Statistical Methodology 26 (2015) 34–60

Contents lists available at ScienceDirect

Statistical Methodology

journal homepage: www.elsevier.com/locate/stamet

Sequential negative binomial problems and statistical ecology: A selected review with new directions

Nitis Mukhopadhyay [a,*], Swarnali Banerjee [b]

[a] Department of Statistics, University of Connecticut, Storrs, CT 06269-4120, USA

[b] Department of Mathematics and Statistics, Old Dominion University, Norfolk, VA 23529, USA

ARTICLE INFO

Article history:
Received 14 July 2014
Received in revised form
27 February 2015
Accepted 27 February 2015
Available online 11 March 2015

MSC:
62P12
62L05
62L12

Keywords:
Average infestation
Coefficient of variation
Ecology
Environmental statistics
First-order properties
Fixed sample methods
Insect count
Known thatch parameter
Second-order properties
Sequential methods
SPRT
Thatch parameter
Two-stage methods
Unknown thatch parameter

ABSTRACT

Count data is abundant in entomology, more broadly, in statistical ecology. In 1949, Frank Anscombe pioneered the role of negative binomial (NB) modeling while working with insect count data. Since then, the spectrum of available research methods has grown immensely in more than past sixty years in involving count data modeled by a NB distribution. NB distribution also finds extensive use in agriculture, insect infestation, soil and weed science, etc. In this paper we have used a real dataset on potato beetle infestation (Beall, 1939) to illustrate smooth data collection under various sequential inferential procedures to draw important and practical conclusions.

We begin by selectively reviewing a majority of influential research methods for a number of formulations and their executions in the context of fixed-sample-size inferential procedures (Section 2). Subsequently, we elaborate on purely sequential and two-stage sampling methodologies for data collection (Sections 3 and 4). In Section 5, we summarize some major results with their interpretations including large-sample first- and second-order properties as appropriate. The illustrations of all the sequential inferential procedures on the real dataset gives interesting insights (Section 6). We also propose a number of selected directions for future research of substantial nature (Section 7). Finally, our own R codes are provided in the Appendix.

* Corresponding author.
E-mail addresses: nitis.mukhopadhyay@uconn.edu (N. Mukhopadhyay), swarnali009@gmail.com (S. Banerjee).

http://dx.doi.org/10.1016/j.stamet.2015.02.006

Contents

1. Introduction

Analysis of data in statistical ecology often requires an assumption about the distribution of count. For example, Anscombe [1] modeled insect count with a *negative binomial* (NB) distribution. He assumed a NB distribution with two parameters μ and κ, $0 < \mu, \kappa < \infty$:

$$f(x; \mu, \kappa) \equiv P(X = x) = \left(1 + \frac{\mu}{\kappa}\right)^{-\kappa} \frac{\Gamma(\kappa + x)}{x!\Gamma(\kappa)} \left(\frac{\mu}{\mu + \kappa}\right)^{x}, \quad x = 0, 1, 2, \ldots. \tag{1.1}$$

This distribution is referred to as NB(μ, κ). One may also refer to Johnson and Kotz [31], Boswell and Patil [11].

We note that μ represents the mean or average count whereas κ indicates the extent of clumping or thatching. A smaller (larger) value of κ indicates heavy (light) clumping.

Anscombe [2] compared a NB distribution with a number of other two-parameter distributions. Each distribution was expressed in terms of κ and $p(= \mu/\kappa)$, and consequently the third and fourth