Delta and Jackknife Estimators with Low Bias for Functions of Binomial and Multinomial Parameters

Christopher S. Withers & Saralees Nadarajah

First version: 10 December 2013
Delta and jackknife estimators with low bias for functions of binomial and multinomial parameters

by

Christopher S. Withers
Applied Mathematics Group
Industrial Research Limited
Lower Hutt, NEW ZEALAND

Saralees Nadarajah
School of Mathematics
University of Manchester
Manchester M13 9PL, UK

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.

Keywords: Binomial distribution; Delta method; Jackknife; Low bias; Multinomial distribution

1 Introduction

The multinomial distribution is the most popular model for multivariate discrete data (Johnson et al., 1997). Its applications are widespread. We mention: models to cluster Internet traffic (Jorgensen, 2004), 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 (Chang and Wang, 2011), fish diet compositions from multiple data sources, statistical alarm method for mobile gamma spectrometry, stylometric analyses, clinical trials (Ganju and Zhou, 2011), 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 (Yamashita et al., 2009), gambling by auctions, automatic image annotation, and probabilities for the first division Spanish soccer league (Diaz-Emparanza and Nunez-Anton, 2010).

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.

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

$$|x| = \sum_{i=1}^{q} x_i, \quad x! = \prod_{i=1}^{q} x_i!, \quad \theta^x = \prod_{i=1}^{q} \theta_i^{x_i},$$

$$\binom{n}{x} = n! / x! \cdot \binom{n}{x_1} = n! / (n - x_1)! = n (n - x_1) \cdots (n - x_1 + 1),$$

$$[N]_x = \prod_{i=1}^{q} \binom{N_i}{x_i}.$$ 

The basic model assumed is

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

where

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

with the density function

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

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

Sections 2 and 3 are based on Withers (1987). They also give a method for expanding $E[t(\hat{\theta})]$ in powers of $n^{-1}$ for a wide class of estimators $\hat{\theta}$. Here, we apply these results to the UE

$$\hat{\theta} = N/n. \quad \text{(1.3)}$$

Since

$$\theta_q = 1 - \sum_{i=1}^{p} \theta_i,$$
where \( p = q - 1 \), a function of \( \theta \), 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

\[
\partial_i = \partial/\partial \theta_i, \quad t_{i_1 \ldots i_t}(\theta) = \partial_{i_1} \cdots \partial_{i_t} t(\theta)
\]

for \( 1 \leq i, i_1, \ldots, i_t \leq p \). For an extension of Withers (1987) to more than one sample, we refer the readers to Withers and Nadarajah (2008).

### 2 Unbiased estimators

For \( x \in \mathbb{N}^q \), an UE of \( \theta^x \) is

\[
\hat{\theta}(x) = [N]_x/[n]_x
\]

\[
= \prod_{i=1}^p \frac{[N_i(N_i - 1) \cdots (N_i - x_i + 1)]}{[n(n-1) \cdots (n - x)]}
\]

\[
= \prod_{i=1}^p \left[ \hat{\theta}(\hat{\theta}_i - n^{-1}) \cdots (\hat{\theta}_i - (x_i - 1)n^{-1}) \right] / \left[ (1 - n^{-1}) \cdots (1 - (x_i - 1)n^{-1}) \right],
\]

as noted in Example 2.7 of Withers (1987). For,

\[
\mathbb{E} \left[ s^N \right] = (s^t\theta)^n
\]  

(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

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

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

We now suppose that \( t(\theta): \mathbb{S}_{q-1} \to \mathbb{R} \) has a Taylor series expansion about \( 0 \in \mathbb{N}^q \),

\[
t(\theta) = \sum \{ \theta^x t_x(0)/x! : x \in \mathbb{N}^q \}.
\]

**Example 2.1** For the binomial distribution, we take \( q = 2, x = (1,0) \) and \( x = (2,0) \). We have

\[
N_1/n - N_1(N_1 - 1)/n(n-1)
\]

as an UE of \( \theta_1 - \theta_1^2 \). Equivalently, taking \( x = (1,1) \), an UE of \( \theta_1\theta_2 \) is \( N_1N_2/n(n-1) \).

