UNIVERSIDAD AUTÓNOMA AGRARIA ANTONIO NARRO

SUBDIRECCIÓN DE POSTGRADO

DESIGUALDADES PROBABILÍSTICAS Y NORMALIDAD

ASINTÓTICA EN TEORÍA DE MUESTREO

Tesis

Que presenta IMELDA CAROLINA CERDA DELGADO

como requisito parcial para obtener el Grado de

MAESTRA EN ESTADÍSTICA APLICADA

Saltillo, Coahuila　　　　Octubre de 2020

DESIGUALDADES PROBABILÍSTICAS Y NORMALIDAD ASINTÓTICA EN TEORÍA DE MUESTREO

Tesis

Elaborada por IMELDA CAROLINA CERDA DELGADO como requisito parcial para obtener el Grado de MAESTRA EN ESTADÍSTICA APLICADA con la supervisión y aprobación del Comité de Asesoría

Dr. Rolando Cavazos Cadena
Asesor Principal

Dr. Mario Cantú Sifuentes
Asesor

M. C. Luis Rodríguez Gutiérrez
Asesor

Dr. Marcelino Cabrera De la Fuente
Subdirector de Postgrado
UAAAN

Saltillo, Coahuila

Octubre de 2020

Acknowledgement

A Dios

Por permitirme llegar hasta esta etapa de mi vida

A todas las personas que contribuyeron en la culminación de este trabajo de investigación, especialmente a mis profesores, Dr. Rolando Cavazos Cadena y Dr. Mario Cantú Sifuentes, por compartir sus conocimientos, además del tiempo, espacio y dedicación otorgados para que este trabajo pudiera finalizarse.

Dedication

A mi familia, que siempre ha estado conmigo, acompañándome
y brindándome su apoyo incondicional en cada momento.

A mi padre, Francisco Javier Cerda García

A mi madre, Gloria Rocío Delgado Pérez

A mi hermano, Francisco Javier Cerda Delgado

COMPENDIO

DESIGUALDADES PROBABILÍSTICAS Y NORMALIDAD ASINTÓTICA EN TEORÍA DE MUESTREO

Por

IMELDA CAROLINA CERDA DELGADO

MAESTRÍA EN

ESTADÍSTICA APLICADA

UNIVERSIDAD AUTÓNOMA AGRARIA
ANTONIO NARRO

BUENAVISTA, SALTILLO, COAHUILA, Octubre de 2020

Dr. Rolando Cavazos Cadena –Asesor–

Palabras clave: Desigualdades probabilísticas, Convergencia en probabilidad, Convergencia en distribución, Normalidad asintótica, Desigualdades de Markov y Chebishev, Estimador de razón, Muestreo aleatorio simple.

Este trabajo trata sobre dos ideas básicas en Estadística y su aplicación a la Teoría del Muestreo: (i) Desigualdades probabilísticas, las cuales proporcionan una cota para la probabilidad de que una variable aleatoria tome valores 'grandes' por medio de los momentos de la variable, y (ii) Normalidad asintótica, una propiedad que desempeña un papel importante en la construcción de intervalos de confianza y en la determinación de tamaños de muestra. En el Capítulo 1 se presenta una breve descripción del trabajo y la organización del material subsecuente, mientras que el Capítulo 2 trata sobre las desigualdades de Markov y Chebishev; en este punto se presenta la principal contribución de este trabajo, a saber, la determinación de condiciones necesarias y suficientes para que las cotas proporcionadas por esas desigualdades coincida con la probabilidad bajo estudio. Luego, en el Capítulo 3 se estudian las ideas de convergencia en probabilidad y en distribución, mientras que en el Capítulo 4 se analiza la noción de normalidad asintótica y su invariancia bajo la aplicación de transformaciones diferenciables, resultado que se utiliza en el Capítulo 5 para deteminar, bajo el esquema de muestreo aleatorio simple, la distribución límite del estimador de razón del total poblacional.

ABSTRACT

PROBABILITY INEQUALITIES AND ASYMPTOTIC NORMALITY IN SAMPLING THEORY

BY

IMELDA CAROLINA CERDA DELGADO

MASTER IN

APPLIED STATISTICS

UNIVERSIDAD AUTÓNOMA AGRARIA
ANTONIO NARRO

BUENAVISTA, SALTILLO, COAHUILA, October, 2020

Dr. Rolando Cavazos Cadena –Advisor–

Key Words: Probability inequalities, Convergence in probability, Convergence in distribution, Asymptotic normality, Markov and Chebishev inequalities, Ratio estimator, Simple random sampling

This work is about two basic statistical ideas applied to *Sampling Theory*: (i) Probability inequalities, setting a bound on the probability that a random variable attains large values in terms of the corresponding moments, and (ii) Asymptotic normality, a property that plays an important role in the construction of confidence intervals and the determination of sample sizes. In Chapter 2 Markov and Chebishev inequalities are studied, and the *main contribution* of this work is presented, namely, necessary and sufficient conditions are given so that the bounds provided by the aforementioned relations coincide with the probability under consideration. In Chapter 3 the ideas of convergence in probability, convergence in distribution as well as the (weak) law of large numbers are discussed. Next, Chapter 4 concerns with asymptotic normality and the invariance of this property under differentiable transformations, result that is finally used in Chapter 5 to analyze, under simple random sampling, the limit distribution of the *ratio estimator* of the population total, and to compare it with the classical estimator.

Contents

Chapter 1

Perspective

1.1. Introduction

This work concerns with basic results in *Classical Statistics*, which are discussed in the context of the Theory of Sampling. The exposition includes two main topics: (i) Probability inequalities, setting bounds on the tails of a probability distributions in terms of moments, and (ii) Asymptotic Normality, a property stating that under mild conditions the distribution of a standardized average can be approximated by a normal distribution. The main technical *contribution* of this work concerns the first point: necessary and sufficient conditions are given so that Markov and Chebishev inequalities are sharp, that is, the bound for the probability of large values and the probability itself coincide. Concerning the second point, the basic invariance property is studied, which establishes that if an asymptotically normal sequence is transformed by a differentiable function, then the modified sequence is also asymptotically normal. A very important point of the analysis is the formula relating the new asymptotic variance with the original one. Such a relation will be used to determine the limit variance of the ratio estimator of the population total, a result that allows to make a comparison with the classical estimator for the simple random sampling scheme. In the following lines, the origin of this work is briefly described, and the organization of the subsequent material is outlined.

1.2. The Origin of This Work

This work is a byproduct of the seminar entitled *Mathematical Statistics: Elements of Theory and Examples*, relaunched on July 2016 by the Graduate Program in Statistics at the Universidad Autónoma Agraria Antonio Narro. The basic aims of the project are:

(i) To be a framework were statistical problems can be freely and fruitfully discussed;

(ii) To promote the *understanding* of basic statistical and analytical tools through the analysis and detailed solution of exercises.

(iii) To develop the *writing skills* of the participants, generating an organized set of neatly solved examples, which can used by other members of the program, as well as by the statistical communities in other institutions and countries.

(iv) To develop the *communication skills* of the students and faculty through the regular participation in seminars, were the results of their activities are discussed with the members of the program.

The activities of the seminar are concerned with fundamental statistical theory at an intermediate (non-measure theoretical) level, as in the book *Mathematical Statistics* by Dudewicz and Mishra (1998). When necessary, other more advanced references that have been useful are Lehmann and Casella (1998), Borobkov (1999) and Shao (2002), whereas deeper probabilistic aspects have been studied in the classical text by Loève (1984). On the other hand, statistical analysis requires algebraic and analytical tools, and the basic references on these disciplines are Apostol (1980), Fulks (1980), Khuri (2002) and Royden (2003), which concern mathematical analysis, whereas the algebraic aspects are covered in Graybill (2000, 2001) and Harville (2008). Initially, the project was concerned with the theory of Point Estimation and Hypothesis Testing. During the last two years the seminar has been focused on Sampling Theory at the level of Lohr (2000), Tucker (1992), Hansen *et al.* (2002), and Sarndal *et al.* (1992).

1.3. The Organization

The remainder of this work has been organized as follows: Chapter 2 is concerned with Markov and Chebishev inequalities, relations that set bounds on the tails of a probability distribution via the moments of positive order. Those results are valid under minimal conditions and, practically, are universally valid. However, such a generality naturally implies that the bounds provided by the aforementioned relations are not sharp, that is, generally the bounds are 'far from' the true value. The *main contribution* of this work consists in presenting a detailed analysis on the necessary and sufficient conditions so that the bounds provided by Markov and Chebishev inequalities coincide with the probability under consideration.

In Chapter 3 two ideas of convergence of random variables are discussed, and the relation between these two concepts is studied. The presentation includes the notion of convergence in probability and the concept of convergence in the mean, and the relation between them is analyzed . Also, the (weak) law of large numbers is stated , and the discussion is applied to to a basic problem in problem in survey sampling, namely determining the sample size required to achieve certain precision/confidence combination.

In Chapter 4 , the idea of asymptotic normality is studied, and it is shown that the Central Limit Theorem allows to determine sample sizes which, in spite of being substantially smaller than those obtained via Chebishev inequality, are sufficient to ensure that desired precision with a given confidence level. It is shown that asymptotic normality is preserved under the application of smooth (differentiable) functions.

Finally, the exposition concludes in Chapter 5 where, using the the results previously analyzed, the usual estimator of the population total under simple random sampling is compared with the ratio estimator.

Chapter 2

Probability Inequalities

2.1. Introduction

%vskip -.25in This chapter is concerned with two results about the the tails of a probability distribution, namely, Markov and Chebishev inequalities. The first relation concerns a nonnegative random variable, and sets an upper bound the probability of attaining large values in terms of moments of positive order, whereas the second inequality uses the variance to provide an upper bound for the probability of large deviations of a random variable about its mean. As it will be discussed below, these results are valid under minimal conditions and, practically, are universally valid. However, such a generality naturally implies that the bounds provided by the aforementioned relations are not sharp, that is, generally they are 'far from' the true value. In this chapter the problem of determining necessary and sufficient conditions so that the bounds provided by Markov and Chebishev inequalities coincide with the probability under consideration. The subsequent material has been organized as follows: In Section 2 Markov and Chebishev inequalities are established, whereas in Section 3 examples are used to show that, in general, the strict inequality holds in Markov and Chebishev relations,; however, an example is given to show that the bound in those relations can be attained in some cases. Next, in Section 4. necessary and sufficient conditions are determined so that the bound in Markov inequality coincides with the probability under consideration, whereas a similar result is obtained in Section 5 for Chebishev inequality.

2.2. Chebishev and Markov Inequalities

This section concerns with inequalities involving probability distributions. Using moments of a random variable X, the relations stated in the following theorem provide bounds for the probability of the event that X attains 'large values'.

Theorem 2.2.1. (i) If X is a nonnegative random variable, then for every $a, t \in (0, \infty)$

$$P[X \geq t] \leq \frac{E[X^a]}{t^a}; \tag{2.2.1}$$

this is *Markov inequality*.

(ii) If Y is a random variable with finite mean $\mu_Y = E[Y]$ and standard deviation σ_Y, then

$$P\left[|Y - \mu_Y| \geq k\sigma_Y\right] \leq \frac{1}{k^2}, \quad k > 0; \tag{2.2.2}$$

this relation is known as *Chebishev inequality*.

Proof. (i) Keeping in mind that the inequality $X \geq 0$ always holds, observe that $X^a \geq t^a$ if $X \geq t$, and $X^a \geq 0$ if $X < t$. Using the notation of indicator functions, these relations can be expressed as

$$X^a \geq t^a I[X \geq t] + 0I[X < t^a] = t^a I[X \geq t].$$

Via the monotonicity of the expectation operator this relation leads to

$$E[X^a] \geq t^a E[I[X \geq t]] = t^a P[X \geq t],$$

and (2.2.1) follows.

(ii) Set $X = |Y - \mu_Y|$, $t = k\sigma_Y$ and $a = 2$. Applying part (i) with his data, it follows that

$$\begin{aligned} P[|Y - \mu_Y| \geq k\sigma_Y] &\leq \frac{E[|Y - \mu_Y|^2]}{(k\sigma_Y)^2} \\ &= \frac{E[(Y - \mu_Y)^2]}{k^2\sigma_Y^2} \\ &= \frac{\sigma_Y^2}{k^2\sigma_Y^2} = \frac{1}{k^2}, \end{aligned}$$

completing the argument. □

Slightly different formulation of the inequalities in the above theorem are discussed below.

Remark 2.2.1. Alternative forms of Markov and Chebishev inequalities.

(i) Using the basic property $P[A^c] = 1 - P[A]$ for any event A, note that the inequality $P[A] \leq b$ is equivalent to $P[A^c] \geq 1 - b$, so that (2.2.1) and (2.2.2) can be equivalently written as

$$P[X < t] \geq 1 - \frac{E[X^a]}{t^a}, \tag{2.2.3}$$

and

$$P\left[|Y - \mu_Y| < k\sigma_Y\right] \geq 1 - \frac{1}{k^2}, \quad k > 0, \tag{2.2.4}$$

relations that are alternative forms of Markov and Chebishev inequalities, respectively.

(ii) Given $t > 0$, consider a sequence $\{t_n\} \subset (0, \infty)$ such that $t_n \searrow t$, that is, $t_n \geq t_{n+1}$ for every n and $\lim_{n\to\infty} t_n = t$. In this case, using that a distribution function is continuous form the right, it follows that

$$\lim_{n\to\infty} P[X < t_n] = P[X \leq t];$$

see, for instance, Dudewicz and Mishra (1988). Applying (2.2.3) with t_n instead of t it follows that

$$\begin{aligned}\lim_{n\to\infty} P[X < t_n] &\geq \lim_{n\to\infty}\left(1 - \frac{E[X^a]}{t_n^a}\right)\\ &= 1 - \lim_{n\to\infty}\frac{E[X^a]}{t_n^a}\\ &= 1 - \frac{E[X^a]}{t^a}.\end{aligned}$$

Combining these two last displays it follows that

$$P[X \leq t] \geq 1 - \frac{E[X^a]}{t^a}, \quad t > 0. \tag{2.2.5}$$

Similarly, starting form (2.2.4), it follows that

$$P\left[|Y - \mu_Y| \leq k\sigma_Y\right] \geq 1 - \frac{1}{k^2}, \quad k > 0. \tag{2.2.6}$$

Again, these relations are alternative forms of Markov and Chebichev inequalities.

(iii) If Y is a random variable with finite mean μ_Y and standard deviation σ_Y, then the corresponding standardized random variable Y^* is given by

$$Y^* = \frac{Y - \mu_Y}{\sigma_Y}. \tag{2.2.7}$$

Observing that, for every $k > 0$,

$$|Y^*| \geq k \iff |Y - \mu_Y| \geq k\sigma_Y \quad \text{and} \quad |Y^*| \leq k \iff |Y - \mu_Y| \leq k\sigma_Y$$

form (2.2.2) and (2.2.6) it follows that

$$P[|Y^*| \geq k] \leq \frac{1}{k^2} \quad \text{and} \quad P[|Y^*| \leq k] \geq 1 - \frac{1}{k^2}; \tag{2.2.8}$$

Observe that for $k \leq 1$ these inequalities convey no information.

(iv) Chebishev inequality can be expressed verbally as follows:

The probability that the random variable Y deviates from its mean by k standard deviations or more is less that $1/k^2$.

For instance, $P[|Y - \mu_Y| \geq 4\sigma_Y] \leq 1/4^2 = 1/16 = .0625$ and $P[|Y - \mu_Y| \geq 5\sigma_Y] \leq 1/5^2 = 1/25 = .04$. □

The importance of Theorem 2.2.1 stems form its generality: Chebishev and Markov inequalities are *always* valid for any random variable with finite moments of the appropriate order.

2.3. Two Examples

The generality behind the conclusions of Theorem 2.2.1 has a price, namely, the bounds in Markov and Chebishev inequalities *are not sharp*, that is, *usually* in (2.2.1) and (2.2.2) the inequality is strict. In this section an example will be used to illustrate this phenomenon, and an additional example will be used to show that, *eventually*, it is possible to observe that the equality occurs. To continue note that, regardless of the distribution of X, Chebishev inequality (2.2.2) with $k = 2$ and $k = 3$ yields that

$$P[|X - \mu_X| \geq 2\sigma_X] \leq \frac{1}{2^2} = \frac{1}{4} \quad \text{and} \quad P[|X - \mu_X| \geq 3\sigma_X] \leq \frac{1}{3^2} = \frac{1}{9}. \tag{2.3.1}$$

Example 2.3.1. For $k = 2, 3$, the probability $P[\,|X - \mu_X| \geq k\sigma_X]$ will be computed for several known distributions, and it will be verified that the strict inequalities occur in the above display.

