doi:10.21608/esju.2024.278969.1028 Research Article
The Egyptian Statistical Journal (ESJ), 68 (2): 1-14

Homepage: https://esju.journals.ekb.eg/

The Egyptian Statistical Journal

Print ISSN 0542-1748– Online ISSN 2786-0086

Parametric Estimation under Type I Censoring Geometric Distribution with Missing Data

Naglaa Abdelmoneim Morad*

Received 24 March 2024; revised 10 August 2024; accepted 12 August 2024

Keywords

Geometric Distribution; Missing Data; Maximum Likelihood Estimators; Type-I Censored and Mean Square Error.

Abstract

This manuscript's main objective is to estimate the geometric distribution's parameters using progressive type-I censored data with missing data. The efficiency of estimators is studied. Moreover, the estimators' consistency property is shown. To produce the estimators, the maximum likelihood technique is performed. The geometric distribution is applied to n = 10, 30, and 50 to create samples. The initial values are set as $\theta_1 = 0.1$, θ_2= 0.2 for one geometric population and θ = 0.15 for a general parameter involving two geometric populations. T_0= 5,10, and 15 are the censored times. Missing data (δ=η=0) has a percentage of 0.1, 0.2, 0.3, 0.5, and 0.7. The bias criterion and the root mean square error (RMSE) are applied. The outcomes for each of the one and two populations revealed a number of significant markers. Both the bias and the root mean square error increase in percentage to the amount of missing data. Conversely, the bias and root mean square error decreases with increasing sample size and censoring time.

1. Introduction

The concept of missing data, in which the observed value of a variable is unknown, is connected to the concept of censored data, in which the observed value of a variable is only partially known. It is typical for observational data to be insufficient in some applications. Two categories of data censorship exist: According to Yu et al. (2023), Type I censoring happens when the process end time is predetermined, and Type II censoring occurs when the quantity of observation data exceeds a predetermined level. For long-term survival models, type-I censored sampling is beneficial. Numerous real-world situations make extensive use of the geometric probability distribution. For instance, geometric distribution is used in the banking sector to assess the financial benefits of a certain option through cost-benefit analysis.

The term "**Missing Data**" describes the amount of missing data for a single participant and the amount of missing data for a single variable within a data set. Missing data refers to data that contains various codes indicating a lack of response, such as "don't know," "refused," "unintelligible," and so on. A few research papers feature missing data. After data collection, researchers are left with partial data sets that do not contain all of the information for every study participant. As a result, scientists have developed various methods for estimating the unknown parameters of various models when data is missing, including the maximum likelihood approach, mean imputation, listwise algorithm, and so on (Acock, 2005). The majority of statistical techniques, however, were developed to analyze data sets with no missing information. So, the researcher has two options: remove cases with incomplete data or use approximated values to fill

Corresponding author*: naglaa_morad@yahoo.com
[1] Faculty of Graduate Studies for Statistical Research, Cairo University, Egypt.

in the gaps. This leads to creating a data collection devoid of missing values (Schafer and Graham, 2002). These are only some of the many studies discussing distributions created by multiplying other life distributions. For instance, using progressive type-II censored data, Feyza and Mehmet (2018) computed the parameters of the Exponential-Geometric (EG) distribution. Long (2021) used double Type-I hybrid censored data to estimate the parameters of the Rayleigh distribution. Akhtar et al. (2023) computed the geometric distribution parameter under a type-I censoring scheme using the Bayesian estimation approach. See [Fayomi and Al-Shammari (2018), Seunggeun et al. (2016)]. Goel and Krishna (2022), Mallick and Ram (2018) and Nadeem et al. (2023)].

Numerous studies examined the geometric distribution, while others focused on the missing data. The contribution of this paper is the estimation of the unknown geometric distribution parameters based on type-I censored data, where part of the data is missing. In addition, the estimations' characteristics are evaluated upon. Studies using simulations are conducted for various sample sizes, censored times, and percentages of missing data. The criterion for bias and root mean square error (RMSE) are used.

This article has the following structure: Section 1 covers the introduction, while Section 2 explains the geometric distribution. Section 3 describes estimators' properties. In Section 4, data analysis and simulation are carried out. The conclusion can be found in section 5. Appendices A and B provide the derivatives of the equations and figures.

2. Geometric Distribution

Geometric distribution is a type of discrete probability distribution that represents the probability of the number of successive failures before a success is obtained in a Bernoulli trial. The geometric distribution is widely used in several real-life scenarios. For example, in financial industries, geometric distribution is used for a cost-benefit analysis to estimate the financial benefits of making a certain decision. See Hossain and Begum (2016).

In this section, estimators of parameters under type I censoring samples with partial data missing are developed using the maximum likelihood technique for geometric distributions. A geometric distribution's probability mass function takes the following form:

$$P(X = x_i) = (1-\theta)^{x_i - 1}\theta \qquad x_i = 1,2,\dots \qquad 0 < \theta < 1 \tag{1}$$

$$E(X) = \frac{1}{\theta}, \qquad var(X) = \frac{1-\theta}{\theta^2} \tag{2}$$

$$F(X = x_i) = 1 - (1-\theta)^{x_i} \tag{3}$$

The sample is denoted as $(x_1, x_2, \dots, x_n)$ with parameter θ (unknown).

The actual observed data is $(z_i, \alpha_i, \delta_i)$, $(i = 1,2,\dots,n)$ where $(z_1, z_2, \dots, z_n)$, $(\alpha_1, \alpha_2, \dots, \alpha_n)$ and $(\delta_1, \delta_2, \dots, \delta_n)$ are independent.