Taking \( x = (2,3) \), an UE of \( \theta_1^2\theta_2^3 \) is

\[
\frac{N_1(N_1 - 1)N_2(N_2 - 1)(N_2 - 2)/n(n-1)(n-2)(n-3)(n-4)}{\hat{\theta}_1(\hat{\theta}_1 - n^{-1})\hat{\theta}_2(\hat{\theta}_2 - n^{-1})(\hat{\theta}_2 - 2n^{-1}) / [1(1 - n^{-1})(1 - 2n^{-1})(1 - 3n^{-1})(1 - 4n^{-1})]}
\]

More generally,

\[
\hat{t}_U = \sum \{ \hat{\theta}(x)t_x(0)/x! : x \in \mathbb{N}^q \}
\]

3
is an UE of $t(\theta)$ for large values of $x$. Also, $\hat{t} = t(\hat{\theta})$ and $\hat{t} = \hat{t}_U$ both satisfy

$$\mathbb{E} \left[ \hat{t} \right] = t(\theta) + O\left(n^{-1}\right),$$

$$\text{var} \left( \hat{t} \right) = vn^{-1} + O\left(n^{-2}\right), \quad \mathbb{E} \left[ (\hat{t} - t(\theta))^2 \right] = vn^{-1} + O\left(n^{-2}\right),$$

where

$$v = \sum_{i,k=1}^{p} t_{ij}k_{ij}, \quad k_{ij} = \delta_{ij}\theta_i - \theta_i \theta_j,$$

where $\delta_{ij} = 1$ or $0$ for $i = j$ or $i \neq j$. For expansions for the cumulants of $t(\hat{\theta})$ in powers of $n^{-1}$, we refer the readers to Withers (1983).

**Example 2.2** An UE of the multinomial density function (1.2) is

$$\hat{m}_q(x : n) = \left[ N \right]_x / x!. $$

For example, taking $q = 2$, an UE of the binomial density function,

$$P(N_1 = x_1) = P(N_1 = x_1, N_2 = n - x_1) = \binom{n}{x_1} \theta_1^{x_1} (1 - \theta_1)^{n-x_1},$$

is

$$\left[ N \right]_x / x! = \left[ N_1 \right]_{x_1} [n - N_1]_{n-x_1} / x! (n - x_1)!. $$

We end this section with a trivial but important example.

**Example 2.3 System failure.** We consider a system with $I$ independent components in series in a given time interval. If one component fails then the system fails. We set $p_i = P$(the $i$th component does not fail). So, the system does not fail in that time interval with probability $t(p) = p_1 \cdots p_I$. If we observe $N_i \sim \text{Binomial}(n_i, p_i)$, $i = 1, \ldots, I$

then an UE of $t(p)$ is $t(\hat{p})$, where $\hat{p}_i = N_i / n_i$.

### 3 Delta estimators for the multinomial distribution

A *delta estimator* is the name we give to an estimator obtained by the Taylor series method of Withers (1987). We suppose that $\hat{\theta} \in S_{p-1}$ is now any UE of any parameter $\theta \in S_{p-1}$ with $r$th order cumulants

$$\kappa\left( \hat{\theta}_1, \ldots, \hat{\theta}_r \right) = n^{1-r}k_{\hat{\theta}_1 \cdots \hat{\theta}_r}$$

for $r \geq 2$, where $k_{\hat{\theta}_1 \cdots \hat{\theta}_r}$ is a function of $\theta$. For $a, b, \ldots \in \mathbb{N}$, we set

$$|a| = \sum_{i_1, \ldots, i_a=1}^{p} t_{i_1 \cdots i_a}(\theta)k_{i_1 \cdots i_a},$$

$$|ab| = \sum_{i_1, \ldots, i_a, j_1, \ldots, j_b=1}^{p} t_{i_1 \cdots i_a, j_1 \cdots j_b}(\theta)k_{i_1 \cdots i_a}k_{j_1 \cdots j_b},$$

4
and so on. For example,

\[ |222| = \sum_{i_1 \ldots i_6 = 1}^p t_{i_1 \ldots i_6} (\theta) k^{i_1 i_2} k^{i_3 i_4} k^{i_5 i_6}. \]

