Thailand Statistician
October 2020; 18(4): 403-419
http//:statassoc.or.th
Contributed paper

On Accuracy Properties of Point Estimators for the Ratio of Binomial Proportions

Thuntida Ngamkham*
Department of Mathematics and Statistics, University of Calgary, Calgary, Alberta, Canada.
*Corresponding author; e-mail: thuntida.ngamkham@ucalgary.ca

Received: 9 April 2019
Revised: 22 June 2019
Accepted: 11 August 2019

Abstract

We continue our investigation on the ratio of binomial proportions started in Ngamkham et al. (2016) and Ngamkham (2018). Contrary to our previous research, where we were investigating interval estimations, here we concentrate on point estimation and its accuracy properties. A general problem of the point estimation for a ratio of two proportions according to data from two independent samples is considered. Each sample may be obtained in the framework of direct or inverse binomial sampling. Our goal is to show that the normal approximations (which are relatively simple) for estimates of the ratio are reliable for construction of point estimators. The main criterion of our judgment is the bias and mean squared error. It is shown by statistically modeled data that the scheme of inverse binomial sampling with planning of the size in the second sample is preferred (so-called special case of the direct-inverse sampling scheme). The main accuracy characteristics of estimators corresponding to all possible combinations of sampling schemes are investigated by the Monte Carlo method. Mean values and mean squared errors of point estimators are collected in tables, and some recommendations for an application of the estimators are presented.

Keywords: Inverse binomial sampling, direct binomial sampling, normal asymptotic of an estimator, mean squared error, bias.

1. Introduction

The problem of comparison of Bernoulli trials success probabilities is a topic in biological and medical investigations. In this article, we identify the sample scheme that provides the best accuracy for the point estimation for the ratio of probabilities. The problem of comparison of Bernoulli trials success probabilities is a topic in biological and medical investigations. In this article, we identify the sample scheme that provides the best accuracy for the point estimation for the ratio of probabilities.

The initial objective of the investigation resulting in the present article was only to justify an application of the delta-method and corresponding normal approximations of probability ratio estimates. We were not planning to construct estimators that have better properties than the estimatros presented in the papers of our predecessors. Moreover, we found that only direct-direct and inverse-inverse sampling schemes (defintions are presented below) were considered before. In this article we introduce direct-inverse and inverse-direct schemes. One of the main achivements is that we were able to construct estimators for different sampling schemes by considering the situation when the stopping moment for observations in the second sample depends on the result of observation in the first one, so-called special case of the direct-inverse sampling scheme. Moreover, it appears that estimator formulae that we discuss in this article are much simpler for practical applications than the formulae presented in some of our predecessor's papers. We can achieve the desired precision and reliability of estimators without solving complicated systems of nonlinear equations.

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 ν_1 and ν_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 estimation of the ratio $\theta = p_1 / p_2$. Note that this ratio is usually called odd ratio in biological, medical, and pharmaceutical investigations.

Each sample may be obtained in the framework of direct or inverse binomial sampling schemes; see Figure 1.

Inverse binomial sampling: In this scheme a Bernoulli sequence $Y^{(\nu)} = (Y_1, \ldots, Y_\nu)$ is observed with a stopping time $\nu = \min\{n : \sum_{k=1}^{n} Y_k \geq m\}$, where m is a fixed number of successes.

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.

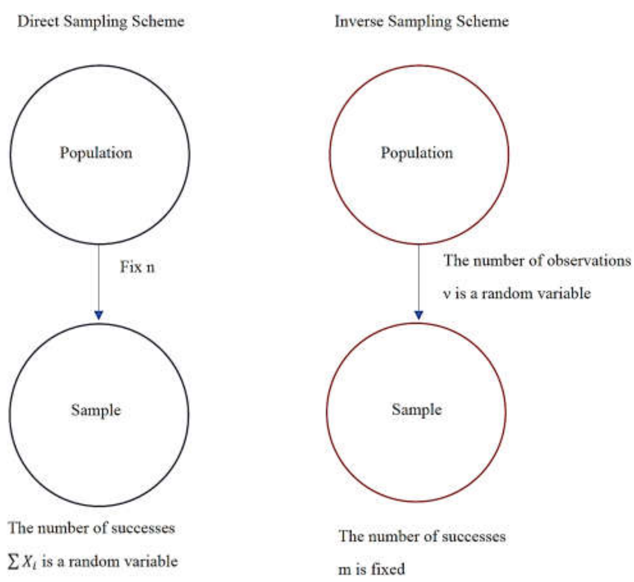

Figure 1 Direct and inverse sampling schemes

In the following, we will 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. In our previous investigations Ngamkham et al. (2016) and Chapters 1 and 2 of Ngamkham

(2018), we concentrated most on confidence estimation for the parameter θ. In this article, we are interested most in point estimators. We refer to the aforementioned references for a detailed literature review pertaining to the estimation of θ. Of note, the first easy-to-calculate methods of θ estimation were suggested by Nether (1957) and Guttman (1958). A survey of these early methods can be found in Sheps (1959). Advantages of unbiased estimators with the uniformly minimal risk are shown in articles by Bennett (1981), Roberts (1993), and Lui (1996).

In this article, we consider only estimators developed previously in Ngamkham et al. (2016) and Chapter 2 Ngamkham (2018). The results for inverse and direct sampling schemes are collected in tables. In each cell of the tables, the following characteristics are presented: the true values of the parameter θ, the variance of the estimator, and its quadratic risk. For each value we generated 10^4 random numbers with the Bernoulli and/or negative binomial (Pascal) distribution with various values of parameters (success probabilities) p_1 and p_2.

We consider the problem of estimation of the probabilities ratio $\theta = p_1 / p_2$ for five schemes of Bernoulli trials: direct- direct, direct- inverse, inverse- direct, inverse- inverse and the special case of the direct-inverse (when n_1 is given, but ν_2 is found by the number of success from the first sample).

2. Point Estimator for the Ratio of Binomial Proportions

The material presented in this section can be found in Ngamkham et al. (2016) and Chapter 2 of Ngamkham (2018). We present it here to fix the notation, and make the article more self-contained.

For a solution of the problems stated in the introduction, it is necessary to construct estimates, preferably unbiased, for the ratio of binomial proportions $\theta = p_1 / p_2$ for all possible combinations of two schemes of Bernoulli trials.

In the following we use the following notation. Direct binomial sampling: a random vector $X^{(n)} = (X_1, \ldots, X_n)$ with Bernoulli components and fixed number of observations n is observed. In the case of direct binomial sampling, we use the statistic $\overline{X}_n = T / n,$ where $T = \sum_{k=1}^{n} X_k$. Inverse binomial sampling: a Bernoulli sequence $Y^{(\nu)} = (Y_1, \ldots, Y_\nu)$ is observed with a stopping time

$$\nu = \min\{n : \sum_{k=1}^{n} Y_k \geq m\}.$$

That is, the components of the sequence $Y_1, Y_2, \ldots$ are observed until the given number m of successes will appear. In the case of inverse binomial sampling we use the statistic $\overline{Y}_m = \nu / m.$

To estimate the ratio of binomial proportions $\theta = p_1 / p_2,$ we need to consider two samples. The following are all possible combinations of the sampling schemes we can apply.

Direct-direct $D(n_1, p_1) - D(n_2, p_2)$: Both samples are obtained in the scheme of direct sampling with probabilities p_1 and p_2 of successes and sample sizes n_1 and n_2, respectively.

Direct-inverse $D(n, p_1) - I(m, p_2)$: The first sample is obtained by the scheme of direct binomial sampling with a probability p_1 of a success and fixed sample size n, while the second sample is obtained by the scheme of inverse binomial sampling with the probability p_2 and stopping time which is defined by the fixed number m of successes in the sample.

Inverse-direct $I(m, p_1) - D(n, p_2)$: The first sample obtained by the scheme of inverse binomial sampling with parameters (p_1, m), and the second sample obtained by the scheme of direct sampling with parameters (n, p_2).

Inverse-inverse $I(m_1, p_1) - I(m_2, p_2)$: In this case both samples are obtained in the scheme of inverse binomial sampling with parameters (p_1, m_1) and (p_2, m_2), respectively.

In Table 1, we present all estimators of the ratio of two proportions $\theta = p_1 / p_2$ for these four possible sampling schemes, which were derived in Ngamkham et al. (2016) and Chapter 2 of Ngamkham (2018).