(a) If $X \sim \mathcal{N}(0, 1)$ then $\mu_X = 0$ and $\sigma_X = 1$, so that

$$P[|X - \mu_X| \geq 2\sigma_X] = 0.04550026 < \frac{1}{4} \quad \text{and} \quad P[|X - \mu_X| \geq 3\sigma_X] = 0.002699796 < \frac{1}{9}.$$

(b) If $X \sim \mathcal{P}(4)$ (Poisson distribution with mean 4), then $\mu_X = \lambda = 4$ and $\sigma_X^2 = \lambda = 4$, so that $\sigma_X = 2$ and then

$$\begin{aligned} P[|X - \mu_X| \geq 2\sigma_X] &= P[|X - 4| \geq 2(2)] \\ &= P[|X - 4| \geq 4] \\ &= [PX \geq 8] + P[X = 0] \\ &= 0.05113362 + 0.01831564 = 0.06944925 < \frac{1}{4}, \end{aligned}$$

whereas

$$\begin{aligned} P[|X - \mu_X| \geq 3\sigma_X] &= P[|X - 4| \geq 3(2)] \\ &= P[|X - 4| \geq 6] \\ &= [PX \geq 10] = 0.002839766 < \frac{1}{9}. \end{aligned}$$

(c) If $X \sim \mathcal{B}(10, 0.4)$ (Binomial distribution with $n = 10$ repetitions and success probability $p = 0.4$), then $\mu_X = 10(0.4) = 4$ and $\sigma_X^2 = 10(0.4)(0.6) = 2.4$, so that $\sigma_X = 1.549193$ and then

$$\begin{aligned} P[|X - \mu_X| \geq 2\sigma_X] &= P[|X - 4| \geq 2(1.549193)] \\ &= P[|X - 4| \geq 3.098387] \\ &= [PX \geq 8] + P[X = 0] \\ &= 0.01229455 + 0.006046618 = 0.01834117 < \frac{1}{4}; \end{aligned}$$

also

$$\begin{aligned} P[|X - \mu_X| \geq 3\sigma_X] &= P[|X - 4| \geq 3(1.549193)] \\ &= P[|X - 4| \geq 4.64758] \\ &= P[X \geq 9] = 0.001677722 < \frac{1}{9}. \end{aligned}$$

In all of the above cases the strict inequalities hold in (2.2.2), showing explicitly that, generally, $P[|X - \mu_X| \geq k\sigma_X]$ is strictly less than $1/k^2$. □

In contrast with the previous example, the following one shows that the equality *may* hold in Chebishev inequality, that is, $P[\,|X - \mu_X| \geq k\sigma_X]$ may be equal to the upper bound $1/k^2$ in (2.2.2).

Example 2.3.2. Let the distribution of the random variable X be determined by

$$f_X(x) = \begin{cases} 1/8 & \text{if } x = -1, \\ 6/8 & \text{if } x = 0, \\ 1/8 & \text{if } x = 1. \end{cases}$$

In this context, it will be shown that Chevishev inequality becomes an equality when $k = 2$. To achieve this goal, note that

$$\mu_X = E[X] = -1\frac{1}{8} + 0\frac{6}{8} + 1\frac{1}{8} = 0,$$

whereas

$$E[X^2] = (-1)^2\frac{1}{8} + 0^2\frac{6}{8} + (1)^2\frac{1}{8} = \frac{2}{8} = \frac{1}{4},$$

and then

$$\sigma_X^2 = E[X^2] - (E[X])^2 = \frac{1}{4}, \quad \sigma_X = \frac{1}{2}.$$

Therefore, Chebishev inequality with $k = 2$ yields that

$$P[|X| \geq 1] = P\left[|X - 0| \geq 2\left(\frac{1}{2}\right)\right] = P\left[|X - \mu_X| \geq k\sigma_X\right] \leq \frac{1}{k^2} = \frac{1}{4},$$

since

$$P[|X| \geq 1] = P[X = 1] + P[X = -1] = 1/8 + 1/8 = 1/4,$$

it follows that equality holds in Chebishev inequality with $k = 2$. □

2.4. Conditions for Equality: I

The examples en the previous section naturally leads to the following question: When does the equality occur in Chebishev inequality? In this section necessary and sufficient conditions on the distribution of a random variable X are stated so that equality holds in the basic relations (2.2.1) and (2.2.2). The analysis starts with Markov inequality.

Theorem 2.4.1. Let X be a random variable such that $E[\,|X|^a] < \infty$ where $a > 0$. Given $t > 0$, the following conditions (i) and (ii) are equivalent:

(i) $P[|X| \geq t] = \dfrac{E[|X|^a]}{t^a}$

(ii) With probability 1, $|X|$ attains just the values 0 and t , that is, $P[|X| = t] + P[X = 0] = 1$.

The proof of this theorem is based on the following auxiliary result.

Lemma 2.4.1. For a given random variable X

$$X \geq 0 \quad \text{and} \quad E[X] = 0 \Rightarrow P[X = 0] = 1.$$

Proof. Suppose that $X \geq 0$ and $E[X] = 0$. In this case $X = |X|$, since $X \geq 0$, and given $n > 0$, Markov inequality yields that

$$P\,[X \geq 1/n] \leq \frac{E[X]}{1/n} = \frac{0}{1/n} = 0.$$

Thus, from $[X > 0] = \bigcup_{n=1}^{\infty}[X \geq 1/n]$ it follows that

$$P[X > 0] = \sum_{n=1}^{\infty} P[X \geq 1/n] = \sum_{n=1}^{\infty} 0 = 0,$$

so that $P[X \leq 0] = 1$ and then, recalling that $X \geq 0$, it follows that $P[X = 0] = 1$. □

Proof of Theorem 2.4.1. (i) $\Rightarrow$ (ii): Suppose that

$$P[|X| \geq t] = \frac{E[|X|^a]}{t^a}, \tag{2.4.1}$$

where a and t are positive real numbers. Note that

$$|X|^a \;=\; |X|^a I[|X| \geq t] + |X|^a I[t > |X|] \;\geq\; |X|^a I[|X| \geq t] \;\geq\; t^a I[|X| \geq t].$$

Using that the expectation operator is monotone, it follows that

$$\begin{aligned} E[X|^a] &= E[|X|^a I[|X| \geq t]] + E[|X|^a I[t > |X|]] \\ &\geq E[|X|^a I[|X| \geq t]] \\ &\geq t^a E[I[|X| \geq t]] \\ &= t^a P[|X| \geq t] \end{aligned}$$

where the relation $E[I[|X| \geq t]] = P[|X| \geq t]$ was used to set the last equality. Since $E[|X|^a] = t^a P[|X \geq t]$, by (2.4.1), the extreme terms in the the above display are equal, and then all of the terms coincide with $t^a P[|X| \geq t]$. Hence,

$$\begin{aligned} t^a P[|X| \geq t] &= E[X|^a] \\ &= E[|X|^a I[|X| \geq t]] + E[|X|^a I[t > |X|]] \\ &= E[|X|^a I[|X| \geq t]] \\ &= t^a E[I[|X| \geq t]] \\ &= t^a P[|X| \geq t]. \end{aligned} \tag{2.4.2}$$

The third equality in this display, which is given by

$$E[|X|^a I[|X| \geq t]] + E[|X|^a I[t > |X|]] = E[|X|^a I[|X| \geq t]],$$

is equivalent to $E[|X|^a I[t > |X|]] = 0$. From this point, applying Lemma 2.4.1 with $|X|^a I[t > |X|]$ instead of X, it follows that

$$P[|X|^a I[t > |X|] = 0] = 1. \tag{2.4.3}$$

Observe now that $|X|^a I[t > |X|] = 0$ is equivalent to $|X| = 0$ or $I[t > |X|] = 0$, and that $I[t > |X|] = 0$ means that the relation $t > |X|$ does not occur, and in that case $t \leq |X|$ holds. Therefore,

$$|X|^a I[t > |X|] = 0 \iff |X| = 0 \text{ or } t \leq |X|,$$

so that

$$[|X|^a I[t > |X|] = 0] = [|X| = 0] \cup [t \leq |X|]. \tag{2.4.4}$$

To continue, note that (2.4.3) implies that $P[A] = P[A \cap [|X|^a I[t > |X|] = 0]]$ for every event A; applying this equality with the event $A = [t > |X|]$, via the previous display it follows that

$$\begin{aligned} P[t > |X|] &= P[[t > |X|] \cap [|X|^a I[t > |X|] = 0]] \\ &= P[[t > |X|] \cap \{|X| = 0] \cup [t \leq |X|]\}] \\ &= P[\{[t > |X|] \cap |X| = 0]\} \cup \{[t > |X|] \cap [t \leq |X|]\}]; \end{aligned}$$

Since $[t > |X|] \cap [t \leq |X|] = \emptyset$ and $[t > |X|] \cap [|X| = 0] = [|X| = 0]$, the above display leads to

$$P[t > |X|] = P[|X| = 0]. \tag{2.4.5}$$

Next, the fourth equality in (2.4.2) states that $E[|X|^a I[|X| \geq t]] = t^a E[I[|X| \geq t]]$, so that

$$E[(|X|^a - t^a)I[|X| \geq t]] = 0.$$

Since the random variable $(|X|^a - t^a)I[|X| \geq t]$ is nonnegative, from Lemma 2.4.1 it follows that

$$P[(|X|^a - t^a)I[|X| \geq t] = 0] = 1. \tag{2.4.6}$$

Now, observe that $(|X|^a - t^a)I[|X| \geq t] = 0$ is equivalent to

$$(|X|^a - t^a) = 0 \quad \text{or} \quad I[|X| \geq t] = 0;$$

since $(|X|^a - t^a) = 0$ is equivalent to $|X| = t$ and $I[|X| \geq t] = 0$ means that the event $[|X| < t]$ occurs, it follows that

$$[|X| = t] \cup [|X| < t] = [(|X|^a - t^a)I[|X| \geq t] = 0].$$

On the other hand, note that (2.4.6) implies that $P[A] = P[A \cap [(|X|^a - t^a)I[|X| \geq t] = 0]]$ for every event A. Using this equality with the event $[|X| \geq t]$ instead of A, via the above display it follows that

$$\begin{aligned} P|X| \geq t] &= P[[|X| \geq t] \cap [(|X|^a - t^a)I[|X| \geq t] = 0]] \\ &= P[[|X| \geq t] \cap \{[|X| = t] \cup [|X| < t]\}] \\ &= P[\{[|X| \geq t] \cap [|X| = t]\} \cup \{[|X| \geq t] \cap [|X| < t]\}]; \end{aligned}$$

hence, observing that $[|X| \geq t] \cap [|X| = t] = [|X| = t]$ and $[|X| \geq t] \cap [|X| < t] = \emptyset$, it follows that

$$P|X| \geq t] = P[|X| = t].$$

Combining this equality with (2.4.5) it follows that

$$P[|X| = t] + P[|X| = 0] = P[|X| \geq t] + P[t > |X|] = 1,$$

which is property (ii).

(ii) $\Rightarrow$ (i): Suppose that $P[|X| = t] + P[X = 0] = 1$. In this case $P[|X| \geq t] = P[|X| = t]$ and

$$E[|X|^a] = 0^a P[|X| = 0] + t^a P[|X| = t] = t^a P[|X| = t] = t^a P[|X| \geq t],$$

and then $P[|X| = t] = \dfrac{E[|X|^a]}{t^a}$. □

Remark 2.4.1. Theorem 2.4.1 states that the equality occurs in Markov inequality just in the very special case that the distribution of $|X|$ is concentrated on two points, namely 0 and $t > 0$. For instance, suppose that $X \sim Bernoulli(p)$, so that $P[X = 1] = p = 1 - P[X = 0]$. In this case, the distribution of $|X|$ is concentrated on 0 and 1, and then , $P[|X| \geq 1] = E[|X|^a]/1^a = E[|X|^a]$ holds for every $a > 0$, but $P[|X| \geq t] < E[|X|^a]/t^a$ for every $t \neq 1$ and $a > 0$. Next, define $Y = X + 1$ so that the distribution of Y is concentrated on the points 1 and 2. In this case, it follows form Theorem 2.4.1 that $P[Y \geq t] < E[|Y|^a]/t^a$ for every $t, a > 0$, that is, the strict inequality occurs in (2.4.1); the reason is that Y is concentrated on two points which are nonull. □

2.5. Conditions for Equality: II

In this section Theorem 2.4.1 will be used to determine necessary and sufficient conditions so that equality holds in Chebishev inequality.

Theorem 2.5.1. Let X be a random variable with finite mean and variance μ_X and $\sigma_X^2 > 0$, respectively. For a positive real number k, the following conditions (i) and (ii) are equivalent.

(i) The equality $P[|X - \mu_X| \geq k\sigma_X] = \dfrac{1}{k^2}$ holds for some $k \geq 1$;

(ii) There exist $a, b \in \mathbb{R}$ such that $a < b$ and

$$P[X = a] = P[X = b] = p > 0 \quad \text{and} \quad P\left[X = \frac{a+b}{2}\right] = 1 - 2p; \tag{2.5.1}$$

(iii) For exactly one $k > 0$, the equality $P[|X - \mu_X| \geq k\sigma_X] = \dfrac{1}{k^2}$ holds.

Proof. Let X^* be the standardized version of X, that is,

$$X^* = \frac{X - \mu_X}{\sigma_X} \tag{2.5.2}$$

so that

$$E[|X^*|^2] = E\left[\left(\frac{X - \mu_X}{\sigma_X}\right)^2\right] = 1,$$

and

$$|X - \mu_X| \geq k\sigma_X \iff |X^*| \geq k.$$

(i) $\Rightarrow$ (ii): Suppose that $P[|X - \mu_X| \geq k\sigma_X] = 1/k^2$ and note that the two previous displays yield that

$$P[|X^*| \geq k] = 1/k^2 = E[|X^*|^2]/k^2.$$

From this point, an application of Theorem 2.4.1 yields that the distribution of $|X^*|$ is supported on the values 0 and k, that is,

$$1 = P[X^* = 0] + P[|X^*| = k].$$

Via (2.5.2) it follows that $X^* = 0$ if and only if $X = \mu_X$ whereas $|X^*| = k$ is equivalent to $|X - \mu_X| = k\sigma_X$, that is, $X = \mu_X + k\sigma_X$ or $X = \mu_X - k\sigma_X$. Combining these relations with (2.5.2) it follows that that

$$\begin{aligned} 1 &= P[X - \mu_X = 0] + P[|X - \mu_X| = k\sigma_X] \\ &= P[X = \mu_X] + P[X = \mu_X + k\sigma_X] + P[X = \mu_X - k\sigma_X] \end{aligned} \tag{2.5.3}$$

Setting

$$a = \mu_X - k\sigma_X \quad \text{and} \quad b = \mu_X + k\sigma_X, \quad \text{so that} \quad \mu_X = \frac{a+b}{2},$$

(2.5.3) leads to

$$1 = P[X = a] + P[X = b] + P[X = (a+b)/2]. \tag{2.5.4}$$

Observe now that

$$\begin{aligned} \frac{a+b}{2} &= \mu_X \\ &= E[X] \\ &= aP[X = a] + bP[X = b] + (1 - P[X = a] - P[x = b])\frac{a+b}{2} \\ &= \frac{a-b}{2}P[X = a] + \frac{b-a}{2}P[X = b] + \frac{a+b}{2}. \end{aligned}$$

It follows that $0 = \dfrac{a-b}{2}P[X = a] + \dfrac{b-a}{2}P[X = b]$, that is, $(b-a)P[X = a] = (b-a)P[X = b]$; since $b > a$ this last relation leads to

$$P[X = a] = P[X = b] =: p,$$

and then (2.5.4) immediately implies that $P[X = (a+b)/2] = 1 - 2p$; observing that $p > 0$, since otherwise $P[X = (a+b)/2] = 1$ and then $\sigma_X = 0$, assertion (ii) follows.