Assume that:

$$z_i = \min(T_0, x_i), \quad \alpha_i = I\{x_i \le T_0\} - I\{x_i > T_0\} \tag{4}$$

be the censoring or status indicator for T_0, where T_0 be a predetermined time to terminate the experiment.

$$\alpha_i = \begin{cases} 1 & x_i \le T_0, z_i \text{ is observed} \\ -1 & x_i > T_0, z_i \text{ is censored} \end{cases} \tag{5}$$

and

$$\delta_i = \begin{cases} 1 & z_i \text{ is observed with } p(\delta_i = 1) = p \\ 0 & \text{otherwise} \end{cases} \tag{6}$$

2.1 Estimation of Parameters in Case of One Population

The maximum likelihood function has the form:

$$L(\theta) = \prod_{i=1}^{n} (f(z_i; \theta))^{A_i} (S(z_i; \theta))^{B_i} \tag{7}$$

where

$$S(z_i; \theta) = 1 - F(z_i) = (1 - \theta)^{z_i} \tag{8}$$

be survival function, and

$$A_i = \alpha_i \delta_i (\alpha_i \delta_i + 1)/2 \tag{9}$$

$$B_i = \alpha_i \delta_i (\alpha_i \delta_i - 1)/2 \tag{10}$$

The estimator of parameter θ:

$$\hat{\theta} = \frac{\sum_{i=1}^{n} A_i}{\sum_{i=1}^{n} (A_i + B_i) z_i} \tag{11}$$

It is presented two geometric populations for n independent observations.
The first sample: the sample is denoted as $(x_1, x_2, \dots, x_n)$ with parameter θ_1 (unknown). The actual observed data is $(z_i, \alpha_i, \delta_i)$ $(i = 1,2, \dots, n)$, where $(z_1, z_2, \dots, z_n)$, $(\alpha_1, \alpha_2, \dots, \alpha_n)$ and $(\delta_1, \delta_2, \dots, \delta_n)$ are independent. Similarly, the second sample: the sample is denoted as $(y_1, y_2, \dots, y_n)$ with parameter θ_2 (unknown). The actual observed data is (M_i, β_i, η_i) where $(M_1, M_2, \dots, M_n)$, $(\beta_1, \beta_2, \dots, \beta_n)$ and $(\eta_1, \eta_2, \dots, \eta_n)$ are independent. Assume that:

$$M_i = \min(T_0, y_i), \; \beta_i = I\{y_i \le T_0\} - I\{y_i > T_0\} \tag{12}$$

be the censoring or status indicator for T_0

Where

$$\beta_i = \begin{cases} 1 & y_i \le T_0, \; M_i \text{ is observed} \\ -1 & y_i > T_0, \; M_i \text{ is censored} \end{cases} \tag{13}$$

And

$$\eta_i = \begin{cases} 1 & M_i \text{ is observed with } p(\eta_i = 1) = p \\ 0 & \text{otherwise} \end{cases} \tag{14}$$

The estimator of parameter θ_1, θ_2:

$$\hat{\theta}_1 = \frac{\sum_{i=1}^{n} A_i}{\sum_{i=1}^{n} (A_i + B_i) z_i}, \qquad \hat{\theta}_2 = \frac{\sum_{i=1}^{n} C_i}{\sum_{i=1}^{n} (C_i + D_i) M_i} \tag{15}$$

Appendix A1 contains the equations' derivation.

2.2 Estimation of Parameters in Case of Two Populations

The bivariate geometric distribution, covered in the next section, is an extension of the univariate geometric distribution.

Assume $\theta_1 = \theta_2 = \theta$ where θ is unknown, the likelihood function of θ is

$$L(\theta) = \prod_{i=1}^{n} (f(z_i;\theta_1))^{A_i} (S(z_i;\theta_1))^{B_i} \prod_{i=1}^{n} (f(M_i;\theta_2))^{C_i} (S(M_i;\theta_2))^{D_i} \tag{16}$$

The estimator of parameter θ:

$$\hat{\theta} = \frac{\frac{1}{2}(\sum_{i=1}^{n} \alpha_i^2 \delta_i^2 + \sum_{i=1}^{n} \alpha_i \delta_i + \sum_{i=1}^{n} \beta_i^2 \eta_i^2 + \sum_{i=1}^{n} \beta_i \eta_i)}{\sum_{i=1}^{n} \alpha_i^2 \delta_i^2 z_i + \sum_{i=1}^{n} \beta_i^2 \eta_i^2 M_i} \tag{17}$$

Appendix A2 contains the equations' derivation.

3. Properties of Estimators

The consistency of estimators is demonstrated in this section.

Since, $\{\alpha_i \delta_i, 1 \le i \le n\}$ are independently identically distributed variables, so

$$\frac{1}{n}\sum_{i=1}^{n} \alpha_i \delta_i \xrightarrow{a.s.} E(\alpha_i \delta_i) \tag{18}$$

$$\hat{\theta} = \frac{2(1-(1-\theta)^{T_0})}{2\frac{1}{\theta}(1-(1-\theta)^{T_0})} \xrightarrow{a.s.} \theta \tag{19}$$

Appendix A3 has the proof of this.

4. Data Simulation and Analysis

A simulation study used R software to assess the estimated parameters' mathematical derivations and behavior. The simulation study is performed for different choices of sample size (n) and censoring time (T_0) . Samples of size n = 10, 30 and 50 and censoring time $T_0 = 5, 10,$ and 15 by using the probability of missing = 0.1, 0.2, 0.3,0.5 and 0.7 have been generated from geometric distribution with missing data. The procedure was executed with 1000 iterations, with initial parameter values $(\theta_1 = 0.1)$ and $(\theta_2 = 0.2)$ chosen for one population. Furthermore, the initial value for the two populations is $(\theta = 0.15)$.

