 | 0.813 | 0.902 | 0.862 | 0.942 | 0.951 | 0.942 | 0.951 | 0.949 |
| | 0.15 | 6.010 | 5.561 | 3.986 | 12.165 | 5 | 2.403 | 10.357 | 3.162 | 1.659 |
| | | 6.143 | 5.161 | 4.140 | 11.782 | 5.787 | 2.402 | 12.050 | 2.499 | 0.852 |
| | | 0.568 | 0.811 | 0.902 | 0.857 | 0.941 | 0.953 | 0.940 | 0.954 | 0.954 |
| | 0.05 | 1.716 | 1.582 | 1.124 | 3.443 | 1.384 | 0.647 | 2.870 | 0.816 | 0.413 |
| | | 1.687 | 1.416 | 1.136 | 3.231 | 1.601 | 0.638 | 3.335 | 0.629 | 0.194 |
| | | 0.566 | 0.810 | 0.904 | 0.859 | 0.942 | 0.952 | 0.939 | 0.953 | 0.950 |
| 50 | 0.1 | 3.627 | 3.346 | 2.383 | 7.262 | 2.960 | 1.372 | 6.093 | 1.725 | 0.881 |
| | | 3.577 | 3.001 | 2.394 | 6.807 | 3.441 | 1.380 | 7.057 | 1.255 | 0.421 |
| | | 0.565 | 0.811 | 0.905 | 0.860 | 0.942 | 0.951 | 0.940 | 0.952 | 0.951 |
| | 0.15 | 5.736 | 5.304 | 3.802 | 11.599 | 4.681 | 2.185 | 9.740 | 2.758 | 1.405 |
| | | 5.683 | 4.765 | 3.827 | 10.896 | 5.415 | 2.208 | 11.359 | 2.051 | 0.665 |
| | | 0.565 | 0.814 | 0.904 | 0.853 | 0.942 | 0.952 | 0.938 | 0.953 | 0.954 |
| | 0.05 | 1.697 | 1.553 | 1.112 | 3.391 | 1.355 | 0.626 | 2.807 | 0.784 | 0.390 |
| | | 1.655 | 1.379 | 1.113 | 3.168 | 1.566 | 0.612 | 3.268 | 0.605 | 0.184 |
| | | 0.567 | 0.813 | 0.904 | 0.855 | 0.940 | 0.952 | 0.939 | 0.954 | 0.953 |
| 100 | 0.1 | 3.579 | 3.296 | 2.355 | 7.129 | 2.872 | 1.329 | 5.908 | 1.663 | 0.826 |
| | | 3.489 | 2.927 | 2.358 | 6.667 | 3.342 | 1.343 | 6.861 | 1.317 | 0.387 |
| | | 0.567 | 0.813 | 0.903 | 0.855 | 0.941 | 0.952 | 0.939 | 0.952 | 0.952 |
| | 0.15 | 5.694 | 5.238 | 3.742 | 11.373 | 4.567 | 2.118 | 9.420 | 2.643 | 1.318 |
| | | 5.560 | 4.658 | 3.750 | 10.618 | 5.290 | 2.100 | 10.931 | 2.043 | 0.623 |

The results for small probabilities of success in Table 12 have the coverage probability close to the nominal.

## 8. Comparison of logarithmic confidence estimator accuracy for different sampling schemes

For the inverse binomial sampling scheme with parameters ($p$, $m$), the mean sample size is $E(\nu) = m/p$. If the observations are obtained in the direct sampling scheme with the same probability $p$ of success and sample size $n = m/p$, then, on average, it is equivalent to the inverse sampling scheme in terms of the experimental cost. The variance of the estimator $\widetilde{\rho}_{m_1,m_2}$ coincides with the variance of the estimator $\widetilde{\rho}_{n_1,n_2}$, if $m_1 = n_1p_1$ and $m_2 = n_2p_2$. Therefore, *the direct-direct and inverse-inverse schemes are equivalent in the same sense regarding asymptotic precision of the cross-product ratio estimation*. The same conclusion is true for all pairs of sampling schemes with the corresponding substitution of $m$ by $np$.