(ii) $\Rightarrow$ (iii): Suppose that (2.5.1) holds and note that

$$\begin{aligned} \mu_X = E[X] &= P[X = a]a + P[X = b]b + P[X = (a+b)/2](a+b)/2 \\ &= pa + pb + (1-2p)(a+b)/2 \\ &= (a+b)/2 \end{aligned}$$

and

$$\begin{aligned} \sigma_X^2 &= E[(X - \mu_X)^2] \\ &= P[X = a](a - \mu_X)^2 + P[X = b])(b - \mu_X)^2 + P[X = (a+b)/2](\mu_X - (a+b)/2)^2 \\ &= p(a - (a+b)/2)^2 + p(b - (a+b)/2)^2 \\ &= p(b-a)^2/2; \end{aligned}$$

hence

$$\sigma_X = \sqrt{\frac{p}{2}}(b-a).$$

Now set

$$k = \sqrt{\frac{1}{2p}}, \quad \text{so that} \quad k\sigma_X = \sqrt{\frac{1}{2p}}\sqrt{\frac{p}{2}}(b-a) = \frac{b-a}{2}.$$

Since the distribution of X is concentrated on $\{a, b, (a+b)/2 = \mu_X\}$, it follows that, with probability 1,

$$[|X - \mu_X| \geq k\sigma_X = (b-a)/2] = [X = a] \cup [X = b],$$

whereas

$$[|X - \mu_X| \geq \tilde{k}\sigma_X] = \begin{cases} [X = a] \cup [X = b], & \text{if } \tilde{k} \leq k, \\ \emptyset, & \text{if } \tilde{k} > k. \end{cases}$$

Hence,

$$P[|X - \mu_X| \geq k\sigma_X] = P[X = a] + P[X = b] = 2p = \frac{1}{k^2},$$

and

$$P[|X - \mu_X| \geq \tilde{k}\sigma_X] = \begin{cases} P[X = a] + P[X = b] = 2p = \dfrac{1}{k^2} < \dfrac{1}{\tilde{k}^2}, & \text{if } \tilde{k} < k, \\ 0 < \dfrac{1}{\tilde{k}^2}, & \text{if } \tilde{k} > k. \end{cases}$$

These two last displays show that the equality $P[|X - \mu_X| \geq t\sigma_X] = 1/t^2$ occurs just for the single value $t = k$, establishing (iii).

(iii) $\Rightarrow$ (i): This part is clear. □

Remark 2.5.1. Note that the conclusion of Theorem 2.5.1 can be summarized as follows: Equality in (2.2.2) occurs just when the following two conditions are satisfied:

(i) The distribution of X must be concentrated on three points, a, b and the midpoint $(a+b)/2$.

(ii) Points a and b are attained with the same probability, say p.

Under these conditions equality occurs in (2.2.2) only if $k = 1/\sqrt{2p}$. □

Example 2.5.1. (a) In the three cases considered in Example 2.3.1, the distribution of X is not concentrated on a set of three elements, so that the strict inequality must *always* occur in (2.2.2), a fact that was confirmed by direct calculations.

(b) Consider a random variable $X \sim \mathit{Bernoulli}(p)$. In this case, the distribution of X is supported on the three points $a = 0, b = 1$ and $1/2 = (a+b)/2$, so that requirement (i) in Remark 2.5.1 is satisfied; in fact, the distribution is supported on the two points 0 and 1, but 1/2 can be safely included in the support set. If $p \neq 1/2$, then $P[X = 1] = p \neq P[X = 0] = 1 - p$, and then requirement (ii) in the above Remark is not fulfilled, so that $P[|X - \mu_X| \geq k\sigma_X] < 1/k^2$ for every

$k > 0$. On the other hand, when $p = 1/2$, then $P[X = 1] = P[X = 0] = 1/2 = p$ and requirement (ii) in Remark 2.5.1 is satisfied, so that $P[|X - \mu_X| \geq k\sigma_X] = 1/k^2$ occurs when $k = 1/\sqrt{2p} = 1$, whereas $P[|X - \mu_X| \geq k\sigma_X] < 1/k^2$ for every positive $k \neq 1$. □

Chapter 3

Convergence of Random Variables

3.1. Introduction

In this chapter two ideas of convergence of random variables are discussed, and the relation between these two concepts is studied. The presentation starts introducing the notion of convergence in probability in Section 2, which is illustrated via a detailed example. Next, in Section 3 the notion of convergence in the mean is formulated, and it is shown that this idea is stronger than convergence in probability, that is, it is proved that, if $\{X_n\}$ converges in the mean to the random variable X, then $\{X_n\}$ also converges to X in probability. However, an example is used to show that the converse of this result is not true, so that the ideas of convergence in probability and convergence in the mean are not equivalent notions. The presentation continues in Section 4 with the (weak) law of large numbers, which can be described as follows: Given a sequence of $X_1, X_2, X_3, \ldots$ of independent and identically distributed random variables with finite mean and variance, then the average Y_n of the first n observations $X_1, X_2, \ldots, X_n$ converges in probability to the population mean. Finally, the exposition concludes in Section 5 with a discussion about an important problem in survey sampling, namely determining the sample size required to achieve certain precision/confidence combination. All of the results presented below are applications of Markov and Chebichev inequalities studied in Chapter 2.

3.2. Convergence in Probability

In this section the notion of convergence in probability for random variables is formulated. Intuitively, given a random variable X, a sequence $\{X_n\}$ of random variables converges to X if the difference between X_n and X becomes arbitrarily small with high probability whenever n is large enough.

Definition 3.2.1. Let X and $X_n, n = 1, 2, 3, \ldots$, be random variables defined on the same probability space. The sequence $\{X_n\}$ *converges in probability* to X if

$$\lim_{n\to\infty} P[|X_n - X| > \varepsilon] = 0 \quad \text{for every } \varepsilon > 0. \tag{3.2.1}$$

The notation

$$X_n \xrightarrow{\text{P}} X \tag{3.2.2}$$

will be used to indicate that $\{X_n\}$ converges in probability to X.

Note that in the above definition it is sufficient to require that (3.2.1) occurs for each ε small enough, say $\varepsilon \in (0, 1)$ or, more generally, $\varepsilon \in (0, \delta)$ where $\delta > 0$ is arbitrary. As already noted, (3.2.1) states that, with a probability as near to 1 as desired, the difference between X_n and X will be less than any positive amount ε prescribed beforehand if n is large enough.

Example 3.2.1. (i) Let X be a random variable, and for each n define X_n as follows:

$$X_n = \begin{cases} X, & \text{if } X \leq n \\ 0, & \text{if } X > n \end{cases}$$

Thus, $|X - X_n| = 0$ when $X < n$, so that, for every $\varepsilon > 0$,

$$[|X - X_n| > \varepsilon] \subset [X > n]$$

and then $P[|X - X_n| > \varepsilon] \leq P[X > n] = 1 - F_X(n) \to 0$ as $n \to \infty$, and it follows that $X_n \xrightarrow{\text{P}} X$.

(ii) Let $X_1, X_2, X_3, \ldots$ be independent and identically distributed random variables with distribution $\mathcal{U}(0, 1)$ (the uniform distribution (0,1)). Set

$$Y_n = \max\{X_1, X_2, \ldots, X_n\}$$

It will be shown that $\{Y_n\}$ converges in probability to the constant random variable equal to 1:

$$Y_n \xrightarrow{\text{P}} 1. \tag{3.2.3}$$

To achieve this goal observe that, with probability 1, the values of Y_n belong to $(0, 1)$ and then, for every $\varepsilon \in (0, 1)$,

$$[|Y_n - 1| > \varepsilon] = [Y_n \leq 1 - \varepsilon] = [X_i \leq 1 - \varepsilon,\ i = 1, 2, \ldots, n],$$

so that

$$P[|Y_n - 1| > \varepsilon] = P[X_i \leq 1 - \varepsilon,\ i = 1, 2, \ldots, n]$$
$$= P[X_1 \leq 1 - \varepsilon] P[X_2 \leq 1 - \varepsilon] \cdots P[X_n \leq 1 - \varepsilon]$$
$$= (1 - \varepsilon)(1 - \varepsilon) \cdots (1 - \varepsilon)$$
$$= (1 - \varepsilon)^n$$

and it follows that $P[|Y_n - 1| > \varepsilon] \to 0$ as $n \to \infty$, establishing (3.2.3).

(iii) Let $X_1, X_2, X_3, \ldots$ be a sequence of independent and identically distributed random variables with mean μ and variance σ^2. It will be proved that

$$Y_n := \frac{2}{n(n+1)} \sum_{i=1}^{n} i X_i \xrightarrow{\text{P}} \mu.$$

To achieve this goal observe that

$$E[Y_n] = \frac{2}{n(n+1)} \sum_{i=1}^{n} i E[X_i] = \frac{2}{n(n+1)} \sum_{i=1}^{n} i\mu$$
$$= \mu \frac{2}{n(n+1)} \sum_{i=1}^{n} i = \mu \frac{2}{n(n+1)} \frac{n(n+1)}{2} = \mu$$

whereas

$$E[[(Y_n - \mu)^2] = \text{Var}\,[Y_n] = \text{Var}\left[\frac{2}{n(n+1)} \sum_{i=1}^{n} i X_i\right]$$
$$= \left(\frac{2}{n(n+1)}\right)^2 \sum_{i=1}^{n} \text{Var}\,[i X_i] = \frac{4}{n^2(n+1)^2} \sum_{i=1}^{n} i^2 \text{Var}\,[X_i]$$
$$= \frac{4}{n^2(n+1)^2} \sum_{i=1}^{n} i^2 \sigma^2 = \sigma^2 \frac{4}{n^2(n+1)^2} \frac{n(n+1)(2n+1)}{6} = \sigma^2 \frac{4n+2}{3n(n+1)}$$

and it follows that

$$\lim_{n \to \infty} E[(Y_n - \mu)^2] = 0.$$

Therefore, an application of Markov inequality yields that, for every $\varepsilon > 0$,

$$P[|Y_n - \mu| > \varepsilon] \leq \frac{E[(Y_n - \mu)^2]}{\varepsilon^2} \to 0 \quad \text{as } n \to \infty,$$

and then $Y_n \xrightarrow{\text{P}} \mu$. □

3.3. Convergence in the Mean

In this section Markov inequality will be used to provide a sufficient criterion for convergence in probability. The proof of the following result extends the argument used to analyze Example 3.2.1(iii) in the previous section.

Lemma 3.3.1. Let X and $X_n, n = 1, 2, 3, \ldots$, be random variables defined on the same probability space, and suppose that, for some $a > 0$,

$$E[|X_n - X|^a] \to 0 \quad \text{as } n \to \infty. \tag{3.3.1}$$

In this case,

$$X_n \xrightarrow{\text{P}} X.$$

Proof. Given $\varepsilon > 0$ observe that Markov inequality yields that

$$P[|X_n - X| > \varepsilon] \le \frac{E[|X_n - X|^a]}{\varepsilon^a} \to 0 \quad \text{as } n \to \infty,$$

so that $X_n \xrightarrow{\text{P}} X$. □

When (3.3.1) holds, the sequence $\{X_n\}$ *converges to* X *in the mean of order* a. Thus, the main conclusion of the lemma states that

$$\begin{gathered}\text{'if } \{X_n\} \text{ converges to } X \text{ in the mean of order } a > 0, \\ \text{then } \{X_n\} \text{ converges to } X \text{ in probability'.}\end{gathered} \tag{3.3.2}$$

The most frequent application of the above lemmas occurs when X is a constant .

Example 3.3.1. Let $X_1, X_2, X_3, \ldots$ be a sequence of independent and identically distributed random variables with mean μ and variance σ^2. In this context, it will be verified that

$$Y_n := \frac{6}{n(n+1)(2n+1)} \sum_{i=1}^{n} i^2 X_i \xrightarrow{\text{P}} \mu.$$

To establish this conclusion observe that

$$\begin{aligned} E[Y_n] &= \frac{6}{n(n+1)(2n+1)} \sum_{i=1}^{n} i^2 E[X_i] = \frac{6}{n(n+1)(2n+1)} \sum_{i=1}^{n} i^2 \mu \\ &= \mu \frac{6}{n(n+1)(2n+1)} \sum_{i=1}^{n} i^2 = \mu \frac{6}{n(n+1)(2n+1)} \frac{n(n+1)(2n+1)}{6} = \mu. \end{aligned}$$

Next, it will be shown that $\{Y_n\}$ converges to μ in the mean of order 2, that is,

$$E[/Y_n - \mu)^2] \to 0 \quad \text{as } n \to \infty.$$

To achieve this goal, note that

$$\begin{aligned} E[[(Y_n - \mu)^2] = \text{Var}\,[Y_n] &= \text{Var}\left[\frac{6}{n(n+1)(2n+1)} \sum_{i=1}^{n} i^2 X_i\right] \\ &= \left(\frac{6}{n(n+1)(2n+1)}\right)^2 \sum_{i=1}^{n} \text{Var}\left[i^2 X_i\right] \\ &= \sigma^2 \left(\frac{6}{n(n+1)(2n+1)}\right)^2 \sum_{i=1}^{n} i^4 \end{aligned} \tag{3.3.3}$$

To continue, a bound for $\sum_{i=1}^{n} i^4$ will be obtained using the integral $\int_0^1 x^4\,dx$. Note that

$$\begin{aligned}
\frac{1}{5} &= \int_0^1 x^4\,dx \\
&= \sum_{k=1}^{n} \int_{(k-1)/n}^{k/n} x^4\,dx \\
&\geq \sum_{k=1}^{n} \int_{(k-1)/n}^{k/n} \left(\frac{k-1}{n}\right)^4 dx \\
&\geq \sum_{k=1}^{n} \frac{1}{n}\left(\frac{k-1}{n}\right)^4 = \frac{1}{n^5}\sum_{k=1}^{n}(k-1)^4 = \frac{1}{n^5}\sum_{i=1}^{n-1} i^4
\end{aligned}$$

Hence, $\displaystyle\sum_{i=1}^{n-1} i^4 \leq \frac{n^5}{5}$ and then

$$\sum_{i=1}^{n} i^4 = \sum_{i=1}^{n-1} i^4 + n^4 \leq \frac{n^5}{5} + n^4 \leq \frac{6n^5}{5}.$$

Combining this relation with the inequality $n(n+1)(2n+1) \geq n^3$, it follows that

$$\left(\frac{6}{n(n+1)(2n+1)}\right)^2 \sum_{i=1}^{n} i^4 \leq \left(\frac{6}{n^3}\right)^2 \frac{6}{5} n^5 = \frac{216}{5n}$$

and, via (3.3.3) it follows that $E[(Y_n-\mu)^2] \leq 216/(5n) \to 0$. Therefore, an application of Lemma 3.3.1 with μ and the sequence $\{Y_n\}$ instead of X and $\{X_n\}$, respectively, and $a = 2$ yields that $Y_n \xrightarrow{\text{P}} \mu$. □

After a glance at (3.3.2), the following question naturally arises:

Is it true that if $\{X_n\}$ converges in probability to X, then

$\{X_n\}$ converges to X in the mean of some order $a > 0$?

It will be shown in the example below that the answer to this question is negative.

Example 3.3.2. (i) Consider the following function

$$f(x) = \frac{\theta-1}{x\log(x)^\theta}, \quad x \geq e, \qquad f(x) = 0 \qquad x < e,$$

where $\theta > 1$.

(i) It will be verified that $f(x)$ is a density function. To this end, observe that $f(x) \geq 0$ for every $x \in \mathbb{R}$, and

$$\int_{\mathbb{R}} f(x)\,dx = \int_e^\infty f(x)\,dx = \int_e^\infty \frac{\theta-1}{x\log(x)^\theta}\,dx. \tag{3.3.4}$$

To compute the right-hand side, observe that , for $x > e$,

$$\begin{aligned}\frac{d}{dx}\left[-\log(x)^{1-\theta}\right] &= -(1-\theta)\log(x)^{-\theta}\frac{d}{dx}\log(x)\\ &= (\theta-1)\frac{1}{\log(x)^{\theta}}\frac{1}{x}\\ &= \frac{\theta-1}{x\log(x)^{\theta}} = f(x)\end{aligned}$$

Thus, $-\log(x)^{1-\theta} = -\log(x)^{1-\theta}$ is an antiderivative of $f(x)$ on the interval $[e,\infty)$, and then (3.3.4) yields that

$$\begin{aligned}\int_{\mathbb{R}} f(x)\,dx &= \int_e^\infty f(x)\,dx\\ &= -\log(x)^{1-\theta}\Big|_{x=e}^{\infty}\\ &= \lim_{b\to\infty}\left[-\log(b)^{1-\theta} - (-\log(e)^{1-\theta}\right]\\ &= \lim_{b\to\infty}\left[-\log(b)^{1-\theta}+1\right];\end{aligned}$$

on the other hand, since $\log(b)\to\infty$ as $b\to\infty$, and $1-\theta<0$, it follows that $\lim_{b\to\infty}\log(b)^{1-\theta} = 0$, and then the above display yields that $\int_{\mathbb{R}} f(x)\,dx = 1$. Hence, $f(\cdot)$ is a density function.