The parameters' bias and the root mean squared error (RMSE) are the two criteria applied. The RMSE is frequently used to assess the estimators' accuracy.

$$\text{Bias} = E(estimator - true\ value) \tag{20}$$

$$\text{RMSE} = \sqrt{\text{E}(estimator - true\ value)^2} \tag{21}$$

Table 1. The Root Mean Square Error of Estimators for One Population

Percentage of Missing	Sample size	10			30			50		
		$T_0 = 5$	$T_0 = 10$	$T_0 = 15$	$T_0 = 5$	$T_0 = 10$	$T_0 = 15$	$T_0 = 5$	$T_0 = 10$	$T_0 = 15$
10%	$\theta_1 = 0.1$	0.08123423	0.0653529	0.059338	0.0480208	0.0358469	0.0314643	0.0408044	0.0293922	0.0256515
	$\theta_2 = 0.2$	0.18769656	0.1687898	0.1643198	0.1075174	0.0899333	0.0864639	0.0944193	0.0765441	0.0738106
20%	$\theta_1 = 0.1$	0.08809086	0.0719653	0.0673647	0.0507445	0.0379868	0.0335261	0.0423556	0.0302721	0.0270998
	$\theta_2 = 0.2$	0.20307634	0.1894466	0.1889709	0.1145862	0.0950947	0.0936162	0.0978264	0.0794103	0.0751598
30%	$\theta_1 = 0.1$	0.09544632	0.0794858	0.0759737	0.0527447	0.0399249	0.0368239	0.0441475	0.0322335	0.0286199
	$\theta_2 = 0.2$	0.23943475	0.2181972	0.2132393	0.1202497	0.1002497	0.097678	0.10005	0.0832046	0.0801062
50%	$\theta_1 = 0.1$	0.09544632	0.0867064	0.0887243	0.0653948	0.0399249	0.0463442	0.0441475	0.0322335	0.0286199
	$\theta_2 = 0.2$	0.270958	0.2145338	0.2314541	0.2553516	0.2191731	0.2075262	0.2254205	0.1905654	0.1790768
70%	$\theta_1 = 0.1$	0.09544632	0.0933702	0.0998599	0.0743623	0.0399249	0.047473	0.0441475	0.0322335	0.0286199
	$\theta_2 = 0.2$	0.299177	0.2407764	0.2483365	0.3405127	0.2932979	0.2767548	0.302686	0.2563345	0.2402499

Table 2. The Bias of Estimators for One Population

Percentage of Missing	Sample size	10			30			50		
		$T_0 = 5$	$T_0 = 10$	$T_0 = 15$	$T_0 = 5$	$T_0 = 10$	$T_0 = 15$	$T_0 = 5$	$T_0 = 10$	$T_0 = 15$
10%	$\theta_1 = 0.1$	0.006467	0.004186	0.003451	0.00226	0.001259	0.00097	0.001632	0.000847	0.000645
	$\theta_2 = 0.2$	0.034525	0.02792	0.026461	0.011329	0.007926	0.007326	0.008737	0.005742	0.005339
20%	$\theta_1 = 0.1$	0.007605	0.005075	0.004447	0.002524	0.001414	0.001102	0.001758	0.000898	0.00072
	$\theta_2 = 0.2$	0.040415	0.035172	0.034996	0.012867	0.008862	0.008589	0.009379	0.00618	0.005536
30%	$\theta_1 = 0.1$	0.008928	0.006192	0.005657	0.002726	0.001562	0.001329	0.00191	0.001018	0.000803
	$\theta_2 = 0.2$	0.056182	0.046658	0.044562	0.014171	0.009849	0.00935	0.00981	0.006785	0.006289
50%	$\theta_1 = 0.1$	0.008928	0.007368	0.007715	0.004191	0.001562	0.002105	0.00191	0.001018	0.000803
	$\theta_2 = 0.2$	0.07195	0.045104	0.0525	0.0639	0.047076	0.042206	0.049798	0.035589	0.031427
70%	$\theta_1 = 0.1$	0.008928	0.008544	0.009773	0.005419	0.001562	0.002209	0.00191	0.001018	0.000803
	$\theta_2 = 0.2$	0.087717	0.056814	0.060438	0.11363	0.084303	0.075061	0.089786	0.064393	0.056566

Table 3. The Root Mean Square Error of Estimators for Two Population

Percentage of Missing	Sample size	10			30			50		
		$T_0 = 5$	$T_0 = 10$	$T_0 = 15$	$T_0 = 5$	$T_0 = 10$	$T_0 = 15$	$T_0 = 5$	$T_0 = 10$	$T_0 = 15$
10%		0.077737	0.058387	0.05201	0.053122	0.035114	0.028953	0.048021	0.029833	0.023816
20%		0.082735	0.063182	0.057219	0.055767	0.036742	0.031267	0.049315	0.030557	0.0247427
30%		0.086499	0.067617	0.061976	0.057871	0.038562	0.033615	0.050359	0.032016	0.0265085
50%		0.090105	0.071777	0.066393	0.0599	0.040299	0.035811	0.051381	0.033411	0.0281638
70%		0.093574	0.07571	0.070534	0.061863	0.041964	0.037879	0.052383	0.034751	0.0297271

Table 4. The Bias of Estimators for Two Population

Percentage of Missing	Sample size	10			30			50		
		$T_0 = 5$	$T_0 = 10$	$T_0 = 15$	$T_0 = 5$	$T_0 = 10$	$T_0 = 15$	$T_0 = 5$	$T_0 = 10$	$T_0 = 15$
10%		0.005922	0.003341	0.002651	0.002766	0.001208	0.000822	0.00226	0.000872	0.000556
20%		0.006708	0.003912	0.003209	0.003048	0.001323	0.000958	0.002383	0.000915	0.0006
30%		0.007332	0.004481	0.003764	0.003282	0.001457	0.001107	0.002485	0.001005	0.000689
50%		0.007957	0.005049	0.00432	0.003516	0.001592	0.001257	0.002587	0.001094	0.000777
70%		0.008581	0.005617	0.004876	0.00375	0.001726	0.001406	0.002689	0.001183	0.000866

The tables above and the graphics below allow for drawing several important conclusions for populations one and two. These include:

1. As sample size increases, bias and root mean square error (RMSE) decrease.
2. As the percentage of missing data increases, bias and root mean square error (RMSE) increase.
3. As censoring time increases, bias and root mean square error (RMSE) decrease.
4. The root mean square error (RMSE) and bias values are generally very small, except for the values of θ_2 for one population for sample sizes 30 and 50 for percentages of missing data equal to 50% and 70%, which are greater than the other values.

5. Conclusion

The main goal of this article is to estimate the unknown parameters of the geometric distribution with progressive type-I censorship and missing data. Furthermore, properties of estimator consistency were determined. The maximum likelihood approach was used to find estimators for the model's parameters in one and two populations.