Table 1 Estimators of the ratio of two proportions $\theta = p_1 / p_2$ and their approximations for all possible combinations of direct and inverse sampling schemes

	Second Sample Direct	Second Sample Inverse
First Sample Direct	$\hat{\theta}_{n_1,n_2} = \dfrac{\bar{X}_{n_1}(n_2+1)}{n_2\bar{X}_{n_2}+1}$ $\approx \tilde{\theta}_{n_1,n_2} = \bar{X}_{n_1} / \bar{X}_{n_2}$	$\hat{\theta}_{n,m} = \bar{X}_n \bar{Y}_m$
First Sample Inverse	$\hat{\theta}_{m,n} = \dfrac{(m-1)(n+1)}{(m\bar{Y}_m - 1)(n\bar{X}_n + 1)}$ $\approx \tilde{\theta}_{m,n} = [\bar{Y}_m \bar{X}_n]^{-1}$	$\hat{\theta}_{m_1,m_2} = \dfrac{(m_1-1)\bar{Y}_{m_2}}{m_1\bar{Y}_{m_1} - 1}$ $\approx \tilde{\theta}_{m_1,m_2} = \bar{Y}_{m_2} / \bar{Y}_{m_1}$

For large values of n and m, all four estimates of probability ratio θ are continuous functions of statistics $\bar{X}_n$ and $\bar{Y}_m$ with finite second moments; therefore the estimates are asymptotically normal. In Ngamkham et al. (2016) and Chapter 2 of Ngamkham (2018), we found the asymptotic of the mean and variance of these estimates, for which we explore the standard delta method. The method is based on a Taylor series expansion in the neighborhoods of the mean values of the statistics $\bar{X}_n$ and $\bar{Y}_m$. The results are summarized in Table 2, which provides both the values of means and variances for normal approximation to the statistics $\tilde{\theta}$ for all possible combinations of direct and inverse sample schemes, where statistics $\tilde{\theta}$ are the stochastic approximations to the corresponding statistics $\hat{\theta}$.

We also consider the interesting special case of the direct-inverse sampling scheme. As we will see in the following, this method performs the best.

For special case of the direct-inverse $D(n, p_1) - I(\nu, p_2)$, the first sample is obtained by the direct sampling method with parameters (n, p_1). The (random) sample size for the second sample that depends on m and the following sampling plan for the second sample can be suggested: Repeat observations until reaching the same number of successes as in the first experiment, that is, set $m = T = \sum_{k=1}^{n} X_k$. In this case the new suggested estimate of the parameter θ is

$$\hat{\theta}_n = \bar{X}_n \bar{Y}_T = \frac{\nu_T}{n}.$$

The estimate $\hat{\theta}_n$ is asymptotically normal with mean θ and variance $\dfrac{\theta(2p_2^{-1}-\theta-1)}{n}.$

Remark: For the schemes of inverse binomial sampling with parameters (p,m) the mean sample size is $E(\nu)=m/p.$ If the observations are obtained in the scheme of direct sampling with the same probability p of the success and sample size $n=m/p,$ then "on average" it is equivalent to the scheme of inverse sampling from the point of view of the cost for the experiment. Variance of the estimate $\hat{\theta}_{m_1,m_2}$ coincides with variance of the estimate $\hat{\theta}_{n_1,n_2}$, if $m_1=n_1p_1$ and $m_2=n_2p_2$. Therefore, schemes direct-direct and inverse-inverse are equivalent in the same sense from the same point of view of asymptotic precision of the estimates for probability ratio. Of course, the same conclusion is true for all pairs of sampling schemes with the corresponding substitution of m by $np.$ We use this remark in our simulations below for calculations of sample sizes.

Table 2 Asymptotic representations, means, and variances for normal approximation to the statistics $\tilde{\theta}$ for all possible combinations of direct and inverse sampling schemes

	Second Sample Direct	Second Sample Inverse
First Sample Direct	$E(\tilde{\theta}_{n_1,n_2})=\theta$ $Var(\tilde{\theta}_{n_1,n_2})=\theta^2\left[\dfrac{1-p_1}{n_1p_1}+\dfrac{1-p_2}{n_2p_2}\right]$	$E(\tilde{\theta}_{n,m})=\theta$ $Var(\tilde{\theta}_{n,m})=\theta^2\left(\dfrac{1-p_1}{np_1}+\dfrac{1-p_2}{m}\right)$
First Sample Inverse	$E(\tilde{\theta}_{m,n})=\theta$ $Var(\tilde{\theta}_{m,n})=\theta^2\left(\dfrac{1-p_1}{m}+\dfrac{1-p_2}{np_2}\right)$	$E(\tilde{\theta}_{m_1,m_2})=\theta$ $Var(\tilde{\theta}_{m_1,m_2})=\theta^2\left(\dfrac{1-p_1}{m_1}+\dfrac{1-p_2}{m_2}\right)$

3. Numerical Modeling of the Accuracy Properties of the Estimators

For each sampling scheme, 10^4 simulations of the corresponding estimators have been performed. According to the results of the simulations, the mean value and mean squared errors (MSE) were calculated. Simulations are done with the help of the Language and Environment for Statistical Computing R Version 3.5.3 (R Core Team 2019). Tables that contain the values of theoretical variances and MSE were constructed.

3.1. Modeling of the bias of estimators

Tables 3 and 4 are devoted to the simulations results for the direct-direct sampling scheme $D(n_1,p_1)-D(n_2,p_2).$ In each cell of the tables the following characteristics are presented: mean value of estimator $\hat{\theta}$ and the true value of the parameter $\theta=p_1/p_2$.

According to the results presented in Tables 3 and 4, we can conclude that when the sample size is increasing, the mean value of the estimate approaches the true value of the probability ratio. When the probabilities are increasing, the mean value of estimator has faster rate of convergence to the desired parameter $\theta=p_1/p_2$.

Table 3 Mean value of estimator for the direct-direct scheme $D(n_1, p_1) - D(n_2, p_2)$, sample sizes $n_1 = 50,\ n_2 = 50$

	n_2	50									
n_1	p_1/p_2	0.05	0.10	0.20	0.30	0.40	0.50	0.60	0.70	0.80	0.90
50	0.05	0.9464	0.4957	0.2502	0.1676	0.1244	0.1012	0.0830	0.0713	0.0627	0.0561
		1.0000	0.5000	0.2500	0.1677	0.1250	0.1000	0.0833	0.0714	0.0625	0.0556
	0.10	1.8507	0.9955	0.5024	0.3321	0.2508	0.1993	0.1668	0.1421	0.1255	0.1115
		2.0000	1.0000	0.5000	0.3333	0.2500	0.2000	0.1667	0.1429	0.1250	0.1111
	0.20	3.7161	2.0017	1.0030	0.6667	0.5017	0.3985	0.3328	0.2844	0.2495	0.2217
		4.0000	2.0000	1.0000	0.6667	0.5000	0.4000	0.3333	0.2857	0.2500	0.2222
	0.30	5.6089	2.9776	1.5056	0.9997	0.7504	0.6007	0.5010	0.4278	0.3747	0.3349
		6.0000	3.0000	1.5000	1.0000	0.7500	0.6000	0.5000	0.4286	0.3750	0.3333
	0.40	7.3582	3.9871	1.9919	1.3390	0.9994	0.8001	0.6657	0.5729	0.5004	0.4453
		8.0000	4.0000	2.0000	1.3333	1.0000	0.8000	0.6667	0.5714	0.5000	0.4444
	0.50	9.2539	4.9860	2.5042	1.6639	1.2549	1.0016	0.8319	0.7142	0.6271	0.5546
		10.0000	5.0000	2.5000	1.6667	1.2500	1.0000	0.8333	0.7143	0.6250	0.5556
	0.60	11.0323	5.9905	3.0091	1.9888	1.4989	1.2014	1.0004	0.8543	0.7498	0.6663
		12.0000	6.0000	3.0000	2.0000	1.5000	1.2000	1.0000	0.8571	0.7500	0.6667
	0.70	12.9655	6.9498	3.4742	2.3368	1.7526	1.3981	1.1656	0.9997	0.8754	0.7772
		14.0000	7.0000	3.5000	2.3333	1.7500	1.4000	1.1667	1.0000	0.8750	0.7778
	0.80	14.8349	7.9547	3.9930	2.6614	2.0060	1.5981	1.3335	1.1436	1.0017	0.8889
		16.0000	8.0000	4.0000	2.6667	2.0000	1.6000	1.3333	1.1429	1.0000	0.8889
	0.90	16.8723	8.9609	4.4910	2.9942	2.2444	1.7997	1.5015	1.2862	1.1270	0.9996
		18.0000	9.0000	4.5000	3.0000	2.2500	1.8000	1.5000	1.2857	1.1250	1.0000