For example, let us compare the characteristics of the logarithmic confidence interval for the direct-direct sampling scheme with sample sizes $n_1 = n_2 = 100$ with similar characteristics for different sampling schemes. Let the values of probabilities take the values $p_1 = 0.8$ and $p_2 = 0.4$. If we choose $n = 100, m = n_2p_2 = 40$ in the direct-inverse scheme, $m = n_1 \cdot p_1 = 80, n = n_2 = 100$ in the inverse-direct, and $m_1 = n_1p_1 = 80, m_2 = n_2p_2 = 40$ in the inverse-inverse sampling schemes, then all sampling schemes can be considered as equivalent" on average" with respect to the cost for observations.

We use this fact when comparing an estimator accuracy for different sampling schemes.

In the previous section, we provided the analysis of precision and reliability properties of three kinds of logarithmic confidence intervals for each of four possible combinations of sampling

**Table 11.** Coverage probability, width, and standard deviation for CI (5).

| | $m_2$ | 20 | | | 50 | | | 100 | | |
|---|---|---|---|---|---|---|---|---|---|---|
| $m_1$ | $p_1 \setminus p_2$ | 0.2 | 0.5 | 0.8 | 0.2 | 0.5 | 0.8 | 0.2 | 0.5 | 0.8 |
| | | 0.946 | 0.949 | 0.965 | 0.948 | 0.949 | 0.955 | 0.949 | 0.948 | 0.950 |
| | 0.2 | 1.513 | 0.440 | 0.166 | 1.239 | 0.338 | 0.110 | 1.138 | 0.300 | 0.089 |
| | | 0.574 | 0.172 | 0.069 | 0.415 | 0.117 | 0.041 | 0.364 | 0.098 | 0.031 |
| | | 0.949 | 0.951 | 0.968 | 0.953 | 0.953 | 0.957 | 0.953 | 0.953 | 0.954 |
| 20 | 0.5 | 7.345 | 2.067 | 0.738 | 6.367 | 1.694 | 0.517 | 6.015 | 1.559 | 0.439 |
| | | 4.068 | 1.172 | 0.412 | 3.534 | 0.985 | 0.293 | 3.249 | 0.871 | 0.243 |
| | | 0.935 | 0.941 | 0.958 | 0.934 | 0.935 | 0.942 | 0.935 | 0.933 | 0.933 |
| | 0.8 | 54.758 | 14.444 | 4.502 | 51.189 | 13.072 | 3.601 | 50.030 | 12.621 | 3.333 |
| | | 63.389 | 16.592 | 5.082 | 59.016 | 15.108 | 4.124 | 57.821 | 14.705 | 3.823 |
| | | 0.947 | 0.953 | 0.966 | 0.947 | 0.950 | 0.957 | 0.949 | 0.948 | 0.952 |
| | 0.2 | 1.227 | 0.375 | 0.151 | 0.907 | 0.261 | 0.093 | 0.780 | 0.215 | 0.070 |
| | | 0.352 | 0.112 | 0.049 | 0.209 | 0.062 | 0.025 | 0.161 | 0.046 | 0.016 |
| | | 0.948 | 0.952 | 0.968 | 0.950 | 0.949 | 0.958 | 0.950 | 0.951 | 0.952 |
| 50 | 0.5 | 5.406 | 1.613 | 0.630 | 4.226 | 1.178 | 0.402 | 3.779 | 1.012 | 0.313 |
| | | 1.816 | 0.558 | 0.228 | 1.251 | 0.357 | 0.127 | 1.077 | 0.289 | 0.093 |
| | | 0.952 | 0.956 | 0.972 | 0.955 | 0.954 | 0.958 | 0.954 | 0.954 | 0.957 |
| | 0.8 | 30.271 | 8.529 | 2.999 | 26.404 | 6.984 | 2.130 | 24.916 | 6.455 | 1.815 |
| | | 20.657 | 5.890 | 1.938 | 18.612 | 4.813 | 1.407 | 17.398 | 4.555 | 1.253 |
| | | 0.948 | 0.954 | 0.965 | 0.950 | 0.952 | 0.957 | 0.948 | 0.950 | 0.953 |
| | 0.2 | 1.124 | 0.352 | 0.147 | 0.779 | 0.232 | 0.088 | 0.630 | 0.180 | 0.064 |
| | | 0.288 | 0.093 | 0.042 | 0.147 | 0.046 | 0.020 | 0.101 | 0.030 | 0.012 |
| | | 0.948 | 0.954 | 0.966 | 0.949 | 0.950 | 0.957 | 0.949 | 0.949 | 0.953 |
| 100 | 0.5 | 4.756 | 1.467 | 0.599 | 3.436 | 1 | 0.366 | 2.905 | 0.807 | 0.271 |
| | | 1.327 | 0.421 | 0.185 | 0.762 | 0.230 | 0.092 | 0.587 | 0.168 | 0.061 |
| | | 0.949 | 0.954 | 0.969 | 0.951 | 0.952 | 0.958 | 0.953 | 0.953 | 0.955 |
| | 0.8 | 23.153 | 6.803 | 2.602 | 18.598 | 5.109 | 1.692 | 16.894 | 4.478 | 1.352 |
| | | 8.993 | 2.675 | 1.045 | 6.742 | 1.864 | 0.628 | 6.068 | 1.608 | 0.484 |