The two criteria, root mean square error and bias, are calculated numerically. The simulation technique was run with different censoring times, percentages of missing data, and sample sizes.

The results showed that the percentage of missing data increased, so did the root mean square error and bias. In contrast, the root mean square error and bias decrease when the censoring time and sample size increase.

Declaration of interests

The authors declare that they have no conflict of interest.

References

Acock, A.C. (2005). Working with missing values. *Journal of Marriage and Family,* 67(4):1012–1028.

Akhtar, N., Khan, S. A., Amin, M., Khan, A. A., Ali, A., and Manzoor, S. (2023). Bayesian estimation of a geometric distribution using informative priors based on a type-I censoring scheme. *Statistics in Transition*, 24(3): 257-263. doi.org/10.59170/stattrans-2023-047.

Fayomi, A. and Al-Shammari, H. (2018). Estimating the parameters of the exponential-geometric distribution based on progressively type-II censored data. *International Journal of Advanced Statistics and Probability.* 6(1): 37–41. https://doi.org/10.14419/ijasp.v6i1.10450

Feyza G, and Mehmet Y. (2018). Different parameter estimation methods for exponential geometric distribution and its applications in lifetime data analysis, *Biostatistics and Biometrics Open Access Journal;* 8(2): 1-8. doi: 10.19080/BBOAJ.2018.08.555735.

Goel, N. and Krishna, H. (2022). Different methods of estimation in two parameter geometric distribution with randomly censored data. *International Journal of System Assurance Engineering and Management* 13(4): 1652–1665. https://doi.org/10.1007/s13198-021-01520-1

Hossain, F. and Begum, M. (2016). Tests of hypotheses for the parameters of a bivariate geometric distribution. *Biometrics and Biostatistics International Journal*, 4(6): 244-249. https://doi.org/10.15406/bbij.2016.04.00112

Long, B. (2021). Estimation and prediction for the rayleigh distribution based on double type-I hybrid censored data. *Communications in Statistics-Simulation and Computation,* 52(1): 1–15. https://doi.org/10.1080/03610918.2021.1938123

Mallick, A. and Joshi, R. (2018). Parameter estimation and application of generalized inflated geometric distribution. *Journal of Statistical Theory and Applications*. 17(3): 491–519. *https://doi.org/10.2991/jsta.2018.17.3.7*

Nadeem, A., Sajjad, A., Muhammad, A., Akbar, A., Amjad, A. and Sadaf, M. (2023). Bayesian estimation of a geometric distribution using informative priors based on a Type-I censoring scheme. *Statistics in Transition New Series.* 24(3): 257-263. doi: http://dx.doi.org/10.59170/stattrans-2023-047

Schafer, J.L., and Graham, J.W. (2002). Missing data: Our view of the state of the art, *the American Psychological Association,* 7(2):147–177. *https://psycnet.apa.org/doi/10.1037/1082-989X.7.2.147*

Seunggeun, H., Jimin, L. and Robert Y. (2016). Parameter estimation of type-I and type-II hybrid censored data from the log-logistic distribution. *Industrial and Systems Engineering Review,* 4(1): 37–44. doi:10.37266/ISER.2016v4i1.pp37-44

Yu, D., Jin, L., Li, J., Qin, X., Zhu, Z., and Zhang, J. (2023). Monitoring the weibull scale parameter based on type I censored data using a modified EWMA control chart. *Axioms*, 12 (5), 487. doi.org/10.3390/axioms12050487

Appendix (A)

The derivation of equations and the proof of properties are performed in this appendix.

A1:The estimator of parameter θ in the case of one population:

$$L(\theta) = \prod_{i=1}^{n}((1-\theta)^{z_i-1}\theta)^{A_i}((1-\theta)^{z_i})^{B_i} \tag{A1.1}$$

Hence, the logarithm of the likelihood function is given by

$$\ln L(\theta) = \sum_{i=1}^{n}\left[A_i\big((z_i-1)\ \ln(1-\theta)+\ln\theta\big)+B_i z_i \ln(1-\theta)\right] \tag{A1.2}$$

$$\frac{\partial(\ln L(\theta))}{\partial\theta} = \sum_{i=1}^{n}\left[\left(\frac{A_i}{\theta}-\frac{A_i(z_i-1)}{1-\theta}\right)-\frac{B_i z_i}{1-\theta}\right] \tag{A1.3}$$

$$\frac{(1-\theta)\sum_{i=1}^{n}A_i-\theta\sum_{i=1}^{n}A_i(z_i-1)-\theta\sum_{i=1}^{n}B_i z_i}{\theta(1-\theta)} = 0 \tag{A1.4}$$

$$\sum_{i=1}^{n}A_i-\theta\sum_{i=1}^{n}A_i-\theta\sum_{i=1}^{n}A_i z_i+\theta\sum_{i=1}^{n}A_i-\theta\sum_{i=1}^{n}B_i z_i = 0 \tag{A1.5}$$

$$\sum_{i=1}^{n}A_i-\theta\left(\sum_{i=1}^{n}A_i z_i+\sum_{i=1}^{n}B_i z_i\right) = 0 \tag{A1.6}$$

$$\sum_{i=1}^{n}A_i = \theta\left(\sum_{i=1}^{n}A_i z_i+\sum_{i=1}^{n}B_i z_i\right) \tag{A1.7}$$

The estimator of parameter θ:

$$\hat{\theta} = \frac{\sum_{i=1}^{n}A_i}{\sum_{i=1}^{n}(A_i+B_i)z_i} \tag{A1.8}$$

A2: The estimator of parameter θ in the case of two populations:

$$L(\theta) = \prod_{i=1}^{n}((1-\theta_1)^{z_i-1}\theta_1)^{A_i}((1-\theta_1)^{z_i})^{B_i}\prod_{i=1}^{n}((1-\theta_2)^{M_i-1}\theta_1)^{C_i}((1-\theta_1)^{M_i})^{D_i} \tag{A2.1}$$