Table 4 Mean value of estimator for the direct-direct scheme $D(n_1, p_1) - D(n_2, p_2)$, sample sizes $n_1 = 200,\ n_2 = 200$

	n_2	200									
n_1	p_1/p_2	0.05	0.10	0.20	0.30	0.40	0.50	0.60	0.70	0.80	0.90
200	0.05	1.0045	0.5034	0.2503	0.1671	0.1248	0.0998	0.0836	0.0711	0.0626	0.0554
		1.0000	0.5000	0.2500	0.1677	0.1250	0.1000	0.0833	0.0714	0.0625	0.0556
	0.10	1.9986	1.0022	0.5014	0.3323	0.2507	0.1994	0.1668	0.1429	0.1254	0.1113
		2.0000	1.0000	0.5000	0.3333	0.2500	0.2000	0.1667	0.1429	0.1250	0.1111
	0.20	4.0006	2.0066	1.0020	0.6659	0.5000	0.4005	0.3326	0.2853	0.2506	0.2226
		4.0000	2.0000	1.0000	0.6667	0.5000	0.4000	0.3333	0.2857	0.2500	0.2222
	0.30	6.0111	2.9949	1.4984	0.9998	0.7504	0.5988	0.4998	0.4286	0.3748	0.3335
		6.0000	3.0000	1.5000	1.0000	0.7500	0.6000	0.5000	0.4286	0.3750	0.3333
	0.40	8.0126	4.0114	1.9989	1.3330	1.0005	0.7997	0.6668	0.5712	0.5005	0.4441
		8.0000	4.0000	2.0000	1.3333	1.0000	0.8000	0.6667	0.5714	0.5000	0.4444
	0.50	9.9168	5.0194	2.4998	1.6688	1.2507	1.0003	0.8321	0.7144	0.6243	0.5551
		10.0000	5.0000	2.5000	1.6667	1.2500	1.0000	0.8333	0.7143	0.6250	0.5556
	0.60	12.0218	6.0089	3.0030	2.0021	1.4996	1.1999	0.9994	0.8566	0.7500	0.6667
		12.0000	6.0000	3.0000	2.0000	1.5000	1.2000	1.0000	0.8571	0.7500	0.6667
	0.70	14.0134	7.0006	3.4972	2.3337	1.7506	1.4001	1.1653	1.0003	0.8745	0.7780
		14.0000	7.0000	3.5000	2.3333	1.7500	1.4000	1.1667	1.0000	0.8750	0.7778
	0.80	15.9870	7.9872	4.0165	2.6667	2.0030	1.6002	1.3352	1.1425	1.0000	0.8891
		16.0000	8.0000	4.0000	2.6667	2.0000	1.6000	1.3333	1.1429	1.0000	0.8889
	0.90	17.9400	8.9644	4.4982	3.0001	2.2472	1.8018	1.4994	1.2860	1.1244	1.0003
		18.0000	9.0000	4.5000	3.0000	2.2500	1.8000	1.5000	1.2857	1.1250	1.0000

Tables 5 and 6 are devoted to the simulations results for the direct-inverse scheme $D(n, p_1) - I(m, p_2)$. In each cell of the tables, the following characteristics are presented: mean value of estimator $\hat{\theta}$ and the true value of the parameter $\theta = p_1 / p_2$.

Table 5 Mean value of estimator for the direct-inverse scheme $D(n, p_1) - I(m, p_2)$, sample size of the first sample, number of success in the second sample $n_1 = 50$, $m_2 = n_2 p_2$

	n_2	50									
n_1	p_1/p_2	0.05	0.10	0.20	0.30	0.40	0.50	0.60	0.70	0.80	0.90
50	0.05	0.8047	0.5012	0.2515	0.1669	0.1244	0.1007	0.0846	0.0717	0.0628	0.0550
		1.0000	0.5000	0.2500	0.1677	0.1250	0.1000	0.0833	0.0714	0.0625	0.0556
	0.10	1.6004	0.9974	0.4973	0.3338	0.2513	0.1994	0.1680	0.1434	0.1255	0.1107
		2.0000	1.0000	0.5000	0.3333	0.2500	0.2000	0.1667	0.1429	0.1250	0.1111
	0.20	3.2072	1.9887	1.0083	0.6665	0.5009	0.4005	0.3347	0.2854	0.2498	0.2206
		4.0000	2.0000	1.0000	0.6667	0.5000	0.4000	0.3333	0.2857	0.2500	0.2222
	0.30	4.7548	2.9919	1.4962	1.0003	0.7506	0.5972	0.5021	0.4274	0.3751	0.3334
		6.0000	3.0000	1.5000	1.0000	0.7500	0.6000	0.5000	0.4286	0.3750	0.3333
	0.40	6.4434	4.0329	2.0060	1.3290	0.9982	0.7976	0.6675	0.5713	0.5009	0.4451
		8.0000	4.0000	2.0000	1.3333	1.0000	0.8000	0.6667	0.5714	0.5000	0.4444
	0.50	8.0828	5.0295	2.4999	1.6631	1.2541	0.9981	0.8325	0.7151	0.6254	0.5574
		10.0000	5.0000	2.5000	1.6667	1.2500	1.0000	0.8333	0.7143	0.6250	0.5556
	0.60	9.5788	6.0046	3.0200	2.0046	1.4979	1.2017	1.0033	0.8572	0.7497	0.6678
		12.0000	6.0000	3.0000	2.0000	1.5000	1.2000	1.0000	0.8571	0.7500	0.6667

Table 6 Mean value of estimator for the direct-inverse scheme $D(n, p_1) - I(m, p_2)$, sample size of the first sample, number of success in the second sample $n_1 = 100$, $m_2 = n_2 p_2$

	n_2	100									
n_1	p_1/p_2	0.05	0.10	0.20	0.30	0.40	0.50	0.60	0.70	0.80	0.90
100	0.05	1.0053	0.4987	0.2490	0.1673	0.1249	0.0995	0.0833	0.0717	0.0627	0.0554
		1.0000	0.5000	0.2500	0.1677	0.1250	0.1000	0.0833	0.0714	0.0625	0.0556
	0.10	2.0083	0.9943	0.4992	0.3336	0.2510	0.2008	0.1675	0.1431	0.1242	0.1118
		2.0000	1.0000	0.5000	0.3333	0.2500	0.2000	0.1667	0.1429	0.1250	0.1111
	0.20	4.0038	2.0039	1.0008	0.6697	0.5023	0.3994	0.3326	0.2869	0.2503	0.2228
		4.0000	2.0000	1.0000	0.6667	0.5000	0.4000	0.3333	0.2857	0.2500	0.2222
	0.30	5.9835	2.9991	1.5051	0.9986	0.7462	0.6002	0.4987	0.4270	0.3757	0.3338
		6.0000	3.0000	1.5000	1.0000	0.7500	0.6000	0.5000	0.4286	0.3750	0.3333
	0.40	8.0104	3.9864	1.9972	1.3319	1.0008	0.8015	0.6663	0.5712	0.4997	0.4449
		8.0000	4.0000	2.0000	1.3333	1.0000	0.8000	0.6667	0.5714	0.5000	0.4444
	0.50	10.0043	5.0254	2.5112	1.6621	1.2516	0.9966	0.8329	0.7153	0.6255	0.5553
		10.0000	5.0000	2.5000	1.6667	1.2500	1.0000	0.8333	0.7143	0.6250	0.5556
	0.60	11.9224	6.0105	2.9990	1.9993	1.4984	1.2009	0.9998	0.8581	0.7508	0.6670
		12.0000	6.0000	3.0000	2.0000	1.5000	1.2000	1.0000	0.8571	0.7500	0.6667
	0.70	14.0663	6.9856	3.4988	2.3406	1.7538	1.4011	1.1667	0.9993	0.8745	0.7787
		14.0000	7.0000	3.5000	2.3333	1.7500	1.4000	1.1667	1.0000	0.8750	0.7778
	0.80	15.9258	7.9881	3.9776	2.6614	2.0006	1.6028	1.3345	1.1426	0.9993	0.8895
		16.0000	8.0000	4.0000	2.6667	2.0000	1.6000	1.3333	1.1429	1.0000	0.8889
	0.90	17.9423	9.0322	4.4931	2.9917	2.2501	1.8000	1.4977	1.2869	1.1250	0.9996
		18.0000	9.0000	4.5000	3.0000	2.2500	1.8000	1.5000	1.2857	1.1250	1.0000

According to the results presented in Tables 5 and 6, we can conclude that when the sample size is increasing, the mean value of the estimate approaches the true value of the probability ratio. When the probabilities are increasing, the mean value of estimator has faster rate of convergence to the desired parameter $\theta = p_1 / p_2$.

Tables 7 and 8 are devoted to the simulations results for the inverse-direct sampling scheme $I(m, p_1) - D(n, p_2)$. In each cell of the table, the following characteristics are presented: mean value of estimator $\hat{\theta}$ and the true value of $\theta = p_1 / p_2$.