**Table 12.** Coverage probability, width, and standard deviation for CI (5) for small probabilities.

| | $m_2$ | 20 | | | 50 | | | 100 | | |
|---|---|---|---|---|---|---|---|---|---|---|
| $m_1$ | $p_1 \setminus p_2$ | 0.05 | 0.1 | 0.15 | 0.05 | 0.1 | 0.15 | 0.05 | 0.1 | 0.15 |
| | | 0.944 | 0.944 | 0.945 | 0.946 | 0.946 | 0.946 | 0.947 | 0.946 | 0.947 |
| | 0.05 | 1.362 | 0.655 | 0.419 | 1.119 | 0.533 | 0.339 | 1.026 | 0.489 | 0.310 |
| | | 0.463 | 0.223 | 0.143 | 0.324 | 0.155 | 0.099 | 0.277 | 0.133 | 0.084 |
| | | 0.943 | 0.944 | 0.944 | 0.947 | 0.947 | 0.946 | 0.948 | 0.948 | 0.947 |
| 20 | 0.1 | 2.930 | 1.408 | 0.901 | 2.412 | 1.154 | 0.733 | 2.229 | 1.062 | 0.671 |
| | | 1.030 | 0.496 | 0.319 | 0.726 | 0.351 | 0.224 | 0.634 | 0.302 | 0.193 |
| | | 0.944 | 0.945 | 0.944 | 0.948 | 0.946 | 0.946 | 0.947 | 0.949 | 0.947 |
| | 0.15 | 4.741 | 2.282 | 1.453 | 3.947 | 1.886 | 1.192 | 3.650 | 1.735 | 1.100 |
| | | 1.709 | 0.826 | 0.527 | 1.250 | 0.605 | 0.378 | 1.103 | 0.521 | 0.335 |
| | | 0.946 | 0.946 | 0.946 | 0.950 | 0.949 | 0.948 | 0.949 | 0.948 | 0.948 |
| | 0.05 | 1.114 | 0.540 | 0.348 | 0.826 | 0.397 | 0.254 | 0.711 | 0.340 | 0.216 |
| | | 0.303 | 0.147 | 0.095 | 0.172 | 0.083 | 0.053 | 0.130 | 0.062 | 0.040 |
| | | 0.944 | 0.946 | 0.946 | 0.948 | 0.949 | 0.947 | 0.948 | 0.949 | 0.949 |
| 50 | 0.1 | 2.378 | 1.150 | 0.740 | 1.772 | 0.852 | 0.544 | 1.533 | 0.732 | 0.466 |
| | | 0.659 | 0.317 | 0.206 | 0.379 | 0.183 | 0.118 | 0.291 | 0.139 | 0.089 |
| | | 0.945 | 0.945 | 0.947 | 0.946 | 0.946 | 0.948 | 0.948 | 0.950 | 0.948 |
| | 0.15 | 3.813 | 1.843 | 1.187 | 2.860 | 1.372 | 0.877 | 2.485 | 1.188 | 0.756 |
| | | 1.068 | 0.517 | 0.333 | 0.630 | 0.304 | 0.195 | 0.488 | 0.233 | 0.150 |
| | | 0.947 | 0.945 | 0.949 | 0.947 | 0.950 | 0.949 | 0.949 | 0.950 | 0.950 |
| | 0.05 | 1.027 | 0.498 | 0.322 | 0.711 | 0.343 | 0.221 | 0.577 | 0.277 | 0.177 |
| | | 0.255 | 0.125 | 0.080 | 0.126 | 0.061 | 0.040 | 0.085 | 0.041 | 0.026 |
| | | 0.947 | 0.946 | 0.948 | 0.949 | 0.948 | 0.948 | 0.949 | 0.949 | 0.949 |
| 100 | 0.1 | 2.174 | 1.055 | 0.682 | 1.514 | 0.731 | 0.469 | 1.235 | 0.593 | 0.379 |
| | | 0.543 | 0.265 | 0.171 | 0.273 | 0.133 | 0.085 | 0.185 | 0.089 | 0.057 |
| | | 0.946 | 0.945 | 0.947 | 0.948 | 0.947 | 0.950 | 0.949 | 0.949 | 0.949 |
| | 0.15 | 3.483 | 1.685 | 1.090 | 2.433 | 1.173 | 0.754 | 1.992 | 0.957 | 0.611 |
| | | 0.878 | 0.426 | 0.276 | 0.446 | 0.216 | 0.139 | 0.305 | 0.147 | 0.095 |

