

SURVEY SAMPLING – AS I UNDERSTAND IT
(A Development of Optimality Criterion)

V. P. Godambe, University of Waterloo

This was the Gold Medalist Presentation at the Statistical Society of Canada meetings held in Victoria, 6th June 1988.

For since the fabric of the universe is most perfect and the work of a most wise Creator, nothing at all takes place in the universe in which some rule of maximum or minimum does not appear.

– Leonhard Euler

Introduction

This is a brief overview of the historical development of the optimality criterion in survey sampling theory and practice. The presentation here has been considerably simplified for it takes for granted a fundamental result. In survey sampling set-up the entire data can be effectively summarized by the set of observed units (or individual labels) together with the corresponding variate values as in (1) to follow. This is a basic discovery due to Basu. He (1958) proved that in survey-sampling set-up (1) constitutes a *minimal sufficient statistic*.

Definitions, Notation and the Problem

Survey Population P is a finite collection of individuals (houses, blocks, farms, households, etc.), each bearing a distinctive label i ; we may write

$$P = \{i:1,\dots,N\},$$

where N is the size of P . Variate under study such as income, size, produce, etc. is denoted by y . The value of y associated with the individual i is y_i , $i = 1,\dots,N$.

We want to estimate some *unknown* characteristic, say the mean

$$\bar{Y} = \sum_1^N y_i / N$$

of the population P . For this purpose a sample s of size n is drawn from P ($s \subset P$), using a *sampling design* (simple random sampling or stratified sampling, etc.) and the values y_i , $i \in s$ are ascertained through a survey.

Problem I: To estimate \bar{Y} given the data

$$d = \{(i, y_i) : i \in s\} \tag{1}$$