Table 7 Mean value of the estimator for the inverse-direct scheme $I(m, p_1) - D(n, p_2)$, number of successes in the first sample, size of the second sample $m_1 = n_1 p_1$, $n_2 = 50$

n_2		50									
n_1	p_1/p_2	0.05	0.10	0.20	0.30	0.40	0.50	0.60	0.70	0.80	0.90
	0.05	0.2140	0.2742	0.2744	0.2260	0.1826	0.1485	0.1206	0.1054	0.0938	0.0824
		1.0000	0.5000	0.2500	0.1677	0.1250	0.1000	0.0833	0.0714	0.0625	0.0556
	0.10	0.2864	0.3695	0.3617	0.2941	0.2357	0.1968	0.1634	0.1422	0.1246	0.1099
		2.0000	1.0000	0.5000	0.3333	0.2500	0.2000	0.1667	0.1429	0.1250	0.1111
	0.20	0.5743	0.7358	0.7256	0.5916	0.4756	0.3921	0.3303	0.2850	0.2503	0.2213
		4.0000	2.0000	1.0000	0.6667	0.5000	0.4000	0.3333	0.2857	0.2500	0.2222
	0.30	0.8607	1.1123	1.0837	0.8832	0.7120	0.5856	0.4943	0.4273	0.3744	0.3330
		6.0000	3.0000	1.5000	1.0000	0.7500	0.6000	0.5000	0.4286	0.3750	0.3333
50	0.40	1.1381	1.4747	1.4471	1.1778	0.9502	0.7811	0.6614	0.5699	0.4986	0.4439
		8.0000	4.0000	2.0000	1.3333	1.0000	0.8000	0.6667	0.5714	0.5000	0.4444
	0.50	1.4340	1.8500	1.8126	1.4726	1.1831	0.9772	0.8247	0.7120	0.6229	0.5542
		10.0000	5.0000	2.5000	1.6667	1.2500	1.0000	0.8333	0.7143	0.6250	0.5556
	0.60	1.7240	2.2287	2.1638	1.7617	1.4281	1.1759	0.9915	0.8531	0.7505	0.6670
		12.0000	6.0000	3.0000	2.0000	1.5000	1.2000	1.0000	0.8571	0.7500	0.6667
	0.70	2.0139	2.5962	2.5244	2.0632	1.6561	1.3675	1.1533	0.9948	0.8732	0.7779
		14.0000	7.0000	3.5000	2.3333	1.7500	1.4000	1.1667	1.0000	0.8750	0.7778
	0.80	2.3045	2.9583	2.8890	2.3618	1.9012	1.5614	1.3194	1.1372	0.9980	0.8888
		16.0000	8.0000	4.0000	2.6667	2.0000	1.6000	1.3333	1.1429	1.0000	0.8889

Table 8 Mean value of the estimator for the inverse-direct scheme $I(m, p_1) - D(n, p_2)$, number of successes in the first sample, size of the second sample $m_1 = n_1 p_1$, $n_2 = 100$

n_2		100									
n_1	p_1/p_2	0.05	0.10	0.20	0.30	0.40	0.50	0.60	0.70	0.80	0.90
	0.05	0.2231	0.2674	0.2101	0.1588	0.1232	0.0994	0.0833	0.0716	0.0625	0.0553
		1.0000	0.5000	0.2500	0.1677	0.1250	0.1000	0.0833	0.0714	0.0625	0.0556
	0.10	0.4481	0.5276	0.4259	0.3149	0.2444	0.1988	0.1666	0.1424	0.1245	0.1109
		2.0000	1.0000	0.5000	0.3333	0.2500	0.2000	0.1667	0.1429	0.1250	0.1111
	0.20	0.9016	1.0599	0.8465	0.6283	0.4879	0.3958	0.3318	0.2858	0.2498	0.2223
		4.0000	2.0000	1.0000	0.6667	0.5000	0.4000	0.3333	0.2857	0.2500	0.2222
100	0.30	1.3446	1.5823	1.2696	0.9453	0.7321	0.5935	0.4980	0.4286	0.3744	0.3329
		6.0000	3.0000	1.5000	1.0000	0.7500	0.6000	0.5000	0.4286	0.3750	0.3333
	0.40	1.7963	2.1159	1.6937	1.2561	0.9722	0.7905	0.6628	0.5691	0.4999	0.4446
		8.0000	4.0000	2.0000	1.3333	1.0000	0.8000	0.6667	0.5714	0.5000	0.4444
	0.50	2.2384	2.6456	2.1189	1.5687	1.2214	0.9877	0.8311	0.7123	0.6249	0.5551
		10.0000	5.0000	2.5000	1.6667	1.2500	1.0000	0.8333	0.7143	0.6250	0.5556

Table 8 (Continued)

	n_2					100					
n_1	p_1/p_2	0.05	0.10	0.20	0.30	0.40	0.50	0.60	0.70	0.80	0.90
100	0.60	2.6943	3.1733	2.5391	1.8936	1.4642	1.1867	0.9984	0.8559	0.7494	0.6661
		12.0000	6.0000	3.0000	2.0000	1.5000	1.2000	1.0000	0.8571	0.7500	0.6667
	0.70	3.1404	3.6974	2.9586	2.1975	1.7070	1.3818	1.1617	0.9975	0.8739	0.7779
		14.0000	7.0000	3.5000	2.3333	1.7500	1.4000	1.1667	1.0000	0.8750	0.7778
	0.80	3.5891	4.2309	3.3864	2.5150	1.9498	1.5828	1.3270	1.1403	0.9983	0.8894
		16.0000	8.0000	4.0000	2.6667	2.0000	1.6000	1.3333	1.1429	1.0000	0.8889
	0.90	4.0543	4.7592	3.8093	2.8346	2.1981	1.7806	1.4915	1.2838	1.1237	0.9994
		18.0000	9.0000	4.5000	3.0000	2.2500	1.8000	1.5000	1.2857	1.1250	1.0000

According to the results presented in Tables 7 and 8, we can conclude that when the sample size is increasing, the mean value of the estimate approaches the true value of the probability ratio. When the probabilities are increasing, the mean value of estimator has bigger rate of convergence to the desired parameter $\theta = p_1 / p_2$.

Tables 9 and 10 are devoted to the simulations results for the inverse-inverse sampling scheme $I(m_1, p_1) - I(m_2, p_2)$. In each cell of the tables, the following characteristics are presented: mean value of estimator $\hat{\theta}$ and the true value of the parameter $\theta = p_1 / p_2$.

Table 9 Mean value of the estimator for the inverse-inverse scheme $I(m_1, p_1) - I(m_2, p_2)$, numbers of successes in the first and second samples $m_1 = n_1 p_1$, $m_2 = n_2 p_2$

	n_2					50					
n_1	p_1/p_2	0.05	0.10	0.20	0.30	0.40	0.50	0.60	0.70	0.80	0.90
50	0.05	1.1944	0.7506	0.3740	0.2504	0.1881	0.1482	0.1221	0.1090	0.0943	0.0820
		1.0000	0.5000	0.2500	0.1677	0.1250	0.1000	0.0833	0.0714	0.0625	0.0556
	0.10	1.6131	1.0111	0.5029	0.3357	0.2494	0.2007	0.1667	0.1426	0.1255	0.1114
		2.0000	1.0000	0.5000	0.3333	0.2500	0.2000	0.1667	0.1429	0.1250	0.1111
	0.20	3.2035	2.0040	1.0001	0.6695	0.4998	0.4006	0.3331	0.2859	0.2502	0.2208
		4.0000	2.0000	1.0000	0.6667	0.5000	0.4000	0.3333	0.2857	0.2500	0.2222
	0.30	4.7877	2.9813	1.5067	0.9980	0.7455	0.5989	0.5013	0.4297	0.3725	0.3344
		6.0000	3.0000	1.5000	1.0000	0.7500	0.6000	0.5000	0.4286	0.3750	0.3333
	0.40	6.4675	3.9863	1.9916	1.3332	1.0036	0.8021	0.6644	0.5712	0.4999	0.4438
		8.0000	4.0000	2.0000	1.3333	1.0000	0.8000	0.6667	0.5714	0.5000	0.4444
	0.50	7.9667	5.0221	2.5015	1.6634	1.2541	1.0025	0.8347	0.7155	0.6247	0.5555
		10.0000	5.0000	2.5000	1.6667	1.2500	1.0000	0.8333	0.7143	0.6250	0.5556
	0.60	9.6852	6.0001	3.0119	2.0000	1.5037	1.2019	0.9981	0.8575	0.7513	0.6670
		12.0000	6.0000	3.0000	2.0000	1.5000	1.2000	1.0000	0.8571	0.7500	0.6667
	0.70	11.2849	7.0040	3.5037	2.3336	1.7442	1.4019	1.1668	1.0009	0.8752	0.7782
		14.0000	7.0000	3.5000	2.3333	1.7500	1.4000	1.1667	1.0000	0.8750	0.7778
	0.80	12.7836	7.9859	4.0092	2.6639	2.0006	1.6033	1.3306	1.1436	1.0005	0.8890
		16.0000	8.0000	4.0000	2.6667	2.0000	1.6000	1.3333	1.1429	1.0000	0.8889
	0.90	14.3163	9.0417	4.4810	3.0014	2.2516	1.7984	1.4973	1.2844	1.1255	0.9998
		18.0000	9.0000	4.5000	3.0000	2.2500	1.8000	1.5000	1.2857	1.1250	1.0000