Hence, the logarithm of the likelihood function is given by:

$$\ln L(\theta)=\sum_{i=1}^{n}\left[A_i\big((z_i-1)\ \ln(1-\theta)+\ln\theta\big)+B_i z_i\ln(1-\theta)\right]$$
$$+\sum_{i=1}^{n}\left[C_i\big((M_i-1)\ \ln(1-\theta)+\ln\theta\big)+D_i M_i\ln(1-\theta)\right] \quad \text{(A2.2)}$$

$$\frac{\partial(\ln L(\theta))}{\partial\theta}=\sum_{i=1}^{n}\left[\left(\frac{A_i}{\theta}-\frac{A_i(z_i-1)}{1-\theta}\right)-\frac{B_i z_i}{1-\theta}\right]$$
$$+\sum_{i=1}^{n}\left[\left(\frac{C_i}{\theta}-\frac{C_i(M_i-1)}{1-\theta}\right)-\frac{D_i M_i}{1-\theta}\right] \quad \textit{(A2.3)}$$

$$\frac{(1-\theta)\sum_{i=1}^{n}A_i-\theta\sum_{i=1}^{n}A_i(z_i-1)-\theta\sum_{i=1}^{n}B_i z_i}{\theta(1-\theta)}$$
$$+\frac{(1-\theta)\sum_{i=1}^{n}C_i-\theta\sum_{i=1}^{n}C_i(M_i-1)-\theta\sum_{i=1}^{n}D_i M_i}{\theta(1-\theta)}=0 \quad \textit{(A2.4)}$$

$$\sum_{i=1}^{n}A_i-\theta\sum_{i=1}^{n}A_i-\theta\sum_{i=1}^{n}A_i z_i+\theta\sum_{i=1}^{n}A_i-\theta\sum_{i=1}^{n}B_i z_i+\sum_{i=1}^{n}C_i-\theta\sum_{i=1}^{n}C_i$$
$$-\theta\sum_{i=1}^{n}C_i M_i+\theta\sum_{i=1}^{n}C_i-\theta\sum_{i=1}^{n}D_i M_i=0 \quad \textit{(A2.5)}$$

$$\sum_{i=1}^{n}A_i-\theta\left(\sum_{i=1}^{n}A_i z_i+\sum_{i=1}^{n}B_i z_i\right)+\sum_{i=1}^{n}C_i-\theta\left(\sum_{i=1}^{n}C_i M_i+\sum_{i=1}^{n}D_i M_i\right)=0 \quad \textit{(A2.6)}$$

The estimator of parameter θ:

$$\hat{\theta}=\frac{\sum_{i=1}^{n}A_i+\sum_{i=1}^{n}C_i}{\sum_{i=1}^{n}(A_i+B_i)z_i+\sum_{i=1}^{n}(C_i+D_i)M_i} \quad \textit{(A2.7)}$$

$$\hat{\theta}=\frac{\frac{1}{2}\left(\sum_{i=1}^{n}\alpha_i^2\delta_i^2+\sum_{i=1}^{n}\alpha_i\delta_i+\sum_{i=1}^{n}\beta_i^2\eta_i^2+\sum_{i=1}^{n}\beta_i\eta_i\right)}{\sum_{i=1}^{n}\alpha_i^2\delta_i^2\,z_i+\sum_{i=1}^{n}\beta_i^2\eta_i^2\,M_i} \quad \text{(A2.8)}$$

A3: Properties of Estimators

$$E(\alpha_i\delta_i)=E(\alpha_i)E(\delta_i)=p[(1)P(x_1\le T_0)+(-1)P(x_1>T_0)] \quad \textit{(A3.1)}$$

$$p(x_1\le T_0)=\sum_{x_1=1}^{T_0}(1-\theta)^{x_1-1}\theta=\theta[1+(1-\theta)+(1-\theta)^2+\cdots+(1-\theta)^{T_0-1}]$$
$$=\theta_1\left[\frac{1-(1-\theta)^{T_0}}{1-(1-\theta)}\right]=1-(1-\theta)^{T_0} \quad \textit{(A3.2)}$$

where

$$a+ar+ar^2+\cdots+ar^{n-1}=\sum_{k=0}^{n-1}ar^k=a\frac{1-r^n}{1-r}\qquad r\neq 1 \quad \textit{(A3.3)}$$

and

$$P(x_1 > T_0) = \sum_{x=T_0+1}^{\infty} (1-\theta)^{x_1-1}\theta$$
$$= \theta[(1-\theta)^{T_0} + (1-\theta)^{T_0+1} + (1-\theta)^{T_0+2} + \cdots + (1-\theta)^{\infty}] \quad (A3.4)$$
$$= 1 - p(x_1 \le T_0) = (1-\theta)^{T_0}$$

$$E(\alpha_i\delta_i) = E(\alpha_i)E(\delta_i) = p[1 - 2(1-\theta)^{T_0}] \quad (A3.5)$$

So

$$\frac{1}{n}\sum_{i=1}^{n} \alpha_i\delta_i \xrightarrow{a.s.} p[1 - 2(1-\theta)^{T_0}] \quad (A3.6)$$

Similarly,

$$\frac{1}{n}\sum_{i=1}^{n} \alpha_i^2\delta_i^2 \xrightarrow{a.s.} E(\alpha_i^2\delta_i^2), \quad (A3.7)$$

$$E(\alpha_i^2\delta_i^2) = E(\alpha_i^2)E(\delta_i^2) = [(1)^2 p(x_1 \le T_0) + (-1)^2 P(x_1 > T_0)]\ p$$
$$= [1 - (1-\theta)^{T_0} + (1-\theta)^{T_0}]p = p \quad (A3.8)$$