(ii) Let X be a random variable with density $f(x)$, and set

$$X_n = \begin{cases} X, & \text{if } X \le n\\ 0, & \text{if } X > n\end{cases}.$$

Show that

$$X_n \xrightarrow{\text{P}} X \quad \text{but} \quad E[|X_n - X|^a] \not\to 0 \text{ as } n\to\infty. \tag{3.3.5}$$

To achieve this goal, let $\varepsilon > 0$ and observe that $|X_n - X| > \varepsilon \iff X > n$, so that

$$P[|X_n - X| > \varepsilon] = P[X > n] = 1 - P[X \le n] \to 0 \quad \text{as } n\to\infty,$$

and then

$$X_n \xrightarrow{\text{P}} X. \tag{3.3.6}$$

On the other hand, for each $a > 0$

$$|X_n - X|^a = \begin{cases} X^a, & \text{if } X > n\\ 0, & \text{if } X \le n,\end{cases}$$

so that

$$\begin{aligned}E[|X_n - X|^a] &= \int_n^\infty x^a f(x)\,dx\\ &= \int_n^\infty x^a\frac{\theta-1}{x\log(x)^{\theta}}\,dx\\ &= \int_n^\infty \frac{1}{x^{1-a}\log(x)^{\theta}}\,dx\end{aligned} \tag{3.3.7}$$

To continue, recall that $\log(x)$ increases more slowly than any positive power of x, that is, $\dfrac{\log(x)}{x^b} \to 0$ as $x \to \infty$ for each $b > 0$, and then

$$\frac{\log(x)^\theta}{x^a} = \left(\frac{\log(x)}{x^{a/\theta}}\right)^\theta \to 0 \quad \text{as } x \to \infty.$$

Therefore, there exists N such that

$$\begin{aligned} n > N \Rightarrow \frac{\log(x)^\theta}{x^a} < 1 \Rightarrow\ & \log(x)^\theta < x^a \\ & \Rightarrow x^{1-a}\log(x)^\theta < x^{1-a}x^a \\ & \Rightarrow x^{1-a}\log(x)^\theta < x. \end{aligned}$$

Combining this relation with (3.3.7), it follows that $\dfrac{1}{x^{1-a}\log(x)^\theta} > \dfrac{1}{x}$ when $n > N$, so that

$$n > N \Rightarrow E[|X_n - X|^a] = \int_n^\infty \frac{1}{x^{1-a}\log(x)^\theta}\,dx \geq \int_n^\infty \frac{1}{x}\,dx = \infty,$$

and (3.3.5) follows combining this display with (3.3.6). □

3.4. Weak Law of Large Numbers

The most important application of the idea of convergence in probability occurs when X_n is the average of a sample from a given population, and X is a constant random variable equal to the population mean. The following result, which is known as the *weak law of large numbers*, states that if the sample size n is large enough, then for each $\varepsilon > 0$ the probability of observing that the sample mean $\overline{X}_n$ differs from the sample population μ by more than ε is negligible, that is, goes to zero as n increases.

Theorem 3.4.1. Let $X_1, X_2, X_3, \ldots$ be independent and identically distributed random variables, and suppose that their common distribution has finite mean μ and variance $\sigma^2 \in (0, \infty)$. Define the sample mean of the first n observations by

$$\overline{X}_n := \frac{X_1 + X_2 + \cdots + X_n}{n}.$$

In this case

$$P[\,|\overline{X}_n - \mu| > \varepsilon] \leq \frac{\sigma^2}{n\varepsilon^2}, \tag{3.4.1}$$

and then

$$\overline{X}_n \xrightarrow{\text{P}} \mu \tag{3.4.2}$$

Proof. Observe that

$$\mathrm{Var}\left[\overline{X}_n\right] = E[(\overline{X}_n - \mu)^2] = \frac{\sigma^2}{n}.$$

From this point, Chebishev inequality with $k = \varepsilon/\sqrt{\mathrm{Var}\left[\overline{X}_n\right]} = \varepsilon\sqrt{n}/\sigma$ yields that

$$P[|\overline{X}_n - \mu| \geq \varepsilon] = P[|\overline{X}_n - \mu| \geq k\sigma_{\overline{X}_n}] \leq \frac{1}{k^2} = \frac{\mathrm{Var}\left[\overline{X}_n\right]}{\varepsilon^2} = \frac{\sigma^2}{\varepsilon^2 n},$$

establishing (3.4.1); since $\dfrac{\sigma^2}{\varepsilon^2 n} \to 0$, this last display yields that $P[|\overline{X}_n - \mu| \geq \varepsilon] \to 0$, so that $\overline{X}_n \xrightarrow{\mathrm{P}} 0$. □

Example 3.4.1. (i) Suppose that $X_1, X_2, \ldots$ are independent and identically distributed random variables with $\mathcal{P}(\lambda)$ distribution. In this case the population mean and variance are given by is $\mu = \lambda$ and $\sigma^2 = \lambda$. Thus, $\overline{X}_n \xrightarrow{\mathrm{P}} \lambda$.

(ii) Suppose that $X_1, X_2, \ldots$ are independent and identically distributed random variables with $\mathcal{E}(\lambda)$ distribution (exponential distribution with density $\lambda e^{-\lambda x}$ for $x > 0$.) . In this case the population mean and variance are given by is $\mu = 1/\lambda$ and $\sigma^2 = 1/\lambda$, so that $\overline{X}_n \xrightarrow{\mathrm{P}} \lambda$.

(iii) Let $X_1, X_2, \ldots$ be independent and identically distributed random variables whose common distribution has finite fourth moment and, as usual, let μ and σ^2 stand for the population mean and variance, respectively. Now, consider the following version of the sample variance based on the first n observations:

$$\tilde{s}_n^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2 \tag{3.4.3}$$

It will be shown that

$$\tilde{s}_n^2 \xrightarrow{\mathrm{P}} \sigma^2. \tag{3.4.4}$$

To achieve this goal, note that

$$\tilde{s}_n^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \mu)^2 - (\overline{X}_n - \mu)^2. \tag{3.4.5}$$

and observe that following facts (a) and (b):

(a) The random variables $(X_1 - \mu)^2, (X_2 - \mu)^2, (X_3 - \mu)^2, \ldots$ are independent with the same distribution. Their common expectation is $E[(X_i - \mu)^2] = \sigma^2$, whereas their common variance is finite, since $\mathrm{Var}\left[(X_i - \mu)^2\right] \leq E[\{(X_i - \mu)^2\}^2] < \infty$, where the last inequality stems from the condition $E[X_i^4] < \infty$. Therefore, the law of large numbers in Theorem 3.4.1 yields that

$$\frac{1}{n}\sum_{i=1}^{n}(X_i - \mu)^2 \xrightarrow{\mathrm{P}} \sigma^2.$$

(b) It will be shown that

$$(\overline{X}_n - \mu)^2 \xrightarrow{\text{P}} 0.$$

To this end, let $\varepsilon > 0$ and observe that

$$\begin{aligned} P[|(\overline{X}_n - \mu)^2 - 0| > \varepsilon] &= P[(\overline{X}_n - \mu)^2 > \varepsilon] \\ &= P[|\overline{X}_n - \mu| > \sqrt{\varepsilon}] \to 0 \quad \text{as } n \to \infty \end{aligned},$$

where the convergence is due to the fact that $\overline{X}_n \to \mu$, by Theorem 3.4.1. Next, use (3.4.5) to obtain

$$\begin{aligned} |\tilde{s}_n^2 - \sigma^2| > \varepsilon &\iff \left| \left(\frac{1}{n} \sum_{i=1}^n (X_i - \mu)^2 - (\overline{X}_n - \mu)^2 \right) - \sigma^2 \right| > \varepsilon \\ &\iff \left| \left(\frac{1}{n} \sum_{i=1}^n (X_i - \mu)^2 - \sigma^2 \right) - (\overline{X}_n - \mu)^2 \right| > \varepsilon \\ &\implies \left| \frac{1}{n} \sum_{i=1}^n (X_i - \mu)^2 - \sigma^2 \right| > \frac{\varepsilon}{2} \quad \text{or} \quad (\overline{X}_n - \mu)^2 > \frac{\varepsilon}{2} \end{aligned}$$

so that

$$[|\tilde{s}_n^2 - \sigma^2| > \varepsilon] \subset \left[\left| \frac{1}{n} \sum_{i=1}^n (X_i - \mu)^2 - \sigma^2 \right| > \frac{\varepsilon}{2} \right] \cup \left[(\overline{X}_n - \mu)^2 > \frac{\varepsilon}{2} \right],$$

and then

$$P[|\tilde{s}_n^2 - \sigma^2| > \varepsilon] \leq P \left[\left| \frac{1}{n} \sum_{i=1}^n (X_i - \mu)^2 - \sigma^2 \right| > \frac{\varepsilon}{2} \right] + P \left[(\overline{X}_n - \mu)^2 > \frac{\varepsilon}{2} \right].$$

Letting n go to ∞, properties (a) and (b) above yield that the two terms in the right-hand side of this display converge to 0 as $n \to \infty$, so that $P[|\tilde{s}_n^2 - \sigma^2| > \varepsilon] \to 0$, establishing (3.4.4). □

If $W_n \xrightarrow{\text{P}} a$, where a is a constant, the it is said that W_n estimates a consistently. Thus, the law of large numbers in Theorem 3.4.1 states that $\overline{X}_n$ *estimates* μ *consistently*, or that $\{\overline{X}_n\}$ *is a consistent sequence of estimators of* μ, whereas part (iii) in Example 3.4.1 states that $\tilde{s}_n^2$ estimates the population variance σ^2 consistently.

3.5. Sample Size in a Survey

The ideas studied so far will be now used to study a basic the problem in the design of a sampling survey, namely determining the size of the sample to achieve a desired precision/confidence goal. Suppose that it is required to approximate μ with an error of at most $\varepsilon > 0$ (this number ε represents the precision). Of course, when the observed data involve randomness, generally no interesting assertion can be made with complete certainty, but usually there exists the possibility of making an incorrect statement based on the observations. Thus, the question is how to guarantee that the difference between μ and the approximation $\overline{X}_n$ does not exceed ε with 'high probability' say γ, which is close to 1 (this number γ is the confidence level). Thus, the problem is

- To determine a sample size n such that $P[|\overline{X}_n - \mu| \leq \varepsilon] \geq \gamma$.

This problem can be studied using (3.4.1). First, note that $P[|\overline{X}_n - \mu| \leq \varepsilon] \geq \gamma$ is equivalent to

$$P[|\overline{X}_n - \mu| > \varepsilon] < 1 - \gamma$$

Next observe that, by (3.4.1), the above inequality holds if $\dfrac{\sigma^2}{\varepsilon^2 n} < 1 - \gamma$ which yields

$$\frac{\sigma^2}{\varepsilon^2(1-\gamma)} < n \tag{3.5.1}$$

and any sample size n satisfying this inequality suffices to ensure that, with probability at least γ, the difference $|\overline{X}_n - \mu|$ does not exceed ε. Note that to determine n satisfying (3.5.1) it is necessary to know the population variance σ^2, or at least an upper bound for such a figure.

(ii) Let $X_1, X_2, \ldots, X_n, \ldots$ be random variables with *Bernoulli*(p) distribution (the Bernoulli distribution with parameter $p \in [0, 1]$), so that $\mu = p$ and $\sigma^2 = p(1-p)$. Suppose that p is unknown and will be approximated with $\overline{X}_n$, the sample mean based on n observations. Since $\sigma^2 = p(1-p)$ depends on p, σ^2 is unknown and an upper bound must be determined before using (3.5.1). To find an upper bound for the variance σ^2, set

$$f(p) = p(1-p)$$

and recall that $p \in [0, 1]$. Now, the maximum value attained by f in $[0, 1]$ will be determined. To this end, first note that f is differentiable everywhere, and that $f'(p) = 1 - 2p = 0$ has the unique solution $p = 1/2$. Thus, f attains its maximum at $p = 1/2$ or at the extreme points 0 and 1 of its domain. Since $f(0) = 0 = f(1)$ and $f(1/2) = 1/4$ it follows that the maximum of f is $1/4$ and then

$$\frac{1}{4} \geq p(1-p) = \sigma^2.$$

Consequently, the inequality in (3.5.1) will be satisfied if

$$\frac{1}{4\varepsilon^2(1-\gamma)} = \frac{1/4}{\varepsilon^2(1-\gamma)} < n$$

Assume that it is desired to approximate p with an error at most 0.03 with confidence level 0.95, so that $\gamma = 0.95$ and $\varepsilon = 0.03$. In this case, the above expression yields that

$$\frac{1}{4(0.03)^2(1-0.95)} \leq n$$

and then $5/(0.3)^2 = 50000/9 = 5555.\overline{5} < n$. In short, a sample size of 5556 suffices to ensure that, with probability 0.95 or more, $\overline{X}_n$ and p differ by at most 0.03. It must be observed that (3.5.1) was obtained from Chebishev inequality, which usually is not sharp. Thus, it might be expected that smaller values of n will be sufficient to ensure the desired maximum error with the given confidence level. In practice, samples with size about 1200 are taken and it is ensured that, with confidence 0.95 the difference between p and $\overline{X}_n$ does not exceed 0.03. The method used to obtain this 'reduced' value of n involves the central limit theorem and will be discussed in the next chapter.

Chapter 4

Asymptotic Normality

4.1. Introduction

Throughout the remainder, $X_1, X_2, X_3, \ldots$ is a sequence of independent and identically distributed random vectors whose common distribution has finite moment of order two, at least. This condition is naturally satisfied in sampling theory where the underlying population is a finite set. However, in that context, the exact distribution of the estimators of such quantities as the population total or average is impossible to determine, since the whole set values of the study variable are not known. The material presented below is extremely helpful in that context, since under minimal conditions, the normal distribution can be used to approximate the exact (but unknown) distribution of the statistic under consideration, result that is discussed in Section 2. The approximation result is also relevant to determine the a sample size, which is generally substantially smaller that the one obtained using Chebishev inequality, but allows to achieve a desired precision with a given confidence level, a topic that is presented in Section 3. Next, in Section 4 it is shown that asymptotic normality is preserved under the application of smooth (differentiable) functions, result that is illustrated in Section 5 determining the limit distribution of the coefficient of variation and a risk ratio, and the exposition concludes in Section 6 with some examples concerning *variance stabilizing transformations*, that is, functions that when applied to the relevant statistic have the effect that the variance of the asymptotic distribution is constant.

4.2. Central Limit Theorem

The law of large numbers reveals a fundamental property of the sample mean $\overline{X}_n$, namely, it comes closer to the population mean μ as n increases, so that $\overline{X}_n - \mu$ is 'small' if n is large enough. Recall now that, for a sample of size n from a population with finite variance σ^2, the variance of $\overline{X}_n$ is σ^2/n, and the standardized sample mean for n observations $X_1, X_2, \ldots, X_n$ is given by

$$\overline{X}_n^* = \frac{\overline{X}_n - \mu}{\sqrt{\sigma^2/n}} = \sqrt{n}\frac{\overline{X}_n - \mu}{\sigma}; \tag{4.2.1}$$

note that $\overline{X}_n^*$ is obtained multiplying the 'small' quantity $\overline{X}_n - \mu$ by $\sqrt{n}/\sigma$, which is a 'large' figure. The following classical *Central Limit Theorem* shows that $\sqrt{n}/\sigma$ acts as 'a magnifying glass' allowing to observe the difference in $\overline{X}_n - \mu$ in detail. Essentially, such a result establishes that, for 'any' set $A \subset \mathbb{R}$, the probability that the standardized mean $\overline{X}_n^*$ belongs to a set A can be approximated by the probability that $Z \in A$, where Z is a random variable with the standard normal distribution.

Theorem 4.2.1. Let $X_1, X_2, X_3, \ldots$ be independent and identically distributed random variables with finite mean μ and variance $\sigma^2 \in (0, \infty)$. In this case, , for every interval $A \subset \mathbb{R}$

$$P[\overline{X}_n^* \in A] \to P[Z \in A], \quad \text{where} \quad Z \sim \mathcal{N}(0,1);$$

more explicitly,

$$\lim_{n\to\infty} P[\overline{X}_n^* \in A] = 4.1_A \frac{1}{\sqrt{2\pi}} e^{-z^2/2}\, dz. \tag{4.2.2}$$

A proof of this result can be found in Dudewicz and Mishra (1988), or Ash (2000).