Table 10 Mean value of the estimator for the inverse-inverse scheme $I(m_1, p_1) - I(m_2, p_2)$, numbers of successes in the first and second samples $m_1 = n_1 p_1$, $m_2 = n_2 p_2$

	n_2	200									
n_1	p_1/p_2	0.05	0.10	0.20	0.30	0.40	0.50	0.60	0.70	0.80	0.90
	0.05	0.9987	0.4982	0.2497	0.1665	0.1247	0.0997	0.0834	0.0719	0.0624	0.0556
		1.0000	0.5000	0.2500	0.1677	0.1250	0.1000	0.0833	0.0714	0.0625	0.0556
	0.10	1.9985	1.0011	0.5020	0.3340	0.2511	0.2011	0.1659	0.1425	0.1246	0.1112
		2.0000	1.0000	0.5000	0.3333	0.2500	0.2000	0.1667	0.1429	0.1250	0.1111
	0.20	3.9962	1.9989	1.0001	0.6662	0.4999	0.4002	0.3327	0.2859	0.2503	0.2221
		4.0000	2.0000	1.0000	0.6667	0.5000	0.4000	0.3333	0.2857	0.2500	0.2222
	0.30	6.0164	3.0054	1.4991	0.9985	0.7500	0.5997	0.4996	0.4283	0.3746	0.3331
		6.0000	3.0000	1.5000	1.0000	0.7500	0.6000	0.5000	0.4286	0.3750	0.3333
	0.40	8.0307	4.0158	2.0026	1.3289	0.9989	0.8000	0.6670	0.5710	0.4995	0.4446
		8.0000	4.0000	2.0000	1.3333	1.0000	0.8000	0.6667	0.5714	0.5000	0.4444
200	0.50	9.9780	4.9912	2.5072	1.6709	1.2472	0.9995	0.8337	0.7130	0.6247	0.5555
		10.0000	5.0000	2.5000	1.6667	1.2500	1.0000	0.8333	0.7143	0.6250	0.5556
	0.60	11.9590	6.0048	3.0107	1.9981	1.5029	1.2018	1.0020	0.8570	0.7495	0.6669
		12.0000	6.0000	3.0000	2.0000	1.5000	1.2000	1.0000	0.8571	0.7500	0.6667
	0.70	13.9916	6.9932	3.5006	2.3311	1.7495	1.4011	1.1671	1.0006	0.8752	0.7777
		14.0000	7.0000	3.5000	2.3333	1.7500	1.4000	1.1667	1.0000	0.8750	0.7778
	0.80	15.9731	8.0000	3.9985	2.6671	2.0012	1.6008	1.3335	1.1422	1.0002	0.8885
		16.0000	8.0000	4.0000	2.6667	2.0000	1.6000	1.3333	1.1429	1.0000	0.8889
	0.90	18.0343	8.9851	4.4999	3.0027	2.2454	1.8003	1.5010	1.2848	1.1252	1.0002
		18.0000	9.0000	4.5000	3.0000	2.2500	1.8000	1.5000	1.2857	1.1250	1.0000

According to the results presented in Tables 9 and 10, we can conclude that when the sample size is increasing, the mean value of the estimate approaches the true value of the probability ratio. When the probabilities are increasing, the mean value of estimator has bigger rate of convergence to the desired parameter $\theta = p_1 / p_2$.

Tables 11 and 12 are devoted to the simulations results for the special case of the direct-inverse sampling scheme $D(n, p_1) - I(\nu, p_2)$. In each cell of the tables, the following characteristics are presented: mean value of estimator $\hat{\theta}$ and the true value of the parameter $\theta = p_1 / p_2$.

Table 11 Mean value of estimator for the special case of the direct-inverse sampling scheme $D(n, p_1) - I(\nu, p_2)$ sample size for the first sample $n = 50$

	ν	Number of successes in the first sample									
n	p_1/p_2	0.05	0.10	0.20	0.30	0.40	0.50	0.60	0.70	0.80	0.90
	0.05	1.0670	0.5773	0.3251	0.2427	0.2007	0.1793	0.1643	0.1436	0.1364	0.1353
		1.0000	0.5000	0.2500	0.1677	0.1250	0.1000	0.0833	0.0714	0.0625	0.0556
	0.10	2.0043	1.0115	0.5052	0.3402	0.2559	0.2049	0.1719	0.1475	0.1296	0.1162
		2.0000	1.0000	0.5000	0.3333	0.2500	0.2000	0.1667	0.1429	0.1250	0.1111
50	0.20	4.0196	1.9896	0.9974	0.6633	0.4997	0.3999	0.3347	0.2852	0.2500	0.2225
		4.0000	2.0000	1.0000	0.6667	0.5000	0.4000	0.3333	0.2857	0.2500	0.2222
	0.30	6.0115	3.0058	1.5048	1.0073	0.7480	0.5992	0.5011	0.4294	0.3745	0.3324
		6.0000	3.0000	1.5000	1.0000	0.7500	0.6000	0.5000	0.4286	0.3750	0.3333
	0.40	7.9884	3.9807	1.9928	1.3339	0.9991	0.8017	0.6647	0.5713	0.5007	0.4439
		8.0000	4.0000	2.0000	1.3333	1.0000	0.8000	0.6667	0.5714	0.5000	0.4444

Table 11 (Continued)

	ν	Number of successes in the first sample									
n	p_1/p_2	0.05	0.10	0.20	0.30	0.40	0.50	0.60	0.70	0.80	0.90
	0.50	9.9738	4.9891	2.5008	1.6699	1.2496	0.9979	0.8329	0.7148	0.6260	0.5565
		10.0000	5.0000	2.5000	1.6667	1.2500	1.0000	0.8333	0.7143	0.6250	0.5556
	0.60	11.9717	5.9745	3.0030	1.9956	1.5006	1.2000	1.0011	0.8585	0.7510	0.6667
		12.0000	6.0000	3.0000	2.0000	1.5000	1.2000	1.0000	0.8571	0.7500	0.6667
50	0.70	14.0227	6.9857	3.5105	2.3251	1.7493	1.4004	1.1681	0.9997	0.8766	0.7785
		14.0000	7.0000	3.5000	2.3333	1.7500	1.4000	1.1667	1.0000	0.8750	0.7778
	0.80	16.0122	7.9855	4.0138	2.6685	1.9971	1.6002	1.3328	1.1389	1.0014	0.8899
		16.0000	8.0000	4.0000	2.6667	2.0000	1.6000	1.3333	1.1429	1.0000	0.8889
	0.90	17.9740	8.9974	4.5064	3.0006	2.2493	1.8022	1.4989	1.2852	1.1249	1.0003
		18.0000	9.0000	4.5000	3.0000	2.2500	1.8000	1.5000	1.2857	1.1250	1.0000

Table 12 Mean value of estimator for the special case of the direct-inverse sampling scheme $D(n, p_1) - I(\nu, p_2)$, sample size for the first sample $n = 200$