So

$$\frac{1}{n}\sum_{i=1}^{n} \alpha_i^2\delta_i^2 \xrightarrow{a.s.} p \quad (A3.9)$$

and

$$\frac{1}{n}\sum_{i=1}^{n} \alpha_i^2\delta_i^2\, z_i \xrightarrow{a.s.} E(\alpha_i^2\delta_i^2 z_i) \quad (A3.10)$$

$$E(\alpha_i^2\delta_i^2 z_i) = E(\alpha_i^2\delta_i^2)E(z_i) = p\, E(z_i = min\,(T_0, x_i)) \quad (A3.11)$$

$$E(z_i) = \sum_{z_i=1}^{T_0} z_i(1-\theta)^{Z_i-1}\theta$$
$$= \theta[1 + 2(1-\theta) + 3(1-\theta)^2 + \cdots + T_0(1-\theta)^{T_0-1}] \quad (A3.12)$$

Let

$$S = 1 + 2x + 3x^2 + 4x^3 + \ldots\ldots\ldots\ldots + nx^{(n-1)} \quad (A3.13)$$

Multiply both sides by ‘x’

$$xS = x + 2x^2 + 3x^3 + \ldots\ldots\ldots\ldots + nx^n \quad (A3.14)$$

Subtract xS from S,

$$S\,(1 - x) = 1 + x + x^2 + \ldots\ldots + x^{n-1} - nx^n \quad (A3.15)$$

except for the last term, the rest is a geometric progression with the first term ‘1’. Where

$$1 + x + x^2 + \cdots + x^{n-1} = \frac{1 - x^n}{1 - x} \quad (A3.16)$$

$$S(1-x) = (1 - x^n)/(1 - x) - n\,x^n \quad (A3.17)$$

$$S = \frac{(1-x^n)}{(1-x)^2} - \frac{n\,x^n}{(1-x)} \tag{A3.18}$$

So

$$(z_i) = \theta\left[\frac{(1-(1-\theta)^{T_0})}{\left(1-(1-\theta)\right)^2} - \frac{T_0\,(1-\theta)^{T_0}}{\left(1-(1-\theta)\right)}\right]$$

$$= \theta\left[\frac{(1-(1-\theta)^{T_0})}{\theta^2} - \frac{T_0\,(1-\theta)^{T_0}}{\theta}\right] \tag{A3.19}$$

$$= \frac{1}{\theta}\left[(1-(1-\theta)^{T_0}) - (T_0(1-\theta)^{T_0}\theta)\right]$$

$$= \frac{1}{\theta}(1-(1-\theta)^{T_0}) - T_0(1-\theta)^{T_0}$$

$$\frac{1}{n}\sum_{i=1}^{n}\alpha_i^2\delta_i^2\, z_i \xrightarrow{a.s.} p\left[\frac{1}{\theta}(1-(1-\theta)^{T_0}) - T_0(1-\theta)^{T_0}\right] \tag{A3.20}$$

For simplicity, assume that $T_0(1-\theta)^{T_0}$ tends to zero, so

$$\frac{1}{n}\sum_{i=1}^{n}\alpha_i^2\delta_i^2\, z_i \xrightarrow{a.s.} \frac{p}{\theta}(1-(1-\theta)^{T_0}) \tag{A3.21}$$

The estimator of θ:

$$\hat{\theta} = \frac{\sum_{i=1}^{n} A_i}{\sum_{i=1}^{n}(A_i + B_i) z_i} = \frac{\sum_{i=1}^{n}\alpha_i^2\delta_i^2 + \sum_{i=1}^{n}\alpha_i\delta_i}{2\sum_{i=1}^{n}\alpha_i^2\delta_i^2\, z_i} \tag{A3.22}$$

Multiply equation (54) by 1/n:

$$\hat{\theta} = \frac{\frac{1}{n}\left(\sum_{i=1}^{n}\alpha_i^2\delta_i^2 + \sum_{i=1}^{n}\alpha_i\delta_i\right)}{\frac{2}{n}\sum_{i=1}^{n}\alpha_i^2\delta_i^2\, z_i} \tag{A3.23}$$

By substitution

$$\hat{\theta} = \frac{p + p(1-2(1-\theta)^{T_0})}{2\frac{p}{\theta}(1-(1-\theta)^{T_0})} = \frac{p(2-2(1-\theta)^{T_0})}{2\frac{p}{\theta}(1-(1-\theta)^{T_0})} \tag{A3.24}$$

$$\hat{\theta} = \frac{2(1-(1-\theta)^{T_0})}{2\frac{1}{\theta}(1-(1-\theta)^{T_0})} \xrightarrow{a.s.} \theta \tag{A3.25}$$

Appendix (B)

This appendix contains the graphs for the bias and the root mean square error.

