Journal of Multivariate Analysis 118 (2013) 138–147

Contents lists available at SciVerse ScienceDirect

Journal of Multivariate Analysis

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

Note(s)

Delta and jackknife estimators with low bias for functions of binomial and multinomial parameters

Christopher S. Withers [a], Saralees Nadarajah [b,*]

[a] Applied Mathematics Group, Industrial Research Limited, Lower Hutt, New Zealand
[b] School of Mathematics, University of Manchester, Manchester M13 9PL, UK

ARTICLE INFO

Article history:
Available online 17 February 2013

AMS subject classification:
62F
62G

Keywords:
Binomial distribution
Delta method
Jackknife
Low bias
Multinomial distribution

ABSTRACT

An estimator is said to be *of order* $s > 0$ if its bias has magnitude n^{-s}, where n is the sample size. We give *delta estimators* and *jackknife estimators* of order four for smooth functions of the parameters of a multinomial distribution. An unbiased estimator is given for its density function. We also give a jackknife estimator of *any* order for smooth functions of the binomial parameter.

The jackknife estimator of order s has a simpler form than the delta estimator of order s. On the other hand, the jackknife estimator, like the bootstrap, requires $\sim n^{s-1}$ calculations while the delta estimator of order s requires only $\sim n$ calculations.

Examples include the log odds ratio, the survival function and the Shannon information or entropy.

1. Introduction

The multinomial distribution is the most popular model for multivariate discrete data [7]. Its applications are widespread. We mention: models to cluster Internet traffic [8], funding source and research report quality in nutrition practice-related research, crash-prediction models for multilane roads, pollen counts, changepoints in the north Atlantic tropical cyclone record, magazine and Internet exposure, genome analysis [1], fish diet compositions from multiple data sources, statistical alarm method for mobile gamma spectrometry, stylometric analyses, clinical trials [5], impacts of movie reviews on box office, amount individuals withdraw at cash machines, soil microbial community, longline hook selectivity for red tilefish Branchiostegus japonicus in the East China Sea [15], gambling by auctions, automatic image annotation, and probabilities for the first division Spanish soccer league [3].

The aim of this note is to provide estimators with low bias for smooth functionals of multinomial parameters. An estimator is said to be *of order* s if its bias has magnitude n^{-s}, written $O(n^{-s})$ or $\sim n^{-s}$, where n is the sample size.

In Section 2, we give unbiased estimators (UEs) for analytic functions of the parameters of a multinomial distribution, such as its density function. In Sections 3 and 4, we give *delta estimators* and *jackknife estimators* of order four for smooth functions of the parameters of a multinomial distribution. Section 4 also gives a jackknife estimator of *any* order for smooth functions of the binomial parameter. Examples include the log odds ratio, the survival function and the Shannon's entropy. One of these examples is illustrated by means of a simulation study in Section 6.

* Corresponding author.
E-mail address: saralees.nadarajah@manchester.ac.uk (S. Nadarajah).

0047-259X/$ – see front matter © 2013 Elsevier Inc. All rights reserved.
doi:10.1016/j.jmva.2013.02.006

We use the following notation. We set $\mathbb{N} = \{0, 1, 2, \ldots\}$. For $q \geq 2$, $x \in \mathbb{N}^q$, $N \in \mathbb{N}^q$, $\theta \in \mathbb{S}_{q-1}$, and $n \in \mathbb{N}$, where $\mathbb{S}_q$ denotes the standard q-simplex, we set

$$|x| = \sum_{i=1}^{q} x_i, \qquad x! = \prod_{i=1}^{q} x_i!, \qquad \theta^x = \prod_{i=1}^{q} \theta_i^{x_i},$$
$$\binom{n}{x} = n!/x!, \qquad [n]_{x_1} = n!/\left(n - x_1\right)! = n\left(n - x_1\right) \cdots \left(n - x_1 + 1\right),$$
$$[N]_x = \prod_{i=1}^{q} [N_i]_{x_i}.$$

The basic model assumed is

$$N \sim \text{Multinomial}_q(n; \theta), \tag{1.1}$$

where

$$\sum_{i=1}^{q} N_i = n, \qquad \sum_{i=1}^{q} \theta_i = 1$$

with the density function

$$P(N = x) = \begin{cases} \binom{n}{x} \theta^x = m_q(x : n, \theta), & \text{if } x \in \mathbb{N}^q \text{ and } |x| = n, \\ 0, & \text{otherwise.} \end{cases} \tag{1.2}$$

We let $t(\cdot) : \mathbb{S}_{q-1} \to \mathbb{R}$ denote a smooth function with all of the required partial derivatives at θ finite (that is, the partial derivatives required by the estimation method being used).

Sections 2 and 3 are based on Withers [13]. They also give a method for expanding $\mathbb{E}\left[t(\widehat{\theta})\right]$ in powers of n^{-1} for a wide class of estimators $\widehat{\theta}$. Here, we apply these results to the UE

$$\widehat{\theta} = N/n. \tag{1.3}$$

Since

$$\theta_q = 1 - \sum_{i=1}^{p} \theta_i,$$

where $p = q - 1$, a function of θ, say $t(\theta)$, is actually a function of $\theta_1, \ldots, \theta_p$. That is, $t(\cdot) : \mathbb{S}_{p-1} \to \mathbb{R}$. With this understanding, we can write

$$\partial_i = \partial/\partial\theta_i, \qquad t_{\cdot i_1 \cdots i_r}(\theta) = \partial_{i_1} \cdots \partial_{i_r} t(\theta)$$

for $1 \leq i, i_1, \ldots, i_r \leq p$. For an extension of Withers [13] to more than one sample, we refer the readers to Withers and Nadarajah [14].

2. Unbiased estimators

For $x \in \mathbb{N}^q$, an UE of θ^x is

$$\begin{aligned} \widehat{\theta}_{(x)} &= [N]_x/[n]_{|x|} \\ &= \left\{ \prod_{i=1}^{p} \left[N_i\left(N_i - 1\right) \cdots \left(N_i - x_i + 1\right)\right] \right\} \Big/ \left[n(n-1) \cdots (n - |x| + 1)\right] \\ &= \left\{ \prod_{i=1}^{p} \left[\widehat{\theta}_i\left(\widehat{\theta}_i - n^{-1}\right) \cdots \left(\widehat{\theta}_i - \left(x_i - 1\right) n^{-1}\right)\right] \right\} \Big/ \left[1\left(1 - n^{-1}\right) \cdots \left(1 - (|x| - 1)\, n^{-1}\right)\right], \end{aligned}$$

as noted in Example 2.7 of Withers [13]. For,

$$\mathbb{E}\left[s^N\right] = \left(s'\theta\right)^n \tag{2.1}$$

for $s_i > 0,\ i = 1, 2, \ldots, q$. Multiplying (2.1) by $(\partial/\partial\theta)^x = \prod_{i=1}^{q} (\partial/\partial\theta_i)^{x_i}$, we obtain

$$\mathbb{E}\left[[N]_x s^{N-x}\right] = [n]_x \theta^x \left(s'\theta\right)^{n-x},$$

and so the result follows by setting $s_i \equiv 1$.