	ν	Number of successes in the first sample									
n	p_1/p_2	0.05	0.10	0.20	0.30	0.40	0.50	0.60	0.70	0.80	0.90
	0.05	0.9992	0.5030	0.2499	0.1667	0.1247	0.1001	0.0831	0.0718	0.0625	0.0556
		1.0000	0.5000	0.2500	0.1677	0.1250	0.1000	0.0833	0.0714	0.0625	0.0556
	0.10	1.9901	1.0023	0.4976	0.3332	0.2507	0.2002	0.1665	0.1430	0.1249	0.1113
		2.0000	1.0000	0.5000	0.3333	0.2500	0.2000	0.1667	0.1429	0.1250	0.1111
	0.20	3.9930	2.0058	0.9971	0.6680	0.4997	0.4003	0.3348	0.2861	0.2504	0.2221
		4.0000	2.0000	1.0000	0.6667	0.5000	0.4000	0.3333	0.2857	0.2500	0.2222
	0.30	6.0155	3.0054	1.4995	0.9998	0.7512	0.5985	0.4997	0.4273	0.3750	0.3331
		6.0000	3.0000	1.5000	1.0000	0.7500	0.6000	0.5000	0.4286	0.3750	0.3333
	0.40	8.0101	4.0097	1.9993	1.3354	1.0007	0.8013	0.6659	0.5715	0.5007	0.4441
		8.0000	4.0000	2.0000	1.3333	1.0000	0.8000	0.6667	0.5714	0.5000	0.4444
200	0.50	9.9916	4.9985	2.4979	1.6655	1.2476	0.9990	0.8341	0.7141	0.6256	0.5552
		10.0000	5.0000	2.5000	1.6667	1.2500	1.0000	0.8333	0.7143	0.6250	0.5556
	0.60	11.9974	5.9993	3.0002	2.0010	1.5013	1.2003	1.0004	0.8576	0.7502	0.6663
		12.0000	6.0000	3.0000	2.0000	1.5000	1.2000	1.0000	0.8571	0.7500	0.6667
	0.70	13.9938	7.0043	3.5029	2.3326	1.7500	1.3999	1.1659	1.0006	0.8750	0.7774
		14.0000	7.0000	3.5000	2.3333	1.7500	1.4000	1.1667	1.0000	0.8750	0.7778
	0.80	15.9917	8.0105	4.0022	2.6679	1.9978	1.6003	1.3331	1.1431	0.9992	0.8887
		16.0000	8.0000	4.0000	2.6667	2.0000	1.6000	1.3333	1.1429	1.0000	0.8889
	0.90	18.0132	8.9973	4.4956	3.0015	2.2484	1.7985	1.4997	1.2859	1.1252	1.0004
		18.0000	9.0000	4.5000	3.0000	2.2500	1.8000	1.5000	1.2857	1.1250	1.0000

According to the results presented in Tables 11 and 12, we can conclude that when the sample size is increasing, the mean value of the estimate approaches the true value of the probability ratio. When the probabilities are increasing, the mean value of estimator has bigger rate of convergence to the desired parameter $\theta = p_1 / p_2$.

3.2. Modeling of the mean squared error (MSE) for the estimators

Tables 13 and 14 are devoted to the continuation of our simulation results for the direct-direct sampling scheme $D(n_1, p_1) - D(n_2, p_2)$. In each cell of the tables, the following characteristics are presented: varaince of estimator and MSE.

Table 13 MSE for the estimator of direct-direct sampling scheme $D(n_1, p_1) - D(n_2, p_2)$, sample sizes $n_1 = 50$, $n_2 = 50$

n_2		50									
n_1	p_1/p_2	0.05	0.10	0.20	0.30	0.40	0.50	0.60	0.70	0.80	0.90
	0.05	0.7600	0.1400	0.0288	0.0119	0.0064	0.0040	0.0027	0.0020	0.0015	0.0012
		0.7718	0.1895	0.0330	0.0131	0.0067	0.0041	0.0028	0.0021	0.0015	0.0012
	0.10	2.2400	0.3600	0.0650	0.0252	0.0131	0.0080	0.0054	0.0038	0.0029	0.0022
		2.1004	0.4393	0.0728	0.0279	0.0135	0.0080	0.0054	0.0039	0.0029	0.0022
	0.20	7.3600	1.0400	0.1600	0.0563	0.0275	0.0160	0.0104	0.0072	0.0053	0.0041
		6.0740	1.3931	0.1822	0.0595	0.0284	0.0162	0.0106	0.0074	0.0053	0.0040
	0.30	15.3600	2.0400	0.2850	0.0933	0.0431	0.0240	0.0150	0.0101	0.0073	0.0054
		12.9035	3.0310	0.3275	0.0979	0.0444	0.0249	0.0155	0.0104	0.0074	0.0055
	0.40	26.2400	3.3600	0.4400	0.1363	0.0600	0.0320	0.0193	0.0126	0.0088	0.0064
		21.7980	4.9785	0.5080	0.1442	0.0630	0.0316	0.0200	0.0126	0.0089	0.0063
50	0.50	40.0000	5.0000	0.6250	0.1852	0.0781	0.0400	0.0231	0.0146	0.0098	0.0069
		32.5263	7.3448	0.8478	0.2065	0.0804	0.0411	0.0234	0.0145	0.0098	0.0069
	0.60	56.6400	6.9600	0.8400	0.2400	0.0975	0.0480	0.0267	0.0161	0.0103	0.0069
		45.3788	10.0623	0.9535	0.2686	0.1004	0.0509	0.0271	0.0164	0.0104	0.0070
	0.70	76.1600	9.2400	1.0850	0.3007	0.1181	0.0560	0.0298	0.0171	0.0104	0.0065
		60.4410	14.2003	1.3329	0.3170	0.1262	0.0593	0.0299	0.0176	0.0103	0.0067
	0.80	98.5600	11.8400	1.3600	0.3674	0.1400	0.0640	0.0326	0.0177	0.0100	0.0057
		75.3682	17.6839	1.6004	0.4003	0.1513	0.0677	0.0335	0.0181	0.0099	0.0056
	0.90	123.840	14.7600	1.6650	0.4400	0.1631	0.0720	0.0350	0.0178	0.0091	0.0044
		104.418	20.7373	2.2011	0.5109	0.1758	0.0743	0.0373	0.0186	0.0095	0.0045

Table 14 MSE for the estimator of direct-direct sampling scheme $D(n_1, p_1) - D(n_2, p_2)$, sample sizes $n_1 = 200$, $n_2 = 200$

n_2		200									
n_1	p_1/p_2	0.05	0.10	0.20	0.30	0.40	0.50	0.60	0.70	0.80	0.90
	0.05	0.1900	0.0350	0.0072	0.0030	0.0016	0.0010	0.0007	0.0005	0.0004	0.0003
		0.2450	0.0377	0.0074	0.0030	0.0016	0.0010	0.0007	0.0005	0.0004	0.0003
	0.10	0.5600	0.0900	0.0163	0.0063	0.0033	0.0020	0.0013	0.0010	0.0007	0.0006
		0.6388	0.1035	0.0164	0.0064	0.0033	0.0021	0.0014	0.0010	0.0007	0.0006
	0.20	1.8400	0.2600	0.0400	0.0141	0.0069	0.0040	0.0026	0.0018	0.0013	0.0010
		2.1429	0.2782	0.0417	0.0147	0.0068	0.0040	0.0025	0.0018	0.0013	0.0010
	0.30	3.8400	0.5100	0.0713	0.0233	0.0108	0.0060	0.0038	0.0025	0.0018	0.0014
		4.9274	0.5491	0.0746	0.0236	0.0106	0.0061	0.0037	0.0026	0.0018	0.0013
	0.40	6.5600	0.8400	0.1100	0.0341	0.0150	0.0080	0.0048	0.0031	0.0022	0.0016
		7.7865	0.9373	0.1121	0.0342	0.0151	0.0080	0.0048	0.0032	0.0022	0.0016
200	0.50	10.0000	1.2500	0.1563	0.0463	0.0195	0.0100	0.0058	0.0036	0.0024	0.0017
		13.4315	1.4262	0.1648	0.0475	0.0199	0.0099	0.0059	0.0035	0.0024	0.0018
	0.60	14.1600	1.7400	0.2100	0.0600	0.0244	0.0120	0.0067	0.0040	0.0026	0.0017
		16.6219	1.9611	0.2173	0.0602	0.0244	0.0122	0.0067	0.0040	0.0026	0.0017
	0.70	19.0400	2.3100	0.2713	0.0752	0.0295	0.0140	0.0075	0.0043	0.0026	0.0016
		23.2074	2.5394	0.2743	0.0763	0.0295	0.0141	0.0073	0.0043	0.0026	0.0016
	0.80	24.6400	2.9600	0.3400	0.0919	0.0350	0.0160	0.0081	0.0044	0.0025	0.0014
		31.3489	3.1079	0.3508	0.0942	0.0356	0.0159	0.0081	0.0044	0.0026	0.0014
	0.90	30.9600	3.6900	0.4163	0.1100	0.0408	0.0180	0.0088	0.0045	0.0023	0.0011
		42.5749	4.1309	0.4474	0.1149	0.0418	0.0182	0.0089	0.0045	0.0023	0.0011

According to the results presented in Tables 13 and 14, it is possible to conclude that the MSE decreases when sample size increases. Similarly, when both probabilities increase, the MSE decreases.

Table 15 is devoted to the continuation of our simulation results for the direct-inverse sampling scheme $D(n, p_1) - I(m, p_2)$. In each cell the following characteristics are presented: varaince of estimator and MSE.