**Table 13.** Comparison of the coverage probability for all sample schemes and types of the confidence intervals.

| Sampling Scheme | Linear | Logarithmic |
|---|---|---|
| Direct-Direct | 0.909 | 0.95167 |
| Direct-Inverse | 0.912 | 0.95101 |
| Inverse-Direct | 0.914 | 0.95227 |
| Inverse-Inverse | 0.910 | 0.95086 |

schemes. Which sampling scheme is best in connection with the situation at hand remains to be answered.

In order to support our further recommendations in the best way, we constructed an additional summary table in which the confidence probabilities of linear and logarithmic confidence intervals for different sampling schemes are compared for the same sample sizes.

As mentioned above, the expected tendency of the confidence intervals to become more precise and accurate when sample sizes increases is observed. This is also true when success probability in each of two samples increasing.

Simulation results show that all sampling schemes have practically identical coverage probability for the logarithmic confidence interval. The best accuracy from the coverage probability point of view occurs with the inverse-inverse sampling scheme. The worst sampling scheme that possesses the smallest coverage probability appears to be the scheme of two independent samples where the first is obtained with the inverse sampling scheme and the second by direct.

To support these findings, we present a table of typical values of the coverage probability for all four sampling schemes and two types of confidence intervals (linear and logarithmic). We remind readers that we choose $n_1 = n_2 = 100, p_1 = 0.8$ and $p_2 = 0.4$ (Table 13).

## 9. Concluding remarks and further research

As we can see from the simulation results presented in Sungboonchoo et al.. (2021), in some cases the linear confidence intervals for the cross-product ratio coefficient have a confidence level lower than nominal. In this article we show that this deficiency may be resolved by considering the logarithmic confidence intervals. Consideration of accuracy and reliability properties of the point estimators for the cross product ratio is also an interesting problem.

## Acknowledgements

The authors are grateful to the reviewers for carefully reading the manuscript and for offering substantial comments and suggestions which enabled them to improve the presentation. The last listed author's research was funded by the subsidy allocated to Kazan Federal University for the state assignment in the sphere of scientific activities, project no. 1.13556.2019/13.1.

## ORCID

Andrei Volodin http://orcid.org/0000-0002-9771-846X

## References

Albert, J. H., and A. K. Gupta. 1983. Estimation in contingency tables using prior information. *Journal of the Royal Statistical Society. Series B (Methodological)* 45:60–69. doi:10.1111/j.2517-6161.1983.tb01231.x.