Remark 4.2.1. Two alternative notations are used to indicate that (4.2.2) holds:

$$\overline{X}_n^* \xrightarrow{\text{d}} Z \quad \text{where} \quad Z \sim \mathcal{N}(0,1),$$

(which is red as '$\overline{X}_n^*$ converges in distribution to Z'), or

$$\overline{X}_n^* \xrightarrow{\text{d}} \mathcal{N}(0,1), \tag{4.2.3}$$

which is red as '$\overline{X}_n^*$ converges to the $\mathcal{N}(0,1)$ distribution'.

(ii) Instead of analyzing the distribution of $\overline{X}_n^*$, frequently it is more convenient to study $\sqrt{n}(\overline{X}_n - \mu) = \sigma\overline{X}_n^*$ Suppose that (4.2.2) holds, and let $A = (a, b)$ a given interval, In this case,

$$\begin{aligned}
P[\sqrt{n}(\overline{X}_n - \mu) \in A] &= P[\sqrt{n}(\overline{X}_n - \mu) \in (a,b)] \\
&= P\left[\sqrt{n}\frac{\overline{X}_n - \mu}{\sigma} \in (a/\sigma, b/\sigma)\right] \\
&= P[\overline{X}_n^* \in (a/\sigma, b/\sigma)] \\
&\to 4.1_{a/\sigma}^{b/\sigma} \frac{1}{\sqrt{2\pi}} e^{-z^2/2}\, dz
\end{aligned}$$

and using the change of variable $z = y/\sigma$ in the last integral it follows that

$$P[\sqrt{n}(\overline{X}_n - \mu) \in A] \to 4.1_a^b \frac{1}{\sqrt{2\pi\sigma^2}} e^{-y^2/[2\sigma^2]}\, dy$$

Therefore,

$$P[\sqrt{n}(\overline{X}_n - \mu) \in A] \to P[W \in (a, b)] = P[W \in A], \quad \text{where } W \sim \mathcal{N}(0, \sigma^2)$$

and the following notation is used for this convergence:

$$\sqrt{n}(\overline{X}_n - \mu) \xrightarrow{\text{d}} W \text{ where } W \sim \mathcal{N}(0, \sigma^2),$$

or

$$\sqrt{n}(\overline{X}_n - \mu) \xrightarrow{\text{d}} \mathcal{N}(0, \sigma^2). \tag{4.2.4}$$

Note that this argument shows that

$$\overline{X}_n^* \xrightarrow{\text{d}} \mathcal{N}(0, 1) \iff \sqrt{n}(\overline{X}_n - \mu) \to \mathcal{N}(0, \sigma^2).$$

□

Example 4.2.1. (i) Suppose that $X_1, X_2, X_3, \ldots$ is a sequence of independent random variables with common distribution $\mathcal{P}(\lambda)$ (the Poisson distribution with parameter λ). In this case $\mu = \lambda = \sigma^2$, and then $\sqrt{n}(\overline{X}_n - \lambda) \xrightarrow{\text{d}} \mathcal{N}(0, \lambda)$.

(ii) If $X_1, X_2, X_3, \ldots$ is a sequence of independent and identically distributed with common distribution *Bernoulli* (p). In this case $\mu = p$ and $p(1 - p) = \sigma^2$. Hence, $\sqrt{n}(\overline{X}_n - p) \xrightarrow{\text{d}} \mathcal{N}(0, p(1 - p))$.
□

The following extension of Theorem 4.2.1 to the multivariate case is a most important result.

Theorem 4.2.2. [Multivariate Central Limit Theorem.] Let $\mathbf{X} = (X_1, X_2, \ldots, X_k)'$ be a random vector with mean μ and variance matrix M, that is,

$$\mu = (\mu^{(1)}, \mu^{(2)}, \ldots, \mu^{(k)})' = (E[X_1], E[X_2], \ldots, E[X_k])'$$

$$M = [m_{i\,j}] = \text{Cov}\,(X_i, X_j)\,.$$

Suppose that $\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3, \ldots,$ is a sequence of independent and identically distributed random vectors with the same distribution as $\mathbf{X}$. In this case, given a region $A \subset \mathbb{R}^k$,

$$\sqrt{n}[\bar{(}\mathbf{X}_n - \mu) \in A] \to 4.1_A \frac{1}{(2\pi)^{n/2}|M|} e^{-\mathbf{x}M^{-1}\mathbf{x}/2}\, d\mathbf{x} \tag{4.2.5}$$

where $\overline{\mathbf{X}}_n = (\mathbf{X}_1 + \mathbf{X}_2 + \cdots + \mathbf{X}_n)/n$ is the sample mean of the first n observation vectors.

The notation

$$\sqrt{n}\ [\overline{\mathbf{X}}_n - \mu] \xrightarrow{\mathrm{d}} \mathcal{N}_k(0, M) \tag{4.2.6}$$

will be used to indicate that (4.2.5) holds.

4.3. Sample Size

Theorem 4.2.1 is a extremely important result in statistics. It is quite general, in the sense that convergence (4.2.2) occurs for any distribution with finite mean and variance. The central limit theorem can be used to analyze the problem of determining the necessary sample size to achieve a specified precision with a given confidence level. As it is discussed below, the bounds for the sample size will be sharper than the ones obtained via Chebishev inequality.

• Suppose that it is required to approximate μ with an error of at most $\varepsilon > 0$ and a confidence level at least $\gamma \in (0, 1)$, so that

$$P[|\overline{X}_n - \mu| \leq \varepsilon] \geq \gamma. \tag{4.3.1}$$

The problem is to determine a sample size n such that this relation is satsified.

To begin with observe that

$$\begin{aligned} P[|\overline{X}_n - \mu| \leq \varepsilon] &= P\left[\left|\frac{\sqrt{n}(\overline{X}_n - \mu)}{\sigma}\right| \leq \frac{\sqrt{n}\varepsilon}{\sigma}\right] \\ &\approx P\left[|Z| \leq \frac{\sqrt{n}\varepsilon}{\sigma}\right] \end{aligned} \tag{4.3.2}$$

where the approximation is based on Theorem 4.2.1 and Z has the standard normal distribution. Next, let z_α be the (right-)percentil of Z of order α, so that $P[Z > z_\alpha] = \alpha$, and then

$$P[|Z| \leq z_\alpha] = 1 - 2\alpha.$$

Selecting α in such a way that $1 - 2\alpha = \gamma$, that is, $\alpha = (1 - \gamma)/2$ it follows that

$$P[|Z| \leq z_{(1-\gamma)/2}] = \gamma.$$

Combining this equality with (4.3.2) it follows that

$$P[|\overline{X}_n - \mu| \geq \varepsilon] \approx \gamma \quad \text{if} \quad \frac{\sqrt{n}\varepsilon}{\sigma} \approx z_{(1-\gamma)/2},$$

and direct calculations solving the above approximation for n lead to

$$n \approx \frac{\sigma^2}{\varepsilon^2} z^2_{(1-\gamma)/2}. \tag{4.3.3}$$

and any sample size n satisfying this relation suffices to ensure that $|\overline{X}_n - \mu|$ does not exceed ε with an approximate probability γ.

Comparing the above expression with relation (3.5.1) obtained from Chebishev inequality, note that the right hand side of (4.3.3) is obtained from (3.5.1) replacing $\dfrac{1}{1-\gamma}$ by $z^2_{(1-\gamma)/2}$. For instance, consider the case $\gamma = 0.95$, and then $1/(1-\gamma) = 20$, whereas $z_{(1-\gamma)/2} = z_{0.025} = 1.959964$, and then $z^2_{(1-\gamma)/2} = 3.841459$, so that the size n obtained from Chebishev inequality in (3.5.1) is $20/3.841459 \approx 5.2$ times the value of n obtained form the approximation of the central limit theorem. As a specific illustration, consider the problem of estimating the parameter of the *Bernoulli* (p) distribution. Suppose that it is desired to estimate p with an error at most $\varepsilon = 0.03$ with probability $\gamma = 0.95$, In this case (4.3.3) yields that

$$n \approx \frac{\sigma^2}{\varepsilon^2} z^2_{(1-0.95)/2} = 1067.072,$$

where the unknown value of σ^2 was replaced by its upper bound $1/4$. This relation states that about 1110 observations are sufficient to ensure that the difference between $\overline{X}_n$ and p will be less than 0.03 with approximate probability 0.95. Note that, in contrast with (3.5.1), the expression (4.3.3) is just an approximation for the value of the sample size n. Thus, it is natural to ask how good is the approximation in (4.3.3). In the present case, the following recommendation has been obtained from empirical (computational) studies: The approximation (4.3.3) is *satisfactory* if $np \geq 30$ and $n(1-p) \geq 30$, that is, if

$$n \min\{p, 1-p\} \geq 30.$$

For instance, for $p = 0.3$, this condition sates that $n(0.03) \geq 30$, or $n \geq 1000$, and then a sample size of $n = 1067$ gives a probability near to 0.95 of observing a difference of at most 0.3 between p and $\overline{X}_n$ when the true (but unknown) value of $\min\{p, 1-p\}$ is 0.3 or larger. □

4.4. Smooth Transformation Theorem

In the section a property of a sequence of estimators $\{\hat{g}_n\}$ of a parametric function $g(\theta)$ is introduced. The idea is to combine the consistency of the estimators $\hat{g}_n$ with the statement that, as n increases and after normalizing the difference between the estimator and the unknown parametric quantity, the resulting sequence is approximately normally distributed with mean $g(\theta)$. The formal definition of this idea is presented below.

Definition 4.4.1. Consider a parametric function $g\colon \Theta \to \mathbb{R}^d$ defined on the parameter space Θ and taking values in $\mathbb{R}^d$, and for each positive integer n let $\hat{g}_n$ be an estimator of $g(\theta)$ based on the first n observations $X_1, X_2, \ldots, X_n$. In this case, the sequence $\{\hat{g}_n\}$ of estimators is *consistent and asymptotically normal* with mean μ and variance $V(\theta)$ if, and only, if

$$\sqrt{n}\ [\hat{g}_n - g(\theta)] \xrightarrow{\text{d}} \mathcal{N}\,(0, V(\theta));$$

where $V(\theta)$ is square nonnegative matrix of order $d \times d$. In this case, $V(\theta)$ is referred to as the asymptotic variance of $\sqrt{n}\ [\hat{g}_n - g(\theta)]$.

The main objective of this section is to establish invariance property of the convergence to normality, which can be roughly stated as follows: If a sequence of random vectors $\{W_n\}$ converges to a (multivariate) normal distribution, and if g is a smooth (differentiable) function, then the transformed sequence $\{g(W_n)\}$ also converges to a normal distribution. This fundamental result is formally stated below in the following theorem.

The following result establishes that convergence to normality is not altered under the application of differentiable transformations.

Theorem 4.4.1. Suppose that $\{W_n\}$ is a sequence of k-dimensional random vectors such that

$$\sqrt{n}\ [W_n - \mu] \xrightarrow{\text{d}} \mathcal{N}_k(0, M)$$

for some nonnegative matrix M of order $k \times k$ and $\mu \in \mathbb{R}^k$. In this case, let g be a function defined on an open set of $\mathbb{R}^k$ containing the vector μ, suppose that g takes value in $\mathbb{R}^d$ and that g is differentiable at μ. In this case

$$\sqrt{n}\ [g(W_n) - g(\mu)] \xrightarrow{\text{d}} \mathcal{N}_d(0, Dg(\mu)MDg(\mu)'),$$

where $Dg(\mu)$ is the (matrix) derivative of g at μ, which has order $d \times k$.

Note that if $\{W_n\}$ is asymtotically normal with mean μ and variance M, this result establishes that $\{g(W_n)\}$ is asymptotically normal with mean $g(\mu)$ and variance $Dg(\mu)MDg(\mu)'$. The following example illustrates the transformation theorem .

Example 4.4.1. Suppose that $X_1, X_2, X_3 \ldots,$ is a sequence of independent and identically distributed random variables with mean μ and variance $\sigma^2 < \infty$. The central limit theorem yields that

$$\sqrt{n}\ [\overline{X}_n - \mu] \xrightarrow{\text{d}} \mathcal{N}\left(0, \sigma^2\right). \tag{4.4.1}$$

Now, the asymptotic distribution of some transformations $\{g(\overline{X}_n)\}$ will be obtained by an application of Theorem 4.4.1.

(i) $g(x) = e^x$. In this case, $g(\overline{X}_n) = e^{\overline{X}_n}$, and observing that $Dg(x) = g'(x) = e^x$, it follows that $Dg(\mu) = e^\mu$. Hence, starting from (4.4.1), an application of Theorem 4.4.1 leads to

$$\sqrt{n}\,\left[e^{\overline{X}_n} - e^\mu\right] \xrightarrow{\mathrm{d}} \mathcal{N}\left(0, e^\mu\sigma^2 e^\mu\right) = \mathcal{N}\left(0, e^{2\mu}\sigma^2\right)$$

(ii) $g(x) = \sin(x)$. For this function, $g(\overline{X}_n) = \sin(\overline{X}_n)$, and $Dg(x) = g'(x) = \cos(x)$, so that $Dg(\mu) = \cos(\mu)$. Thus, (4.4.1), and Theorem 4.4.1 together imply that

$$\sqrt{n}\,[\sin(\overline{X}_n) - \sin(\mu)] \xrightarrow{\mathrm{d}} \mathcal{N}\left(0, \cos(\mu)\,\sigma^2\cos(\mu)\right) = \mathcal{N}\left(0, \cos(\mu)^2\,\sigma^2\right)$$

(iii) Consider now that transformation $g(x) = (e^x, \sin(x))'$. This function transforms $\mathbb{R} = \mathbb{R}^1$ into $\mathbb{R}^2$, and its derivative Dg is the following matrix of order 2×1:

$$Dg(x) = \begin{bmatrix} \dfrac{d}{dx}e^x \\ \dfrac{d}{dx}\sin(x) \end{bmatrix} = \begin{bmatrix} e^x \\ \cos(x) \end{bmatrix}.$$

Therefore,

$$\sqrt{n}\,[g(\overline{X}_n) - g(\mu)) \xrightarrow{\mathrm{d}} \mathcal{N}_2(0, Dg(\mu)\sigma^2 Dg(\mu)');$$

more explicitly,

$$\sqrt{n}\,\left[\begin{pmatrix} e^{\overline{X}_n} \\ \sin(\overline{X}_n) \end{pmatrix} - \begin{pmatrix} e^\mu \\ \sin(\mu) \end{pmatrix}\right] \xrightarrow{\mathrm{d}} \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} e^\mu \\ \cos(\mu) \end{bmatrix}\sigma^2[e^\mu \quad \cos(\mu)]\right)$$
$$= \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \sigma^2\begin{bmatrix} e^{2\mu} & e^\mu\cos(\mu) \\ e^\mu\cos(\mu) & \cos^2(\mu) \end{bmatrix}\right). \qquad \square$$

Example 4.4.2. Let $X_1, X_2, X_3, \ldots$ be independent random variables from the *Bernoulli* (p) population, where the parameter $p \in (0,1)$ is unknown. The first population moment is p, so that the moments estimator of p is $\hat{p}_n = \overline{X}_n$. Since the population variance is $\sigma^2 = p(1-p)$, the central limit theorem yields that

$$\sqrt{n}\,[\overline{X}_n - p] \xrightarrow{\mathrm{d}} \mathcal{N}(0, p(1-p))$$

Consider now the smooth function

$$g(p) = \arcsin(\sqrt{p}),$$

so that

$$D_p g(p) = g'(p) = \frac{d}{dp}\arcsin(\sqrt{p}) = \frac{1}{\sqrt{1-(\sqrt{p})^2}}\frac{1}{2\sqrt{p}} = \frac{1}{2}\frac{1}{\sqrt{p(1-p)}}.$$

An application of Theorem yields that

$$\sqrt{n}\,[\arcsin(\overline{X}_n) - \arcsin(p)] = \sqrt{n}\,[g(\overline{X}_n) - g(p)]$$
$$\xrightarrow{\mathrm{d}} \mathcal{N}(0, Dg(p)p(1-p)Dg(p)) = \mathcal{N}\left(0, \frac{1}{4}\right)$$

notice that the (asymptotic) variance of the transformed mean—$\arcsin(\overline{X}_n)$—does not depend on the value of p; this stabilizing transformation is frequently used when comparing proportions, since an essential assumption in the analysis of variance is that the standard deviations of the different populations being compared are the same. □

.