Table 15 MSE for estimator of direct-inverse sampling scheme $D(n, p_1) - I(m, p_2)$, sample size of the first sample, number of success in the second sample $n_1 = 50$, $m_2 = n_2 p_2$

	n_2	50									
n_1	p_1/p_2	0.05	0.10	0.20	0.30	0.40	0.50	0.60	0.70	0.80	0.90
50	0.05	0.7600	0.1400	0.0288	0.0119	0.0064	0.0040	0.0027	0.0020	0.0015	0.0012
		0.6847	0.1576	0.0306	0.0119	0.0066	0.0041	0.0028	0.0020	0.0015	0.0012
	0.10	2.2400	0.3600	0.0650	0.0252	0.0131	0.0080	0.0054	0.0038	0.0029	0.0022
		2.0677	0.3865	0.0702	0.0250	0.0135	0.0082	0.0054	0.0039	0.0028	0.0023
	0.20	7.3600	1.0400	0.1600	0.0563	0.0275	0.0160	0.0104	0.0072	0.0053	0.0041
		6.8175	1.0921	0.1663	0.0585	0.0281	0.0161	0.0106	0.0071	0.0054	0.0041
	0.30	15.3600	2.0400	0.2850	0.0933	0.0431	0.0240	0.0150	0.0101	0.0073	0.0054
		13.7950	2.1218	0.2957	0.0932	0.0432	0.0237	0.0149	0.0100	0.0073	0.0054
	0.40	26.2400	3.3600	0.4400	0.1363	0.0600	0.0320	0.0193	0.0126	0.0088	0.0064
		23.5420	3.4105	0.4453	0.1391	0.0618	0.0319	0.0194	0.0126	0.0089	0.0063
	0.50	40.0000	5.0000	0.6250	0.1852	0.0781	0.0400	0.0231	0.0146	0.0098	0.0069
		37.0906	5.1023	0.6296	0.1838	0.0784	0.0402	0.0235	0.0146	0.0096	0.0068
	0.60	56.6400	6.9600	0.8400	0.2400	0.0975	0.0480	0.0267	0.0161	0.0103	0.0069
		45.3788	10.0623	0.9535	0.2686	0.1004	0.0509	0.0271	0.0164	0.0104	0.0070
	0.70	76.1600	9.2400	1.0850	0.3007	0.1181	0.0560	0.0298	0.0171	0.0104	0.0065
		60.4410	14.2003	1.3329	0.3170	0.1262	0.0593	0.0299	0.0176	0.0103	0.0067
	0.80	98.5600	11.8400	1.3600	0.3674	0.1400	0.0640	0.0326	0.0177	0.0100	0.0057
		75.3682	17.6839	1.6004	0.4003	0.1513	0.0677	0.0335	0.0181	0.0099	0.0056
	0.90	123.840	14.7600	1.6650	0.4400	0.1631	0.0720	0.0350	0.0178	0.0091	0.0044
		104.418	20.7373	2.2011	0.5109	0.1758	0.0743	0.0373	0.0186	0.0095	0.0045

According to the results presented in Table 15, it is possible to conclude that the MSE decreases when sample size increases. Similarly, when both probabilities increase, the MSE decreases.

Table 16 is devoted to the continuation of our simulation results for the inverse-direct sampling scheme $I(m, p_1) - D(n, p_2)$. In each cell of the tables, the following characteristics are presented: varaince of estimator and MSE.

Table 16 MSE for estimator of the inverse-direct sampling scheme $I(m, p_1) - D(n, p_2)$, number of successes in the first sample, size of the second sample $m_1 = n_1 p_1$, $n_2 = 50$

	n_2	50									
n_1	p_1/p_2	0.05	0.10	0.20	0.30	0.40	0.50	0.60	0.70	0.80	0.90
50	0.05	0.7600	0.1400	0.0288	0.0119	0.0064	0.0040	0.0027	0.0020	0.0015	0.0012
		0.7237	0.2181	0.1486	0.1046	0.0696	0.0435	0.0391	0.0228	0.0201	0.0155
	0.10	2.2400	0.3600	0.0650	0.0252	0.0131	0.0080	0.0054	0.0038	0.0029	0.0022
		2.9576	0.4342	0.0557	0.0287	0.0163	0.0118	0.0076	0.0058	0.0044	0.0034

Table 16 (Continued)

	n_2	50									
n_1	p_1/p_2	0.05	0.10	0.20	0.30	0.40	0.50	0.60	0.70	0.80	0.90
50	0.20	7.3600	1.0400	0.1600	0.0563	0.0275	0.0160	0.0104	0.0072	0.0053	0.0041
		11.7844	1.6520	0.1323	0.0477	0.0275	0.0177	0.0116	0.0088	0.0062	0.0046
	0.30	15.3600	2.0400	0.2850	0.0933	0.0431	0.0240	0.0150	0.0101	0.0073	0.0054
		26.5115	3.6525	0.2446	0.0709	0.0395	0.0235	0.0157	0.0108	0.0080	0.0060
	0.40	26.2400	3.3600	0.4400	0.1363	0.0600	0.0320	0.0193	0.0126	0.0088	0.0064
		47.0791	6.4392	0.4003	0.0979	0.0497	0.0309	0.0195	0.0136	0.0092	0.0069
	0.50	40.0000	5.0000	0.6250	0.1852	0.0781	0.0400	0.0231	0.0146	0.0098	0.0069
		73.6400	10.0239	0.5845	0.1288	0.0638	0.0370	0.0230	0.0151	0.0102	0.0070
	0.60	56.6400	6.9600	0.8400	0.2400	0.0975	0.0480	0.0267	0.0161	0.0103	0.0069
		105.971	14.4582	0.8094	0.1619	0.0776	0.0421	0.0257	0.0158	0.0107	0.0071
	0.70	76.1600	9.2400	1.0850	0.3007	0.1181	0.0560	0.0298	0.0171	0.0104	0.0065
		144.043	19.6196	1.0730	0.1983	0.0920	0.0498	0.0288	0.0167	0.0102	0.0066
	0.80	98.5600	11.8400	1.3600	0.3674	0.1400	0.0640	0.0326	0.0177	0.0100	0.0057
		188.054	25.5623	1.3629	0.2451	0.1040	0.0555	0.0300	0.0173	0.0099	0.0057
	0.90	123.840	14.7600	1.6650	0.4400	0.1631	0.0720	0.0350	0.0178	0.0091	0.0044
		238.355	32.3192	1.6924	0.2830	0.1214	0.0596	0.0327	0.0174	0.0089	0.0044

According to the results presented in Table 16, it is possible to conclude that the MSE decreases when sample size increases. Similarly, when both probabilities increase, the MSE decreases.

Tables 17 is devoted to the continuation of our simulation results for the inverse-inverse sampling scheme $I(m_1, p_1) - I(m_2, p_2)$. In each cell of the tables, the following characteristics are presented: varaince of estimator and MSE.

Table 17 MSE for estimator of the inverse-inverse sampling scheme $I(m_1, p_1) - I(m_2, p_2)$, numbers of successes in the first and second samples $m_1 = n_1 p_1$, $m_2 = n_2 p_2$

	n_2	50									
n_1	p_1/p_2	0.05	0.10	0.20	0.30	0.40	0.50	0.60	0.70	0.80	0.90
50	0.05	0.7600	0.1400	0.0288	0.0119	0.0064	0.0040	0.0027	0.0020	0.0015	0.0012
		5.1606	1.3397	0.3529	0.1602	0.0837	0.0507	0.0352	0.0265	0.0158	0.0164
	0.10	2.2400	0.3600	0.0650	0.0252	0.0131	0.0080	0.0054	0.0038	0.0029	0.0022
		2.3597	0.4695	0.0945	0.0369	0.0191	0.0121	0.0078	0.0059	0.0043	0.0033
	0.20	7.3600	1.0400	0.1600	0.0563	0.0275	0.0160	0.0104	0.0072	0.0053	0.0041
		7.1528	1.1674	0.1833	0.0651	0.0331	0.0182	0.0119	0.0089	0.0062	0.0048
	0.30	15.3600	2.0400	0.2850	0.0933	0.0431	0.0240	0.0150	0.0101	0.0073	0.0054
		14.1460	2.2625	0.3097	0.0997	0.0468	0.0261	0.0167	0.0110	0.0081	0.0061
	0.40	26.2400	3.3600	0.4400	0.1363	0.0600	0.0320	0.0193	0.0126	0.0088	0.0064
		23.4426	3.4706	0.4621	0.1439	0.0623	0.0328	0.0199	0.0133	0.0089	0.0068
	0.50	40.0000	5.0000	0.6250	0.1852	0.0781	0.0400	0.0231	0.0146	0.0098	0.0069
		37.8273	5.1949	0.6461	0.1933	0.0802	0.0421	0.0243	0.0152	0.0103	0.0069
	0.60	56.6400	6.9600	0.8400	0.2400	0.0975	0.0480	0.0267	0.0161	0.0103	0.0069
		51.6918	7.1170	0.8639	0.2363	0.1017	0.0482	0.0272	0.0162	0.0105	0.0071
	0.70	76.1600	9.2400	1.0850	0.3007	0.1181	0.0560	0.0298	0.0171	0.0104	0.0065
		70.0740	9.5893	1.0936	0.3165	0.1187	0.0562	0.0306	0.0181	0.0107	0.0066
	0.80	98.5600	11.8400	1.3600	0.3674	0.1400	0.0640	0.0326	0.0177	0.0100	0.0057
		89.0514	11.7404	1.3399	0.3576	0.1418	0.0650	0.0328	0.0179	0.0101	0.0059
	0.90	123.840	14.7600	1.6650	0.4400	0.1631	0.0720	0.0350	0.0178	0.0091	0.0044
		109.574	14.6581	1.6786	0.4375	0.1702	0.0730	0.0360	0.0179	0.0093	0.0045

According to the results presented in Table 17, it is possible to conclude that the MSE decreases when sample size increases. Similarly, when both probabilities increase, the MSE decreases.