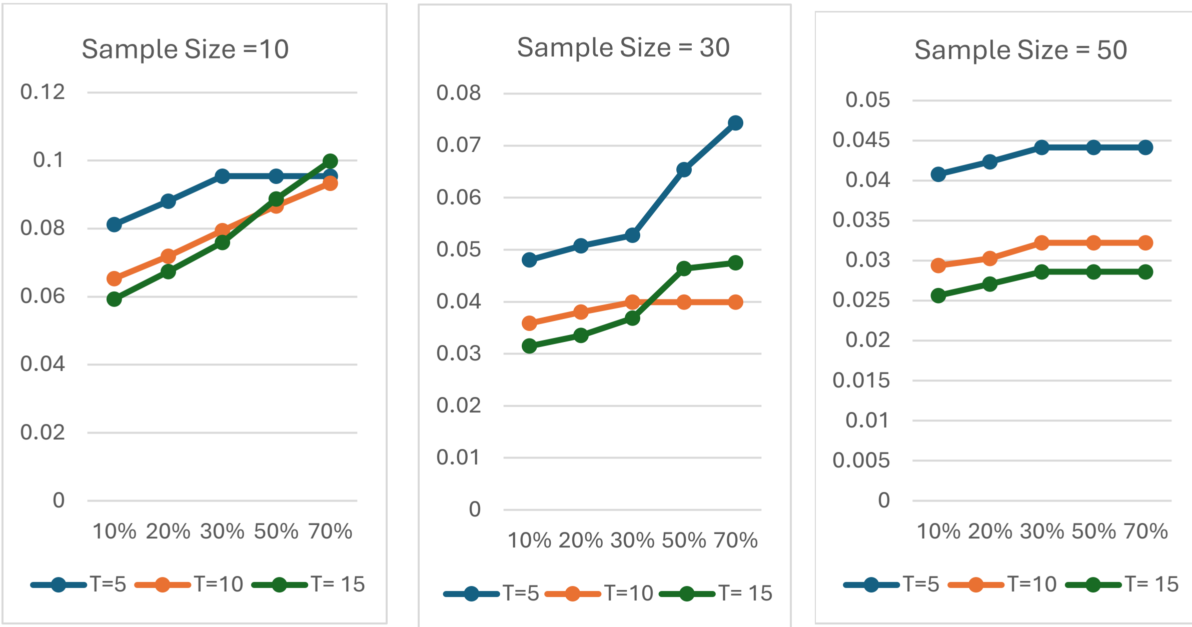

Figure B.1. The Root Mean Square Error of θ_1 for One Population

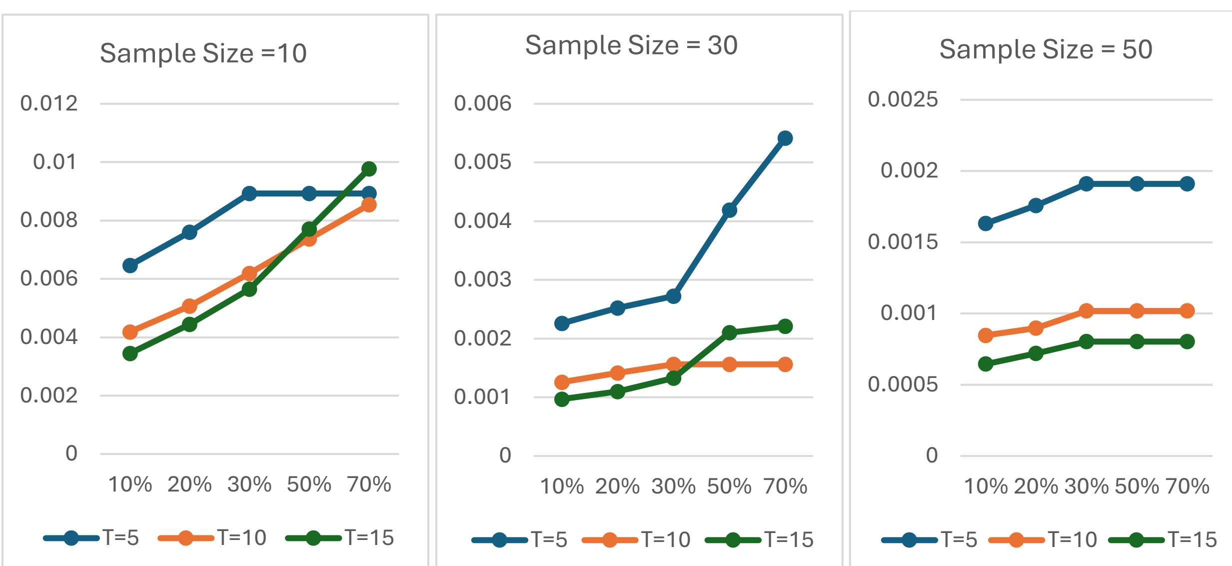

Figure B.2. The Bias of θ_1 for One Population

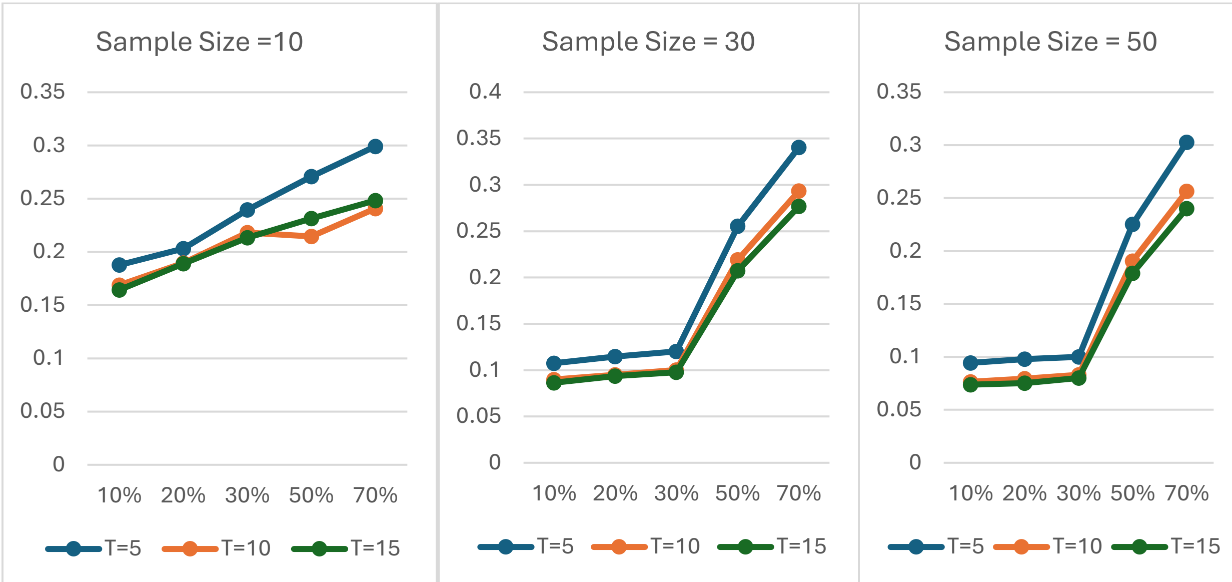

Figure B.3. The Root Mean Square Error of θ_2 for One Population