4.5. Coefficient of Variation and Risk Ratio

In this section Theorem 4.4.1 will be used to find the limit distribution of the coefficient of variation when that data are obtained form a normal population. Throughout the following discussion $X_1, X_2, X_3, \ldots$ are independent and identically distributed random variables with $\mathcal{N}\left(\mu, \sigma^2\right)$ distribution. Now, recall that the sample variance

$$S_n^2 = \sum_{i=1}^{n}(X_i - \overline{X}_n)^2/(n-1)$$

satisfies that

$$\frac{(n-1)S_n^2}{\sigma^2} \sim \chi^2_{n-1}. \tag{4.5.1}$$

Next let $Z_1, Z_2, \ldots, Z_{n-1}$ be independent random variables with standard normal distribution. In the case, since the variables Z_i^2 are independent with mean 1 and variance 2, the central limit theorem yields that

$$\sqrt{n-1}\frac{(Z_1^2 + Z_2^2 + \cdots + Z_{n-1}^2)/(n-1) - 1}{\sqrt{2}} \xrightarrow{\text{d}} \mathcal{N}(0,1);$$

also, it is known that $Z_1^2 + Z_2^2 + \cdots + Z_{n-1}^2 \sim \chi^2_{n-1}$, and using (4.5.1) it is possible to use $(n-1)S_n^2/\sigma^2$ instead of $\sum Z_i^2$ in the above display to obtain

$$\sqrt{n-1}\ \frac{S_n^2/\sigma^2 - 1}{\sqrt{2}} \xrightarrow{\text{d}} \mathcal{N}\left(0, 1\right),$$

which is equivalent to

$$\sqrt{n-1}\ [S_n^2 - \sigma^2] \xrightarrow{\text{d}} \mathcal{N}\left(0, 2\sigma^4\right).$$

Because of the convergence $\sqrt{n}/\sqrt{n-1} \to 1$ as $n \to \infty$, it follows that

$$\sqrt{n}\ [S_n^2 - \sigma^2] \xrightarrow{\text{d}} \mathcal{N}\left(0, 2\sigma^4\right).$$

Consider now the function

$$g(x) = \sqrt{x}, \qquad \text{so that } Dg(x) = g'(x) = 1/[2\sqrt{x}].$$

Via Theorem 4.4.1, the two previous displays yield that

$$\begin{aligned}\sqrt{n}\,[S_n-\sigma] &= \sqrt{n}\,[g(S_n^2)-g(\sigma^2)]\\ &\xrightarrow{\mathrm{d}} \mathcal{N}\left(0, g'(\sigma^2)\,(2\sigma^4)g'(\sigma^2)\right) = \mathcal{N}\left(0,\sigma^2/2\right).\end{aligned} \tag{4.5.2}$$

• The *coefficient of variation*

$$\mathrm{CV} = \frac{\mu}{\sigma}$$

is naturally estimated by

$$\widehat{\mathrm{CV}}_n = \frac{\overline{X}_n}{S_n},$$

which is the maximum likelihood estimator as well as the moments estimator, and the present objective is to determine its asymptotic distribution. To achieve this goal, the following the well-known fact will be used: *for the normal model $\overline{X}_n$ and S_n are independent random variables.* Combining this fact with (4.5.2) and the convergence $\sqrt{n}\,[\overline{X}_n-\mu]\xrightarrow{\mathrm{d}}\mathcal{N}\left(0,\sigma^2\right)$, it follows that

$$\sqrt{n}\,\left[\begin{pmatrix}\overline{X}_n\\ S_n\end{pmatrix}-\begin{pmatrix}\mu\\ \sigma\end{pmatrix}\right]\xrightarrow{\mathrm{d}}\mathcal{N}\left(\begin{bmatrix}0\\0\end{bmatrix},\begin{bmatrix}\sigma^2 & 0\\ 0 & \sigma^2/2\end{bmatrix}\right) \tag{4.5.3}$$

Next, consider the function transforming a vector in $\mathbb{R}^2$ with no-null second component into the a real number specified as follows:

$$g\begin{pmatrix}x_1\\ x_2\end{pmatrix} = \frac{x_1}{x_2}.$$

The derivative of g is the matrix of order 1×2 given by

$$Dg\begin{pmatrix}x_1\\ x_2\end{pmatrix} = [\partial_{x_1}g, \partial_{x_2}g] = [1/x_2,\ \ -x_1/x_2^2],$$

and it follows that

$$g\begin{pmatrix}\mu\\ \sigma\end{pmatrix} = \frac{\mu}{\sigma} = \mathrm{CV},\quad g\begin{pmatrix}\overline{X}_n\\ S_n\end{pmatrix} = \frac{\overline{X}_n}{S_n} = \widehat{\mathrm{CV}}_n,\quad Dg\begin{pmatrix}\mu\\ \sigma\end{pmatrix} = [1/\sigma,\ \ -\mu/\sigma^2],$$

and then

$$Dg\begin{pmatrix}\mu\\ \sigma\end{pmatrix}\begin{bmatrix}\sigma^2 & 0\\ 0 & \sigma^2/2\end{bmatrix}Dg\begin{pmatrix}\mu\\ \sigma\end{pmatrix}' = 1+\frac{\mu^2}{2\sigma^2} = 1+\frac{\mathrm{CV}^2}{2}$$

Thus, starting with (4.5.3), an application of Theorem 4.4.1 with the function g specified above yields that

$$\sqrt{n}\,\left[\widehat{\mathrm{CV}}_n-\mathrm{CV}\right] = \sqrt{n}\,\left[g\begin{pmatrix}\overline{X}_n\\ S_n\end{pmatrix}-g\begin{pmatrix}\mu\\ \sigma\end{pmatrix}\right]\xrightarrow{\mathrm{d}}\mathcal{N}\left(0,1+\frac{\mathrm{CV}^2}{2}\right).$$

Example 4.5.1. Consider samples $X_1, X_2, \ldots, X_n$ and $Y_1, Y_2, \ldots, Y_n$ of the *Binomial*(p_1) and *Binomial*(p_2) populations, respectively. In health studies, each p_i is interpreted as the probability if acquiring some illness and is ‘small’, whereas the ratio

$$r = \frac{p_1}{p_2}$$

is referred as the *risk ratio*. The moments estimator of r based on the two samples of size n is

$$\hat{r}_n = \frac{\overline{X}_n}{\overline{Y}_n},$$

and obtaining an approximation for the distribution of $\hat{r}_n$ for large samples is an interesting and important problem. Notice that $\sqrt{n}\,[\overline{X}_n - p_1] \xrightarrow{\text{d}} \mathcal{N}(0, p_1(1-p_1))$ and $\sqrt{n}\,[\overline{Y}_n - p_1] \xrightarrow{\text{d}} \mathcal{N}(0, p_2(1-p_2))$, by the central limit theorem, and that the independence of the samples implies that

$$\sqrt{n}\,\left[\begin{bmatrix}\overline{X}_n\\ \overline{Y}_n\end{bmatrix} - \begin{bmatrix}p_1\\ p_2\end{bmatrix}\right] \xrightarrow{\text{d}} \mathcal{N}\left(\begin{bmatrix}0\\0\end{bmatrix}, \begin{bmatrix}p_1(1-p_1) & 0\\ 0 & p_2(1-p_2)\end{bmatrix}\right). \tag{4.5.4}$$

Now, consider the function

$$g(p_1, p_2) = \log(p_2/p_1) = \log(p_2) - \log(p_1),$$

and notice that

$$Dg(p_1,p_2) = (\partial_{p_1} g(p_1,p_2),\ \partial_{p_2} g(p_1,p_2)) = \left(\frac{1}{p_1}, \frac{1}{p_2}\right),$$

as well as

$$Dg(p_1,p_2)\begin{bmatrix}p_1(1-p_1) & 0\\ 0 & p_2(1-p_2)\end{bmatrix} Dg(p_1,p_2)' = (1-p_1)/p_1 + (1-p_2)/p_2.$$

and, starting with (4.5.4), an application of Theorem 4.4.1 yields that

$$\sqrt{n}\,[\log(\overline{X}_n/\overline{Y}_n) - \log(p_1/p_2)] = \sqrt{n}\,\left[\begin{bmatrix}\overline{X}_n\\ \overline{Y}_n\end{bmatrix} - \begin{bmatrix}p_1\\ p_2\end{bmatrix}\right] \xrightarrow{\text{d}} \mathcal{N}\left(0, \frac{1-p_1}{p_1} + \frac{1-p_2}{p_2}\right),$$

that is,

$$\sqrt{n}\,[\log(\hat{r}_n) - \log(r)] \xrightarrow{\text{d}} \mathcal{N}\left(0, \frac{1-p_1}{p_1} + \frac{1-p_2}{p_2}\right).$$

□

4.6. Additional Examples

In this section more illustrations of the transformation theorem are analyzed.

Example 4.6.1. (i) Suppose that

$$\sqrt{n}(X_n - \mu) \xrightarrow{\text{d}} \mathcal{N}(0, \sigma^2),$$

and consider the function $g(x) = x^2$. In this case $g(\cdot)$ is differentiable everywhere, so that Theorem 4.4.1 implies that $\{g(X_n)\}$ is asymptotically normal. Since $g'(\mu) = 2\mu$, it follows that

$$\begin{aligned}\sqrt{n}(g(X_n) - g(\mu)) &= \sqrt{n}(X_n^2 - \mu^2)\\ &\xrightarrow{\text{d}} \mathcal{N}(0, g'(\mu)^2\sigma^2) = \mathcal{N}(0, [2\mu]^2\sigma^2) = \mathcal{N}(0, 4\mu^2\sigma^2).\end{aligned}$$

(ii) Suppose that $Y_i \sim \mathcal{P}(\lambda)$, $i = 1, 2, \ldots$, are independent. In this case, $E[Y_i] = \lambda = \mathrm{Var}\,[Y_i]$, and the central limit theorem implies that

$$\sqrt{n}(\overline{Y}_n - \lambda) \xrightarrow{\mathrm{d}} \mathcal{N}(0, \lambda) \tag{4.6.1}$$

Now, consider the function $g(x) = 2\sqrt{x}$, and observe that $g'(x) = 1/\sqrt{x}$, so that $[g'(\lambda)]^2 = [1/\sqrt{\lambda})]^2 = 1/\lambda$, and then

$$\begin{aligned}\sqrt{n}(g(\overline{Y}_n) - g(\lambda)) &= 2\sqrt{n}\left(\sqrt{\overline{Y}_n} - \sqrt{\lambda}\right)\\ &\xrightarrow{\mathrm{d}} \mathcal{N}(0, [g'(\lambda)]^2\lambda) = \mathcal{N}(0, 1).\end{aligned} \tag{4.6.2}$$

An interesting aspect in this example is that the asymptotic variance in (4.6.1) depends on λ (frequently an unknown parameter), whereas it is constant in the above convergence. The mapping $g(x) = 2\sqrt{x}$ is referred to as *a variance stabilizing function* for the Poisson distribution.

(iii) Suppose that $Y_i \sim \mathcal{E}(\lambda)$, $i = 1, 2, \ldots$, are independent; recall the $\mathcal{E}(\lambda)$ stands for the exponential distribution with density $f(x) = \lambda e^{-\lambda x}$ for $x > 0$. In this case,

$$E[Y_i] = 1/\lambda$$

and

$$\mathrm{Var}\,[Y_i] = 1/\lambda^2,$$

and the central limit theorem implies that

$$\sqrt{n}\left(\overline{Y}_n - \frac{1}{\lambda}\right) \xrightarrow{\mathrm{d}} \mathcal{N}\left(0, \frac{1}{\lambda^2}\right) \tag{4.6.3}$$

Now, the function g be given by $g(x) = \ln(x)$, so that $g'(x) = 1/x$, and then $g'(\mu) = g'(1/\lambda) = \lambda$. Therefore, $[g'(\mu)]^2\sigma^2 = [\lambda]^2(1/\lambda^2) = 1$, and then

$$\begin{aligned}\sqrt{n}(g(\overline{Y}_n) - g(1/\lambda)) &= \sqrt{n}\left(\ln(\overline{Y}_n) - \ln(1/\lambda)\right)\\ &\xrightarrow{\mathrm{d}} \mathcal{N}(0, [g'(1/\lambda)]^2/\lambda) = \mathcal{N}(0, 1);\end{aligned} \tag{4.6.4}$$

since the asymptotic variance of $\ln(\overline{Y}_n)$ is constant, the mapping $g(x) = \ln(x)$ is *a variance stabilizing function* for the exponential distribution.

(iv) Let $Y_i \sim \mathit{Bernoulli}\,(p)$, $i = 1, 2, \ldots$, be independent random variables, and observe that $E[Y_i] = p$ and $\mathrm{Var}\,[Y_i] = p(1-p)$. By the central limit theorem, it follows that

$$\sqrt{n}(\overline{Y}_n - p) \xrightarrow{\mathrm{d}} \mathcal{N}(0, p(1-p)) \tag{4.6.5}$$

Now, consider the function $g(x) = 2\arcsin(\sqrt{x})$, $x \in (0, 1)$, and observe that

$$g'(x) = 2\frac{1}{\sqrt{1 - (\sqrt{x})^2}}\left(\frac{1}{2\sqrt{x}}\right) = \frac{1}{\sqrt{x(1-x)}}$$

Therefore,

$$[g'(\mu)]^2\sigma^2 = [g'(p)]^2 p(1-p) = \left[\frac{1}{\sqrt{p(1-p)}}\right]^2 p(1-p) = 1$$

and then Theorem implies that

$$\begin{aligned}\sqrt{n}(g(\overline{Y}_n) - g(\lambda)) &= 2\sqrt{n}\left(\arcsin \overline{Y}_n - \arcsin \lambda\right) \\ &\xrightarrow{\text{d}} \mathcal{N}(0, [g'(p)]^2 p(1-p)) = \mathcal{N}(0,1).\end{aligned} \tag{4.6.6}$$

Thus, $2\arcsin(\cdot)$ is a variance stabilizing function for the Bernoulli family of distributions. □

Chapter 5

Simple Random Sampling

5.1. Introduction

The basic problem studied in Sampling Theory consists in formulating inferences about a whole population $\mathcal{U}$ using knowledge of just one part (a subset) of $\mathcal{U}$. In principle, the population is finite, the subset of the population which is analyzed to state the inferences is called the *sample* and, generally, it is required to accompany the stated conclusions about the population with an assessment of their precision or reliability. Such a requirement can be fulfilled if the analyzed sample is chosen via a random procedure, and this chapter introduces the basic ideas of 'probability sampling schemes'. The subsequent material has been organized as follows: In Section 2 the notions of population, sample and parameter are introduced, and the basic problem in the theory of sampling is formally stated. Next, in Section 3 two general strategies (or schemes) to select a sample are briefly described, and they are illustrated by means of two important schemes, namely, the *simple and Bernoulli* strategies. Then, the concept of sampling (probability) design is formulated in Section 4, and an alternative implementation of the simple design is presented in Section 5. Finally, the chapter concludes in Section 6, which concerns with two notions that will pay important roles in the study of estimation problems, namely, the ideas of inclusion probabilities and membership indicators.