Table 18 is devoted to the continuation of our simulation results for the special case direct-inverse sampling scheme $D(n, p_1) - I(\nu, p_2)$. In each cell of the tables, the following characteristics are presented: varaince of estimator and MSE.

Table 18 MSE for estimator of special case direct-inverse sampling scheme $D(n, p_1) - I(\nu, p_2)$, sample size for the first sample $n = 50$

	ν	Number of successes in the first sample									
n	p_1/p_2	0.05	0.10	0.20	0.30	0.40	0.50	0.60	0.70	0.80	0.90
	0.05	0.7600	0.1850	0.0438	0.0183	0.0097	0.0058	0.0038	0.0026	0.0018	0.0013
		0.6683	0.1840	0.0796	0.0701	0.0668	0.0691	0.0713	0.0639	0.0664	0.0722
	0.10	1.4800	0.3600	0.0850	0.0356	0.0188	0.0112	0.0072	0.0049	0.0034	0.0025
		1.4667	0.3596	0.0844	0.0371	0.0211	0.0143	0.0108	0.0087	0.0070	0.0063
	0.20	2.8000	0.6800	0.1600	0.0667	0.0350	0.0208	0.0133	0.0090	0.0063	0.0044
		2.8243	0.6760	0.1594	0.0652	0.0350	0.0207	0.0132	0.0089	0.0062	0.0045
	0.30	3.9600	0.9600	0.2250	0.0933	0.0488	0.0288	0.0183	0.0122	0.0084	0.0059
		4.0380	0.9650	0.2188	0.0948	0.0481	0.0283	0.0184	0.0125	0.0086	0.0061
	0.40	4.9600	1.2000	0.2800	0.1156	0.0600	0.0352	0.0222	0.0147	0.0100	0.0069
		5.0062	1.2116	0.2748	0.1171	0.0602	0.0353	0.0221	0.0146	0.0099	0.0069
50	0.50	5.8000	1.4000	0.3250	0.1333	0.0688	0.0400	0.0250	0.0163	0.0109	0.0074
		5.8912	1.4372	0.3159	0.1317	0.0681	0.0400	0.0255	0.0163	0.0112	0.0075
	0.60	6.4800	1.5600	0.3600	0.1467	0.0750	0.0432	0.0267	0.0171	0.0113	0.0074
		6.5191	1.5636	0.3597	0.1455	0.0754	0.0433	0.0269	0.0172	0.0112	0.0077
	0.70	7.0000	1.6800	0.3850	0.1556	0.0788	0.0448	0.0272	0.0171	0.0109	0.0069
		7.0476	1.6916	0.3815	0.1525	0.0773	0.0447	0.0273	0.0174	0.0109	0.0070
	0.80	7.3600	1.7600	0.4000	0.1600	0.0800	0.0448	0.0267	0.0163	0.0100	0.0059
		7.4357	1.7560	0.4039	0.1604	0.0772	0.0440	0.0259	0.0164	0.0099	0.0059
	0.90	7.5600	1.8000	0.4050	0.1600	0.0788	0.0432	0.0250	0.0147	0.0084	0.0044
		7.5363	1.7825	0.4088	0.1624	0.0789	0.0432	0.0250	0.0147	0.0086	0.0045

According to the results presented in Table 18, it is possible to conclude that the MSE decreases when sample size increases. Similarly, when both probabilities increase, the MSE decreases.

4. Comparison and Analysis of the Accuracy Properties of the Estimators

As mentioned above, we observe the expected tendency of MSE decreasing with increasing sample sizes and with increasing success probabilities in each of two samples

Simulation results show that the best accuracy from the MSE point of view occurs with the special case of the direct-inverse sampling scheme, where the first sample is obtained by the direct Bernoulli sampling scheme, and the second sample is retrieved with the inverse sampling scheme with the number of successes equals to the number of successes in the first sample.

The worst sampling scheme that possesses the largest MSE appears to be the scheme of two independent samples where the first one is obtained with the Inverse sampling scheme and the second one with the direct.

All other sampling schemes have practically identical MSE. To support these findings, we present Table 19 of typical values of MSE for all five sampling schemes. For a convenience, $n_1 = n_2 = 200$ and $p_1 = 0.2$, $p_2 = 0.05$ are taken. We refer to the remark above, where we stated that for the scheme of inverse-inverse sampling sheme $I(m_1, p_1) - I(m_2, p_2)$ is equivalent to the direct-direct sampling scheme $D(n_1, p_1) - D(n_2, p_2)$, if $m_1 = n_1 p_1 = 40$ and $m_2 = n_2 p_2 = 10$.

Table 19 Comparison of the performance for all sample schemes

Sampling Scheme	Variance	MSE
Direct-Direct $D(n_1, p_1) - D(n_2, p_2)$	1.8400	2.1429
Direct-Inverse $D(n, p_1) - I(m, p_2)$	1.8400	1.9030
Inverse-Direct $I(m, p_1) - D(n, p_2)$	1.8400	6.8242
Inverse-Direct $I(m_1, p_1) - I(m_2, p_2)$	1.8400	1.8792
Special Case $D(n, p_1) - I(\nu, p_2)$	0.7000	0.6998

5. Conclusions

In this paper, we were determined the optimal sampling scheme of Bernoulli trials that provides the best accuracy for the estimation of probability ratio, This is the fifth considered scheme, namely the special case of the direct-inverse sampling scheme $D(n, p_1) - I(\nu, p_2)$.

According to the results presented in the tables, it is possible to make the following ranking (from best to worst) for the quality of estimators for the different sampling schemes:

1. Special Case Direct-Inverse $D(n, p_1) - I(\nu, p_2)$ - the smallest MSE
2. Inverse-Inverse $I(m_1, p_1) - I(m_2, p_2)$
3. Direct-Inverse $D(n, p_1) - I(m, p_2)$
4. Direct-Direct $D(n_1, p_1) - D(n_2, p_2)$
5. Inverse-Direct $I(m, p_1) - D(n, p_2)$ - the largest MSE

The minimal MSE is approximately ten times different from the maximum. For the verification of this statement, we present additional table with the typical values of the MSE for only two sampling schemes. To present different results than in Table 19, the values $n_1 = n_2 = 100$ and $p_1 = 0.15$, $p_2 = 0.04$ are considered in Table 20.

Table 20 Comparison of the performance for two extreme sample schemes

Sampling Scheme	MSE
Inverse-Direct $I(m, p_1) - D(n, p_2)$	6.9353
Special Case $D(n, p_1) - I(\nu, p_2)$	0.6999

Acknowledgements

The author is grateful to the referees for carefully reading the manuscript and for offering extremely valuable comments and suggestions which enabled the author to substantially improve the paper.

References

Bennett BM. On the use of the negative binomial in epidemiology. Biom J. 1981; 23(1): 69-72.

Guttman I. A note on a series solution of a problem in estimation. Biometrika. 1958; 45: 565-567.

Lui KJ. Point estimation on relative risk under inverse sampling. Biom J. 1996; 38(6): 669-680.

Ngamkham T, Volodin A, Volodin I. Confidence intervals for a ratio of Binomial proportions based on direct and inverse sampling schemes. Lobachevskii J Math. 2016; 37(4): 466-494.

Ngamkham T, Confidence interval estimation for the ratio of binomial proportions and random numbers generation for some statistical models. Ph.D. [Thesis]. Southeast Reginal: University of Regina; 2018.

R Core Team. R: a language and environment for statistical computing. Vienna: R Foundation for Statistical Computing; 2019.

Roberts CD. Unbiased estimation of odds ratio by using negative binomial plans. Biom J. 1993; 35(5): 581-587.

Sheps MC. An examination of some methods of comparing several rates of proportions. Biometrics 1959; 15(1): 87-89.