



ELSEVIER

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](#).

© 2015 Elsevier B.V. All rights reserved.

* Corresponding author.

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

Contents

1.	Introduction.....	35
1.1.	Motivating real data illustration.....	36
1.2.	The layout of this paper.....	37
2.	Fixed-sample-size inferential problems.....	37
2.1.	Estimating parameters and fitting a NB model with n fixed.....	37
2.2.	More on estimating parameter κ with n fixed.....	38
3.	Sequential inferential problems: tests of hypotheses.....	38
4.	Sequential inferential problems: estimation.....	41
4.1.	Estimation of μ when κ is known.....	41
4.2.	Estimation of μ when κ is unknown.....	45
4.3.	Estimation of κ	45
5.	Technical details.....	45
5.1.	Sequential bounded risk estimation with κ known.....	45
5.2.	Two-stage bounded risk estimation with κ known-I.....	46
5.3.	Two-stage bounded risk estimation with κ known-II.....	47
5.4.	Two-stage bounded risk estimation with κ unknown.....	47
6.	Illustrations using potato beetle data.....	49
6.1.	Sequential probability ratio test (3.1)–(3.5).....	49
6.2.	Purely sequential estimation of μ with known κ	50
6.3.	Two-stage estimation of μ for known κ	51
6.4.	Two-stage estimation of μ for unknown κ	51
7.	Selected future directions.....	52
7.1.	Possible extensions of a sequential probability ratio test.....	52
7.2.	Sequential multiple hypotheses tests.....	52
7.3.	Sequential selection methods.....	52
7.4.	Sequential and non-sequential simultaneous estimation.....	53
7.5.	Change-point problems.....	53
7.6.	Some additional thoughts.....	53
	Acknowledgments.....	54
	Appendix. The computer codes.....	54
A.1.	Downloadable and executable program generating Table 2.....	54
A.2.	R code for Table 3.....	54
A.3.	R code for Table 4.....	55
A.4.	R code for Table 5.....	56
A.5.	R code for Table 6.....	56
A.6.	R code for Table 7.....	57
A.7.	R code for Table 8.....	58
A.8.	R code for Table 9.....	58
	References.....	58

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 , $0 < \mu, \kappa < \infty$:

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

This distribution is referred to as $NB(\mu, \kappa)$. 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

Download English Version:

<https://daneshyari.com/en/article/1151648>

Download Persian Version:

<https://daneshyari.com/article/1151648>

[Daneshyari.com](https://daneshyari.com)