Biometrika (2022), **xx**, x, *pp*. 1–2

Printed in Great Britain

Estimating the size of a population through repeated sampling: a new view on capture-recapture procedures

BY G. KAUERMANN, N. SAPARGALI

Department of Statistics, Ludwig-Maximilians-Universität München, 80539 Munich, Germany

goeran.kauermann@stat.uni-muenchen.de nurzhan.sapargali@stat.uni-muenchen.de

SUMMARY

There should be a single paragraph summary which should not contain formulae or symbols, followed by some key words in alphabetical order. Typically there are 3–8 key words, which should contain nouns and be singular rather than plural. The summary contains bibliographic references only if they are essential. It should indicate results rather than describe the contents of the paper: for example, 'A simulation study is performed' should be replaced by a more informative phrase such as 'In a simulation our estimator had smaller mean square error than its main competitors.'

Some key words: Capture-recapture estimator; Inclusion probability; Population size.

1. INTRODUCTION

2. MODEL AND LIKELIHOOD

Our starting point is the basic capture-recapture model with equal capture probabilities. Consider a closed population $\mathcal{P} = \{1, \ldots, N\}$ from which we draw $T > 1$ independent samples of size n using some sampling scheme. Define S_t to be the set of individuals included into sample

t and $\pi_i = \text{pr}(i \in S_t)$ to be the inclusion probability of i. Let $X_{i(T)}$ denote the frequency count of i being included into T samples

$$X_{i(T)} = \sum_{t=1}^{T} \mathbb{1}\{i \in S_t\};$$

If $\pi_i = n/N$ for all i, then $X_{1(T)}, \ldots, X_{N(T)}$ are modeled as identically and independently distributed binomial variables

$$\text{pr}(X_{i(T)} = x_i) = \binom{T}{x_i} \left(\frac{n}{N}\right)^{x_i} \left(1 - \frac{n}{N}\right)^{T - x_i}$$

Since we only observe $X_{i(T)} > 0$, it is useful to decompose $\mathcal{P}$ into $S_T^{in} = \{i \in \mathcal{P} : X_{i(T)} > 0\}$ and $S_T^{out} = \{i \in \mathcal{P} : X_{i(T)} = 0\}$