5.2. Population and Random Samples

Consider a set $\mathcal{U}$ with N elements, hereafter referred to as *the population*, whose elements are denoted by U_i, $i = 1, 2, 3, \ldots, N$:

$$\mathcal{U} = \{U_1, U_2, U_3, \ldots, U_N\}. \tag{5.2.1}$$

This set $\mathcal{U}$ is an abstract representation of a class of concrete objects. The elements U_i are refereed to as *the units* and each one of them conveys some information that is of interest to the analyst. Such information is represented by a function

$$\mathcal{Y}: \mathcal{U} \to \mathbb{R}^k,$$

which is referred to as the *study variable*, and

$$Y_i = \mathcal{Y}(U_i), \quad i = 1, 2, 3, \ldots, N \tag{5.2.2}$$

stands for the value that the function $\mathcal{Y}$ associates with U_i. For instance, if the population $\mathcal{U}$ consists of all the oranges in a container, Y_i may represent the amount of juice that can be extracted for the i-th orange U_i, whereas if the units U_i are persons, Y_i might be the weight of the i-th person, or the pair (weight, age) for the i-th person. It is assumed that N, the number of elements of the population, is known, but the function $\mathcal{Y}$ is *unknown*, so that the value Y_i associated with U_i can be determined only after analyzing the unit U_i. Throughout the remainder the interest focuses in two parameters (this is the technical name for a quantity that depends on all the values $Y_1, Y_2, \ldots, Y_N$): the population *total*

$$Y = Y_1 + Y_2 + Y_3 + \cdots + Y_N \tag{5.2.3}$$

and the population *average*

$$\overline{Y} = \frac{Y_1 + Y_2 + Y_3 + \cdots + Y_N}{N}. \qquad (5.2.4).$$

The main problem considered below can be sated as follows:

To estimate the population total or average based only
on $Y_i = \mathcal{Y}(U_i)$ for U_i in a subset S of the population $\mathcal{U}$ (5.2.5)

This problem is important in the common situation that it is impossible, impractical or expensive to examine all of the units in the population to determine the whole set of values $Y_1, Y_2, \ldots, Y_N$, and then compute exactly the value of the parameter. However, it is possible that the available resources (time, budget) allow to examine some units $U_{i_1}, U_{i_2}, \ldots, U_{i_n}$ so that the corresponding $\mathcal{Y}$-values $Y_{i_1}, Y_{i_2}, \ldots, Y_{i_n}$ can be determined, and the problem is to obtain 'a reasonable approximation' of the parameter value using only the information obtained from the analyzed units. The subset of $\mathcal{U}$ for which the values $Y_i = \mathcal{Y}(U_i)$ are determined is called *a sample*, so that the above problem can be stated as follows:

To estimate a population parameter based on the
values Y_i corresponding to units U_i in a sample S. (5.2.6)

Of course, since the parameters $\overline{Y}$ or Y are unknown, when 'approximations' $\hat{Y}$ or $\hat{\overline{Y}}$ are proposed, a measure of the 'estimation error' $|\hat{Y} - Y|$ must be provided. Such an assessment is possible only if the sample used in the analysis was selected via a random mechanism (Lhor, 2010). In this work, all of the samples considered below will be obtained from $\mathcal{U}$ using *simple random sampling*, as described below.

5.3. Simple Random Sampling

In this section a basic random sampling method is briefly described. Let $n < N$ be the desired *sample size* The *simple random sampling scheme* (without replacement), which is used to obtain a sample of size n is as follows:

1. Select a member of the population using a random mechanism assigning probability $1/N$ to each one of the N elements of $\mathcal{U}$;

2. Remove from the population the unit selected in the previous draw and, with equal probability $1/(N-1)$, select from the remaining $N-1$ elements a new member of the population;

$\vdots$

n. Remove from the population the units selected in the $n-1$ draws already performed and, with equal probability $1/(N-n+1)$, select a new element from the remaining $N-n+1$ units.

After these steps, a (random) sequence

$$\tilde{S} = (U_{i_1}, U_{i_2}, \ldots, U_{i_n}) \tag{5.3.1}$$

is obtained, where U_{i_k} is the unit selected in the k-th draw. This is a vector of distinct units taking values on the space

$$\tilde{\mathcal{S}}_n := \{\tilde{s} = (u_1, u_2, \ldots, u_n) \,|\, u_1, u_2, \ldots, u_n \text{ are different elements of } \mathcal{U}\}. \tag{5.3.2}$$

The elements of $\tilde{\mathcal{S}}_n$ are the *ordered samples* without replacement of size n and are also referred to as the *permutations* of size n of the population $\mathcal{U}$. From the above description it follows that

$$P[\tilde{S} = \tilde{s}] = \frac{1}{(N)_n} = \frac{1}{N(N-1)\cdots(N-n+1)}, \quad \tilde{s} \in \tilde{S}_n,$$

that is, all of the ordered samples (permutations) of size n have the same probability of selection. Finally, a set S is immediately determined form $\tilde{S}$ forgetting the order in which the units were selected:

$$S = \{U_{i_1}, U_{i_2}, \ldots, U_{i_n}\}. \tag{5.3.3}$$

This set is a member of the family

$$\mathcal{S}_n = \{s |\, s \text{ is a subset of size } n \text{ of } \mathcal{U}\}$$

which consists of all subsets (samples) of size n of $\mathcal{U}$. Since the elements of a set of size n can be arranged into a sequence in $n!$ forms, it follows that

$$P[S = s] = \frac{n!}{(N)_n} = \frac{1}{\binom{N}{n}}, \quad s \in \mathcal{S}, \tag{5.3.4}$$

so that all of the (unordered) samples of size n have the same probability of selection.

Definition 5.3.1. Given a sample S as in (5.3.3), let

$$y_j = \mathcal{Y}(U_{i_j}), \quad j = 1, 2, \ldots, n \tag{5.3.5}$$

by the information associated to the j-th unit in the sample.

(i) The sample mean (or average) is denoted by $\overline{y}$ and is defined by

$$\overline{y} = \frac{y_1 + y_2 + \cdots + y_n}{n}.$$

(ii) The estimators $\hat{\overline{Y}}$ and $\hat{Y}$ of the population average and total, respectively, are given by

$$\hat{\overline{Y}} = \overline{y}, \qquad \hat{Y} = N\overline{y}. \tag{5.3.6}$$

For every positive integer $r \leq N$, the indicator random variable $I_r \equiv I_r(S)$ of the unit U_r is defined by

$$I_r(S) = \begin{cases} 1, & \text{if } U_r \in S, \\ 0, & \text{if } U_r \notin S. \end{cases} \tag{5.3.7}$$

With this notation, a glance at (5.3.3) and (5.3.5) yields that

$$\sum_{i=1}^{n} y_i = \sum_{r=1}^{N} I_r(S) Y_r$$

and then

$$\hat{\overline{Y}} = \overline{y} = \frac{1}{n} \sum_{r=1}^{N} I_r(S) Y_r, \qquad \hat{Y} = N\overline{y} = \frac{N}{n} \sum_{r=1}^{N} I_r(S) Y_r. \tag{5.3.8}$$

5.4. Mean and Variance of the Estimators

Under the simple random sampling previously introduced, in this section the expectation and variance of $\overline{y}$ will be computed. First, it will be assumed that the study variable is scalar, that is, takes values in $\mathbb{R}$.

Theorem 5.4.1. Under the simple random sampling scheme with sample size n,

$$E[\overline{y}] = \overline{Y}, \quad \text{and} \quad \text{Var}\,[\overline{y}] = \left(\frac{1}{n} - \frac{1}{N}\right) S_Y^2,$$

where

$$S_Y^2 = \frac{1}{N-1}\sum_{k=1}^{N}(Y_k - \overline{Y})^2 \tag{5.4.1}$$

is the population variance of $Y_1, Y_2, \ldots, Y_N$.

The proof of this result relies on the following lemma.

Lemma 5.4.1. Under the simple random sampling scheme with sample size n,

$$\begin{aligned} E[I_r] &= \frac{n}{N} \\ \text{Var}\,[I_r] &= \frac{n}{N}\left(1 - \frac{n}{N}\right) \\ \text{Cov}\,(I_r, I_t) &= -\frac{n}{N}\left(1 - \frac{n}{N}\right)\frac{1}{N-1}, \quad r \neq t \end{aligned}$$

.

Proof. Let $\mathcal{S}_n$ be the family of all subsets of $\mathcal{U}$ with n elements, so that $P[S = s] > 0$ if, and only if, $s \in \mathcal{S}_n$; recall that

$$P[S = s] = \frac{1}{\binom{N}{n}}, \quad s \in \mathcal{S}_n.$$

Next, observe that $I_r = I_r(S)$ is a Bernoulli variable, that is, it attains just the values 0 and 1; moreover, $I_r(S) = 1$ if and only if $U_r \in S$. Since there are $\binom{N-1}{n-1}$ subsets of S with n elements that include U_r, it follows that

$$P[I_r = 1] = \sum_{s\in\mathcal{S}_n:\, U_r\in s} P(S = s] = \sum_{s\in\mathcal{S}_n:\, U_r\in s} \frac{1}{\binom{N}{n}} = \frac{\binom{N-1}{n-1}}{\binom{N}{n}};$$

using the identity

$$\binom{N}{n} = \frac{N}{n}\binom{N-1}{n-1} \tag{5.4.2}$$

it follows that

$$P[I_r = 1] = \frac{n}{N}, \quad \text{and then} \quad I_r \sim \textit{Bernoulli}\,(n/N)\,.$$

Applying the well-known formulas for the mean and variance of the *Bernoulli* (p) distribution, it follows that

$$E[I_r] = \frac{n}{N} \quad \text{and} \quad \text{Var}\,[I_r] = \frac{n}{N}\left(1 - \frac{n}{N}\right).$$

Now, let r and t be two positive different integers less than or equal to N and observe that $I_r(S)I_t(S)$ attains just two values, 0 and 1, and that $I_r(S)I_t(S) = 1$ if, and only if, $U_r \in S$ and $U_t \in S$, so that

$$\begin{aligned} P[I_r(S)I_t(S) = 1] &= P[U_r \in S, U_t \in S] \\ &= \sum_{s\in\mathcal{S}_n:U_r\in s,\, U_t\in s} P[S = s] \\ &= \sum_{s\in\mathcal{S}_n:U_r\in s,\, U_t\in s} \frac{1}{\binom{N}{n}}, \end{aligned}$$

To conclude, recall that there are $\binom{N-2}{n-2}$ samples of size n containing U_j and U_k, so that that

$$P[I_r(S)I_t(S)=1]=\frac{\begin{pmatrix}N-2\\ n-2\end{pmatrix}}{\begin{pmatrix}N\\ n\end{pmatrix}}=\frac{n(n-1)}{N(N-1)},\quad r\neq t,$$

where the last equality is due to the relation $\begin{pmatrix}N\\ n\end{pmatrix}=\frac{N(N-1)}{n(n-1)}\begin{pmatrix}N-2\\ n-2\end{pmatrix}$, which follows applying (5.4.2) twice. Thus, $E[I_rI_t]=\frac{n(n-1)}{N(N-1)}$ and then

$$\begin{aligned}\text{Cov}\,(I_r,I_t)&=E[I_rI_t]-E[I_r]E[I_t]\\ &=\frac{n(n-1)}{N(N-1)}-\frac{n}{N}\frac{N}{N}\\ &=\frac{n}{N}\left(\frac{n-1}{N-1}-\frac{n}{N}\right)=-\frac{1}{N-1}\frac{n}{N}\left(1-\frac{n}{N}\right),\end{aligned}$$

and the proof is complete. □

Proof of Theorem 5.4.1. The argument combines the previous lemma with formula (5.3.8). Note that

$$E[\overline{y}]=\frac{1}{n}E\left[\sum_{k=1}^{N}I_kY_k\right]=\frac{1}{n}\sum_{k=1}^{N}Y_kE\left[I_k\right]=\frac{1}{n}\sum_{k=1}^{N}Y_k\frac{n}{N}=\frac{1}{N}\sum_{k=1}^{N}Y_k=\overline{Y}.$$

On the other hand, to compute the variance of $\overline{y}$ it is convenient to use the following notation,

$$\tau=\frac{n}{N}\left(1-\frac{n}{N}\right).$$

Then, the conclusions of Lemma 5.4.1 yield that t

$$\text{Var}\,[I_k]=\tau\quad\text{and}\quad\text{Cov}\,(I_k,I_t)=-\frac{1}{N-1}\tau.$$

Therefore,

$$\begin{aligned}\text{Var}\,[\overline{y}]&=\text{Var}\left[\frac{1}{n}\sum_{k=1}^{N}Y_kI_k\right]\\ &=\frac{1}{n^2}\text{Var}\left[\sum_{k=1}^{N}I_kY_k\right]\\ &=\frac{1}{n^2}\left[\sum_{k=1}^{N}Y_k^2\text{Var}\,[I_k]+\sum_{1\leq j\neq k\leq N}Y_jY_k\text{Cov}\,(I_i,I_k)\right]\\ &=\frac{1}{n^2}\left[\sum_{k=1}^{N}Y_k^2\tau-\sum_{1\leq j\neq k\leq N}Y_jY_k\frac{\tau}{N-1}\right]\\ &=\frac{\tau}{n^2}\left[\sum_{k=1}^{N}Y_k^2-\frac{1}{N-1}\sum_{1\leq j\neq k\leq N}Y_jY_k\right].\end{aligned}$$

Now observe that

$$N^2\overline{Y}^2 = \left(\sum_{k=1}^{N} Y_k\right)^2 = \sum_{k=1}^{N} Y_k \sum_{j=1}^{N} Y_j = \sum_{1\leq j,k\leq N} Y_j Y_k = \sum_{k=1}^{N} Y_k^2 + \sum_{1\leq j\neq k\leq N} Y_j Y_k$$

so that

$$\sum_{1\leq j\neq k\leq N} Y_j Y_k = N^2\overline{Y}^2 - \sum_{k=1}^{N} Y_k^2$$

and then

$$\begin{aligned}\mathrm{Var}\left[\overline{y}\right] &= \frac{\tau}{n^2}\left[\sum_{k=1}^{N} Y_k^2 - \frac{1}{N-1}\left(N^2\overline{Y}^2 - \sum_{k=1}^{N} Y_k^2\right)\right].\\ &= \frac{\tau}{n^2}\left[\sum_{k=1}^{N} Y_k^2 + \frac{1}{N-1}\sum_{k=1}^{N} Y_k^2 - \frac{N^2}{N-1}\overline{Y}^2\right].\\ &= \frac{\tau}{n^2}\frac{N}{N-1}\left[\sum_{k=1}^{N} Y_k^2 - N\overline{Y}^2\right].\end{aligned}$$

Using the identity $\sum_{k=1}^{N}(Y_k - \overline{Y})^2 = \sum_{k=1}^{N} Y_k^2 - N\overline{Y}^2$, it follows that

$$\begin{aligned}\mathrm{Var}\left[\overline{y}\right] &= \frac{\tau}{n^2}\frac{N}{N-1}\sum_{k=1}^{N}(Y_k - \overline{Y})^2\\ &= \frac{N}{n^2}\tau\frac{1}{N-1}\sum_{k=1}^{N}(Y_k - \overline{Y})^2 \quad = \frac{N}{n^2}\tau S_Y^2.\end{aligned}$$

Finally, observe that

$$\frac{N}{n^2}\tau = \frac{N}{n^2}\frac{n}{N}\left(1 - \frac{n}{N}\right) = \frac{1}{n}\left(1 - \frac{n}{N}\right) = \frac{1}{n} - \frac{1}{N},$$

and the desired expression for $\mathrm{Var}\left[\overline{y}\right]$ follows combining the two last displays. □

Example 5.4.1. Suppose that the study variable is a bi-dimensional vector, say

$$Z_i = \mathcal{Z}(U_i) = \begin{bmatrix} Y_i \\ X_i \end{bmatrix}.$$

In this case, given the sample $S = \{U_{i_1}, U_{i_2}, \ldots, U_{i_n}\}$ the information obtained from the unit j-th unit in the sample is

$$z_j = \mathcal{Z}(U_{i_j}) = \begin{bmatrix} y_j \\ x_j \end{bmatrix},$$

and

$$\overline{z} = \begin{bmatrix} \overline{y} \\ \overline{x} \end{bmatrix}.$$

The objective of this example is to determine

$$\mathbf{V} = \mathrm{Var}\left[\overline{z}\right] = \begin{bmatrix} \mathrm{Var}\left[\overline{y}\right] & \mathrm{Cov}\left(\overline{y}, \overline{x}\right) \\ \mathrm{Cov}\left(\overline{x}, \overline{y}\right) & \mathrm{Var}\left[\overline{x}\right] \end{bmatrix}$$

Of course, from Theorem 5.4.1 it is known that

$$\text{Var}\,[\overline{y}] = \left(\frac{1}{n} - \frac{1}{N}\right) S_Y^2 \quad \text{and} \quad \text{Var}\,[\overline{x}] = \left(\frac{1}{n} - \frac{1}{N}\right) S_X^2,$$

where S_X^2 is as in (5.4.1) with the study variable X_i instead of Y_i. To determine Cov $(\overline{y}, \overline{x})$, define the new study variable,

$$W_i = Y_i + X_i, \quad i = 1, 2, \ldots, N,$$

so that $\overline{W} = \overline{Y} + \overline{X}$ and

$$\begin{aligned}
S_W^2 &= \frac{1}{N-1}\sum_{k=1}^{N}(W_i - \overline{W})^2 \\
&= \frac{1}{N-1}\sum_{k=1}^{N}(Y_i + X_i - [\overline{Y} + t\overline{X}])^2 \\
&= \frac{1}{N-1}\sum_{k=1}^{N}([Y_i - \overline{Y}] + [X_i - \overline{X}])^2 \\
&= \frac{1}{N-1}\left[\sum_{k=1}^{N}(Y_i - \overline{Y})^2 + \sum_{k=1}^{N}(X_i - \overline{X})^2 + 2\sum_{k=1}^{N}(Y_i - \overline{Y})(X_i - \overline{X})\right] \\
&= S_Y^2 + S_X^2 + 2S_{YX}
\end{aligned} \tag{5.4.3}$$

where

$$S_{YX} = \frac{1}{N-1}\sum_{k=1}^{N}(Y_i - \overline{Y})(X_i - \overline{X})$$

The W values obtained form the sample are $w_j = y_j + x_j$, $j = 1, 2, \ldots, n$, and then

$$\overline{w} = \overline{y} + \overline{x}$$

Now, Var $[\overline{w}]$ will be computed in two ways:

- Using the formula for the variance of a sum,

$$\text{Var}\,[\overline{w}] = \text{Var}\,[\overline{y} + \overline{x}] = \text{Var}\,[\overline{y}] + \text{Var}\,[\overline{x}] + 2\text{Cov}\,(\overline{y}, \overline{x})$$

- Applying Theorem 5.4.1 with the study variable W instead of Y

$$\begin{aligned}
\text{Var}\,[\overline{w}] &= \left(\frac{1}{n} - \frac{1}{N}\right) S_W^2 \\
&= \left(\frac{1}{n} - \frac{1}{N}\right) \left[S_Y^2 + S_X^2 + 2S_{YX}\right] \\
&= \left(\frac{1}{n} - \frac{1}{N}\right) S_Y^2 + \left(\frac{1}{n} - \frac{1}{N}\right) S_X^2 + 2\left(\frac{1}{n} - \frac{1}{N}\right) S_{YX} \\
&= \text{Var}\,[\overline{y}] + \text{Var}\,[\overline{x}] + 2\left(\frac{1}{n} - \frac{1}{N}\right) S_{YX}
\end{aligned}$$

where the second equality is due to (5.4.3). Comparing the two last displays, it follows that

$$\mathrm{Cov}\,(\overline{y}, \overline{x}) = \left(\frac{1}{n} - \frac{1}{N}\right) S_{YX}. \tag{5.4.4}$$

To conclude this section, the limit distribution of $\overline{y}$ will be discussed: Suppose that

$$N \to \infty, \quad n \to \infty, \quad \text{and} \quad \frac{n}{N} \to 0. \tag{5.4.5}$$

This condition means that the sample and population sizes under consideration are 'big', but the sample size is 'small' when compared with the population size.

Theorem 5.4.2. Under condition (5.4.5)

$$\sqrt{n}(\overline{y} - \overline{Y}) \xrightarrow{\mathrm{d}} \mathcal{N}(0, S_Y^2). \tag{5.4.6}$$

and

$$\sqrt{n}\left(\begin{bmatrix}\overline{x}\\ \overline{y}\end{bmatrix} - \begin{bmatrix}\overline{X}\\ \overline{Y}\end{bmatrix}\right) \xrightarrow{\mathrm{d}} \mathcal{N}\left(\begin{bmatrix}0\\ 0\end{bmatrix}, \begin{bmatrix}S_X^2 & S_{YX}\\ S_{XY} & S_Y^2\end{bmatrix}\right). \tag{5.4.7}$$

This result was used in Section 3.5 to determine the sample size to achieve a required precision with a desired confidence level.

5.5. Ratio Estimator

Consider the problem of estimating the population total for the study variable $\mathcal{Y}$, which is given in (5.2.3), and suppose that there is other interesting feature of the population units which is given by the quantity X_i for the unit U_i. It is supposed that the total of this additional study variable, given by

$$X = X_1 + X_2 + X_3 + \cdots + X_N,$$

is *known*. Under this condition, define

$$\hat{Y}_R = \frac{\overline{y}}{\overline{x}} X \tag{5.5.1}$$

which is known as the *ratio estimator* for the total Y. The main objective of the section is to compare the mean quadratic error of this estimator with the variance of the usual estimator $\hat{Y}$ defined in (5.3.6). The main result is the following

Theorem 5.5.1. Under (5.4.5),

$$\frac{E[(\hat{Y} - Y)^2]}{E[(\hat{Y}_R - Y)^2]} \approx \frac{S_Y^2}{S_X^2 \left(\frac{\overline{Y}}{\overline{X}}\right)^2 - 2S_{XY}\frac{\overline{Y}}{\overline{X}} + S_Y^2}. \tag{5.5.2}$$

Remark 5.5.1. The above theorem provides a guide to decide which estimator $\hat{Y}$ or $\hat{Y}_R$ to use when estimating the population total Y. Of course, these comments make sense when condition (5.4.5) holds, that is, if the sample and population sizes n and N are large, and n is small compared with N. In this context, (5.5.2) yields that

$$\begin{aligned}
\hat{Y}_R \text{ is preferred to } \hat{Y} &\iff E[(\hat{Y}_R - Y)^2] < E[(\hat{Y} - Y)^2] \\
&\iff S_Y^2 > S_X^2\left(\frac{\overline{Y}}{\overline{X}}\right)^2 - 2S_{XY}\frac{\overline{Y}}{\overline{X}} + S_Y^2 \\
&\iff 2S_{XY}\frac{\overline{Y}}{\overline{X}} > S_X^2\left(\frac{\overline{Y}}{\overline{X}}\right)^2 \\
&\iff \frac{S_{XY}}{S_X S_Y} > \frac{1}{2}\frac{S_X}{S_Y}\frac{\overline{Y}}{\overline{X}} \\
&\iff \rho_{XY} > \frac{1}{2}\frac{CV(X)}{CV(Y)}
\end{aligned}$$

where ρ_{XY} is the population correlation coefficient between x and Y, whereas $CV(X) = S_X/\overline{X}$ and $CV(Y) = S_Y/\overline{Y}$ are the coefficients of variation of X and Y, respectively. Thus, use the ratio estimator if X and Y are 'highly correlated'; if $CV(X)$ and $CV(Y)$ are similar, a correlation larger than $1/2$ is sufficient to ensure that the ratio estimator is preferred to the usual estimator. □

The proof of Theorem 5.5.1 is somewhat technical. To begin with observe that

$$\text{Var}\left[\hat{Y}\right] = \text{Var}\left[N\overline{y}\right] = N^2\text{Var}\left[\overline{y}\right] = N^2\left(\frac{1}{n} - \frac{1}{N}\right)S_Y^2.$$

Thus, under (5.4.5), $\frac{n}{N^2}\text{Var}\left[\hat{Y}\right] = \left(1 - \frac{n}{N}\right)S_Y^2 \to S_Y^2$ and then

$$\frac{n}{N^2}E[(\hat{Y} - Y)^2] = \frac{n}{N^2}\text{Var}\left[\hat{Y}\right] \approx S_Y^2; \tag{5.5.3}$$

note that, since $\hat{Y}$ is unbiased, $\text{Var}\left[\hat{Y}\right]$ coincides with the mean square error $E[(\hat{Y} - Y)^2]$. The next step consists in determining an approximation for the mean square error of $\hat{Y}_R$. The argument relies on the following theorem.

Theorem 5.5.2. Under 5.4.5,

$$\sqrt{n}\left(\frac{\overline{y}}{\overline{x}} - \frac{\overline{Y}}{\overline{X}}\right) \xrightarrow{\text{d}} \mathcal{N}\left(\begin{bmatrix}0\\0\end{bmatrix}, \frac{1}{\overline{X}^2}\left[S_X^2\left(\frac{\overline{Y}}{\overline{X}}\right)^2 - 2S_{XY}\frac{\overline{Y}}{\overline{X}} + S_Y^2\right]\right). \tag{5.5.4}$$

Proof. Set

$$f(x, y) = \frac{y}{x}$$

so that

$$Df(x,y) = (\partial_x f(x,y), \partial f_y(x,y)) = \left(-\frac{y}{x^2}, \frac{1}{x}\right).$$

Combining (5.4.7) with Theorem 4.4.1 it follows that

$$\sqrt{n}\left(f\begin{bmatrix}\overline{x}\\ \overline{y}\end{bmatrix} - f\begin{bmatrix}\overline{X}\\ \overline{Y}\end{bmatrix}\right) \xrightarrow{\text{d}} \mathcal{N}\left(\begin{bmatrix}0\\0\end{bmatrix}, Df(\overline{X},\overline{Y})\begin{bmatrix}S_X^2 & S_{YX}\\ S_{XY} & S_Y^2\end{bmatrix} Df(\overline{X},\overline{Y})'\right). \tag{5.5.5}$$

Next, observe that

(a) $f\begin{bmatrix}\overline{x}\\ \overline{y}\end{bmatrix} = \dfrac{\overline{y}}{\overline{x}}$,

(b) $f\begin{bmatrix}\overline{X}\\ \overline{Y}\end{bmatrix} = \dfrac{\overline{Y}}{\overline{X}}$,

(c) $Df(\overline{X},\overline{Y}) = \left(-\dfrac{\overline{Y}}{\overline{X}^2}, \dfrac{1}{\overline{X}}\right)$, and

(d) The variance in the right-hand side of (5.5.5) simplifies to

$$\begin{aligned} Df(\overline{X},\overline{Y})\begin{bmatrix}S_X^2 & S_{YX}\\ S_{XY} & S_Y^2\end{bmatrix} Df(\overline{X},\overline{Y})' &= \left(-\frac{\overline{Y}}{\overline{X}^2}, \frac{1}{\overline{X}}\right)\begin{bmatrix}S_X^2 & S_{YX}\\ S_{XY} & S_Y^2\end{bmatrix}\begin{pmatrix}-\frac{\overline{Y}}{\overline{X}^2}\\ \frac{1}{\overline{X}}\end{pmatrix} \\ &= S_X^2\left(\frac{\overline{Y}}{\overline{X}^2}\right)^2 - 2S_{XY}\frac{\overline{Y}}{\overline{X}^2}\frac{1}{\overline{X}} + S_Y^2\left(\frac{1}{\overline{X}}\right)^2 \\ &= \frac{1}{\overline{X}^2}\left[S_X^2\left(\frac{\overline{Y}}{\overline{X}}\right)^2 - 2S_{XY}\frac{\overline{Y}}{\overline{X}} + S_Y^2\right] \end{aligned}$$

Thus, (5.5.5) is equivalent to

$$\sqrt{n}\left(\frac{\overline{y}}{\overline{x}} - \frac{\overline{Y}}{\overline{X}}\right) \xrightarrow{\text{d}} \mathcal{N}\left(\begin{bmatrix}0\\0\end{bmatrix}, \frac{1}{\overline{X}^2}\left[S_X^2\left(\frac{\overline{Y}}{\overline{X}}\right)^2 - 2S_{XY}\frac{\overline{Y}}{\overline{X}} + S_Y^2\right]\right)$$

which is the desired conclusion. □

Proof of Theorem 5.5.1. Note that Theorem 5.5.2 implies that

$$\overline{X}\sqrt{n}\left(\frac{\overline{y}}{\overline{x}} - \frac{\overline{Y}}{\overline{X}}\right) \xrightarrow{\text{d}} \mathcal{N}\left(\begin{bmatrix}0\\0\end{bmatrix}, \left[S_X^2\left(\frac{\overline{Y}}{\overline{X}}\right)^2 - 2S_{XY}\frac{\overline{Y}}{\overline{X}} + S_Y^2\right]\right), \tag{5.5.6}$$

so that

$$\begin{aligned} E\left[\frac{\overline{y}}{\overline{x}}\right] &\approx \frac{\overline{Y}}{\overline{X}}, \\ \operatorname{Var}\left[\overline{X}\sqrt{n}\left(\frac{\overline{y}}{\overline{x}} - \frac{\overline{Y}}{\overline{X}}\right)\right] &\approx S_X^2\left(\frac{\overline{Y}}{\overline{X}}\right)^2 - 2S_{XY}\frac{\overline{Y}}{\overline{X}} + S_Y^2. \end{aligned} \tag{5.5.7}$$

On the other hand, using that $\overline{Y}/\overline{X} = Y/X$ and $X = N\overline{X}$, it follows that

$$\frac{\sqrt{n}}{N}\left(\hat{Y}_R - Y\right) = \frac{\sqrt{n}}{N}\left(\frac{\overline{y}}{\overline{x}}X - Y\right) = X\frac{\sqrt{n}}{N}\left(\frac{\overline{y}}{\overline{x}} - \frac{Y}{X}\right) = \overline{X}\sqrt{n}\left(\frac{\overline{y}}{\overline{x}} - \frac{\overline{Y}}{\overline{X}}\right)$$

and then the above display yields that

$$\begin{aligned}
E[\hat{Y}_R] &\approx Y \\
\frac{n}{N^2} E[(\hat{Y}_R - Y)^2] &= E\left[\left(\frac{\sqrt{n}}{N}(\hat{Y}_R - Y)\right)^2\right] \\
&= E\left[\left(\overline{X}\sqrt{n}\left(\frac{\overline{y}}{\overline{x}} - \frac{\overline{Y}}{\overline{\overline{X}}}\right)\right)^2\right] \\
&= \mathrm{Var}\left[\overline{X}\sqrt{n}\left(\frac{\overline{y}}{\overline{x}} - \frac{\overline{Y}}{\overline{\overline{X}}}\right)\right] \\
&\approx S_X^2 \left(\frac{\overline{Y}}{\overline{\overline{X}}}\right)^2 - 2S_{XY}\frac{\overline{Y}}{\overline{\overline{X}}} + S_Y^2
\end{aligned}$$

and combining this fact with (5.5.3) it follows that

$$\frac{E[(\hat{Y} - Y)^2]}{E[(\hat{Y}_R - Y)^2]} = \frac{\frac{n}{N^2}E[(\hat{Y} - Y)^2]}{\frac{n}{N^2}E[(\hat{Y}_R - Y)^2]} \approx \frac{S_Y^2}{S_X^2\left(\frac{\overline{Y}}{\overline{\overline{X}}}\right)^2 - 2S_{XY}\frac{\overline{Y}}{\overline{\overline{X}}} + S_Y^2},$$

and the proof is complete. $\square$

References

[1]. T. M. Apostol (1980), Mathematical Analysis, *Addison Wesley*, Reading, MA

[2]. A. A. Borovkov (1999), Mathematical Statistics, *Gordon and Breach*, New York

[3]. E. Dudewicz y S. Mishra (1998). Mathematical Statistics, *Wiley*, New York.

[4]. W. Fulks (1980), Cálculo Avanzado, *Limusa*, México, D. F.

[5]. F. A. Graybill (2000), Theory and Application of the Linear Model, *Duxbury*, New York.

[6]. F. A. Graybill (2001), Matrices with Applications in Statistics *Duxbury*, New York.

[7]. D. A. Harville (2008), Matrix Algebra Form a Statistician Perspective, *Springer-Verlag*, New York.

[8]. M. H. Hansen, W. N. Hurwitz, W. G. Madow (2003), Sample Survey Methods and Theory : Volume I, Methods and Applications, Wiley (Classics Library), New York

[9]. A. I. Khuri (2002), Advanced Calculus with Applications in Statistics, *Wiley*, New York.

[10]. E. L. Lehmann and G. B. Casella, (1998), Theory of Point Estimation, Springer, New York.

[11]. E. L. Lehmann and J. P. Romano, (1999), Testing Statistical Hypothesis, Springer, New York.

[12]. M. Loève (1984), Probability Theory, I, Springer-Verlag, New York.

[13]. D. C. Montgomery (2011), Introduction to Statistical Quality Control, 6th Edition, Wiley, New York.

[14]. S. Lohr (2010), Sampling: Design and Analysis 2nd Ed. Cengage, San Francisco.

[15]. A. M. Mood, D. C. Boes and F. A. Graybill (1984), Introduction to the Theory of Statistics, McGraw-Hill, New York.

[16]. W. Rudin (1984), Real and Complex Analysis, *McGraw-Hill*, New York.

[17]. H. L. Royden (2003), Real Analysis, *MacMillan*, London.

[18]. J. Shao (2010), Mathematical Statistics, *Springer*, New York.

[19]. C. E. Sarndal, B. Swensson, J. Wretman (1992), Model Assisted Survey Sampling, Springer-Verlag, New York.

[20]. H. Tucker (2002), Mathenatical Methods in Sample Surveys, Word Scientific, Singapore.

[21]. D. Wackerly, W. Mendenhall y R. L. Scheaffer (2009), Mathematical Statistics with Applications, *Prentice-Hall*, New York.