Amini, M., and M. Mahdizadeh. 2017. Nonparametric estimation of the entropy using a ranked set sample. *Communications in Statistics – Simulation and Computation* 46:6719–37. doi:10.1080/03610918.2016.1208229.

Anděl, J. 1973. On interactions in contingency tables. *Aplikace Matematiky* 18:99–109.

Baxter, P. D., and P. R. Marchant. 2010. The cross-product ratio in bivariate lognormal and gamma distributions, with an application to non-randomized trials. *Journal of Applied Statistics* 37 (4):529–36. doi:10.1080/02664760902744962.

Cornfield, J. 1956. A statistical problem arising from retrospective studies. In *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*. Vol. 4. Berkeley, CA: University of California Press,

Gart, J. J. 1962. On the combination of relative risks. *Biometrics* 18 (4):601–10. doi:10.2307/2527905.

Goodman, L. A. 1964. Simultaneous confidence limits for cross-product ratios in contingency tables. *Journal of the Royal Statistical Society. Series B (Methodological)* 26:86–102. doi:10.1111/j.2517-6161.1964.tb00543.x.

Holland, P. W., and Y. J. Wang. 1987. Dependence functions for Continuous bivariate densities. *Communications in Statistics – Theory and Methods* 16:863–76. doi:10.1080/03610928708829408.

Lee, L.-F. 1981. *Empirical Bayes estimators for the cross-product ratios for $2 \times 2$ contingency tables*. Ph.D.thesis, Virginia Polytechnic Institute and State University.

Lehmann, E. L. 1997. *Elements of large sample theory*. New York: Springer-Verlag.

Lin, C.-Y., and M.-C. Yang. 2008. Improved *p*-value tests for comparing two independent binomial proportions. *Communications in Statistics – Simulation and Computation* 38:78–91. doi:10.1080/03610910802417812.

Mahdizadeh, M., and N. R. Arghami. 2012. Quantile estimation using ranked set samples from a population with known mean. *Communications in Statistics – Simulation and Computation* 41:1872–81. doi:10.1080/03610918.2011.624236.

Martín Andrés, A., J. M. Tapia García, and F. Gayá Moreno. 2020. Two-tailed asymptotic inferences for the odds ratio in cross-sectional studies: Evaluation of fifteen old and new methods of inference. *Journal of Biopharmaceutical Statistics* 30:900–915.

McCann, M. H., and J. M. Tebbs. 2009. Simultaneous logit-based confidence intervals for odds ratios in $2 \times k$ classification tables with a fixed reference level. *Communications in Statistics – Simulation and Computation* 38:961–75.

Ngamkham, T. 2018. Confidence interval estimation for the ratio of binomial proportions and random numbers generation for some statistical models. Ph. D. thesis, University of Regina.

Ngamkham, T., A. Volodin, and I. Volodin. 2016. Confidence intervals for a ratio of binomial proportions based on direct and inverse sampling schemes. *Lobachevskii Journal of Mathematics* 37 (4):466–96. doi:10.1134/S1995080216040132.

Niebuhr, T., and M. Trabs. 2019. Profiting from correlations: Adjusted estimators for categorical data. *Applied Stochastic Models in Business and Industry* 35 (4):1090–102. doi:10.1002/asmb.2452.

Schaarschmidt, F., D. Gerhard, and C. Vogel. 2017. Simultaneous confidence intervals for comparisons of several multinomial samples. *Computational Statistics and Data Analysis* 106:65–76. doi:10.1016/j.csda.2016.09.004.

Sungboonchoo, C., T. Ngamkham, W. Panichkitkosolkul, and A. Volodin. 2021. Confidence estimation of the cross-product ratio of binomial proportions under different sampling schemes. *Lobachevskii Journal of Mathematics* 42 (2):1–16.

Wang, Y. J. 1987. The probability integrals of bivariate normal distributions: A contingency table approach. *Biometrika* 74:185–90. doi:10.1093/biomet/74.1.185.

Xu, J. L. 2012. A property of the generalized proportional hazards model. *Far East Journal of Theoretical Statisitcs* 41 (2):149–62.