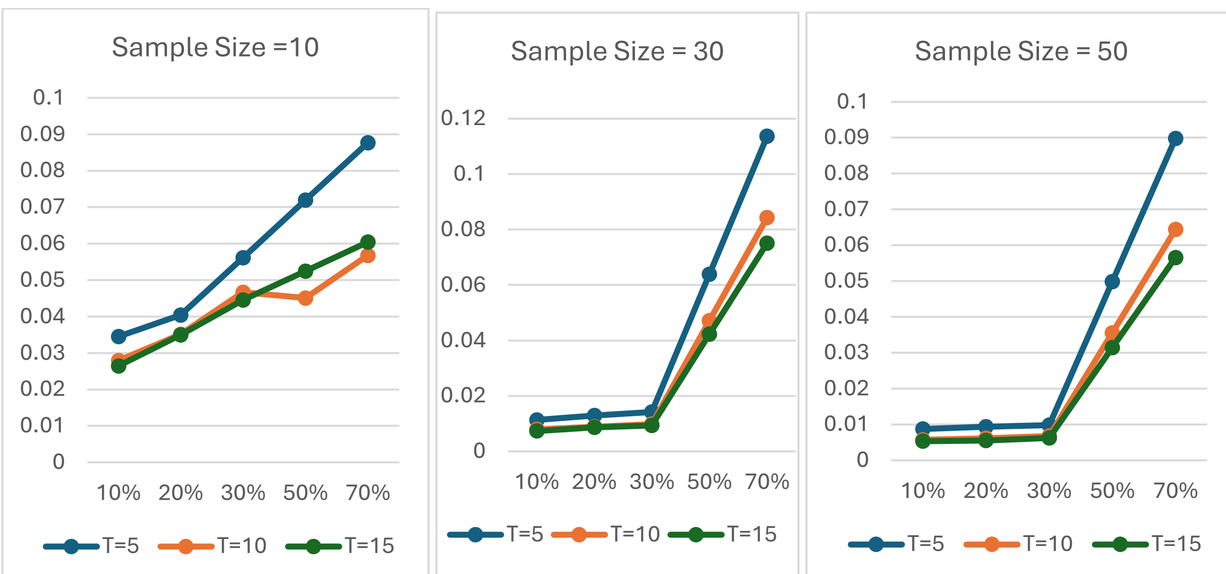

Figure B.4. The Bias of θ_2 for One Population

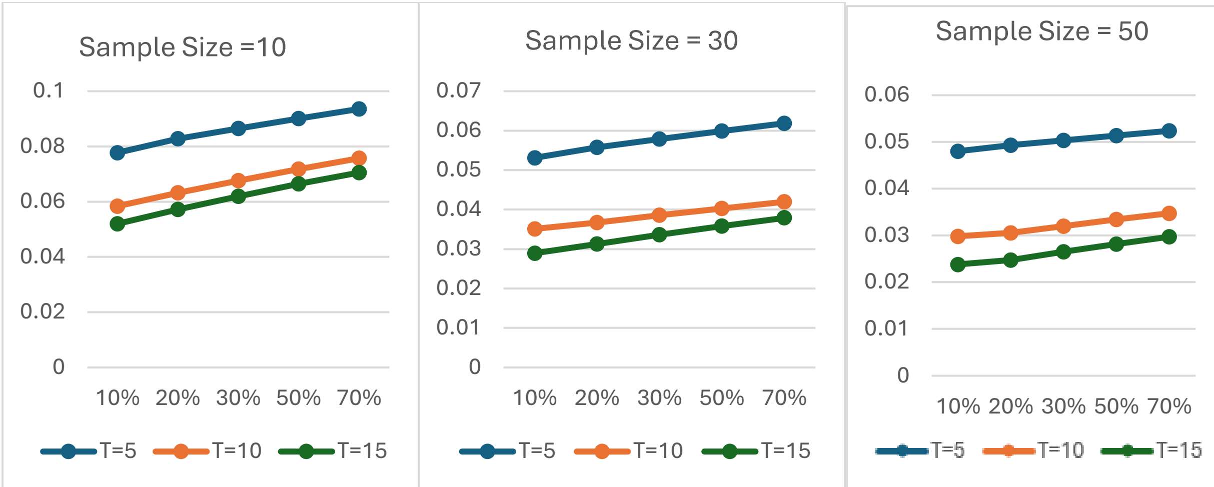

Figure B.5. The Root Mean Square Error of θ for Two Populations

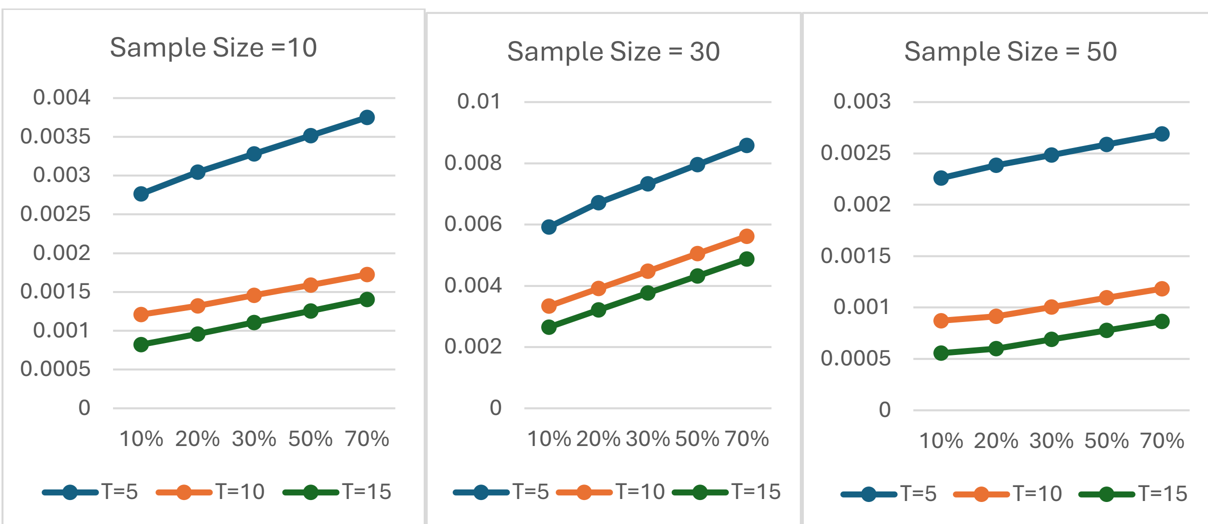

Figure B.6. The Bias of θ for Two Populations