We then have

\[ \mathbb{E} \left[ t (\hat{\theta}) \right] = \sum_{j=1}^{\infty} n^{-j} C_j (\theta, t), \]

where

\[
\begin{align*}
C_1 (\theta, t) &= |2|/2, \\
C_2 (\theta, t) &= |3|/3! + |22|/8, \\
C_3 (\theta, t) &= |4|/4! + |23|/12 + |222|/48,
\end{align*}
\]

and so on. Also for \( s \geq 1 \), an estimator of \( t (\theta) \) with bias \( \sim n^{-s} \), the delta estimator of order \( s \), is given in terms of \( t_i \) of (3.3) by

\[ \hat{t}_{D_s} = t_{D_s} (n, \hat{\theta}) = \sum_{i=0}^{s-1} n^{-i} t_i (\hat{\theta}), \quad (3.2) \]

where

\[
\begin{align*}
t_0 (\theta) &= t (\theta), \\
t_i (\theta) &= - \sum_{j=1}^{i} C_j (\theta, t_{i-j})
\end{align*}
\]

for \( i \geq 1 \). So, the \( j \)th coefficient, \( C_j (\theta, t) \), requires the cumulants of \( \hat{\theta} \) up to order \( j + 1 \) and the partial derivatives of \( t (\theta) \) up to order \( 2j \). Also, the delta estimator of order \( s \), \( \hat{t}_{D_s} \), requires the cumulants of \( \hat{\theta} \) up to order \( s \) and the partial derivatives of \( t (\theta) \) up to order \( 2s - 2 \). Formulas for \( t_i (\theta) \) for \( i = 1, 2, 3 \) are given in Appendix D of Withers (1987).

**Univariate estimators.**

If \( p = 1 \), we write (3.1) as

\[ \kappa_r (\hat{\theta}) = n^{1-r} k_r, \quad r \geq 2, \]

and denote the \( r \)th derivative of any function \( t (\theta) \) by \( t_r \). So,

\[ |ab\cdots| = t_{a+b+\cdots} (\theta) k_a k_b \cdots \]

and \( t_i (\theta) \) of (3.3) can be written as

\[
\begin{align*}
t_1 (\theta) &= - C_1 (\theta, t) = - t_2 v/2, \quad \text{where} \ v = k_2, \\
t_2 (\theta) &= - t_1 t_2 v/2 - t_3 k_3/6 - t_4 v^2/8, \\
t_3 (\theta) &= - t_2 t_2 v/2 - t_1 t_3 k_3/6 - t_4 t_4 v^2/8 - t_5 k_4/24 - t_5 v k_3/12 - t_6 v^3/48.
\end{align*}
\]

From these, or from Appendix D of Withers (1987), we obtain

\[ t_i (\theta) = \sum_{j=2}^{2i} A_{ij} t_j, \quad i \geq 1, \quad (3.4) \]
where
\[
A_{12} = -v/2, \\
A_{22} = v v_{1}/4, \\
A_{23} = v v_{1}/2 - k_{3}/6, \\
A_{32} = -v^{2} v_{4}/16 - v v_{2} v_{3}/4 - v v_{2}^{2}/8 + v_{3} k_{3}/12, \\
A_{33} = -v^{2} v_{2}/4 - v v_{1} v_{2} + 3 v_{2} k_{3}/12 + v_{2} k_{3}^{2}/4, \\
A_{34} = -3 v^{2} v_{2}/8 - 5 v v_{1}^{2}/8 + v_{2} k_{3}/12 + v_{1} k_{3}/4 - k_{4}/24, \\
A_{35} = -v^{2} v_{1}/4 + v k_{3}/12, \\
A_{36} = -v^{3}/48.
\]

To apply (3.2), (3.3) to the multinomial estimator, \( \hat{\theta} \), of (1.3), or (3.4) to the binomial estimator, one simply substitutes

\[
k^{ij} = \delta_{ij} \theta_{i} - \theta_{i} \theta_{j},
\]

\[
k^{ijk} = \begin{cases} 
\theta_{i} (1 - \theta_{i}) (1 - 2 \theta_{i}), & \text{if } i = j = k, \\
-\theta_{i} (1 - 2 \theta_{i}) \theta_{k}, & \text{if } i = j \neq k, \\
2 \theta_{i} \theta_{j} \theta_{k}, & \text{if } i, j, k \text{ are distinct}, 
\end{cases}
\]

\[
k^{ijkl} = \begin{cases} 
\theta_{i} \bar{\theta}_{l} (1 - 6 \theta_{i} \bar{\theta}_{l}), & \text{if } i = j = k = l, \\
-\theta_{i} (1 - 6 \theta_{i} \bar{\theta}_{l}) \theta_{l}, & \text{if } i = j = k \neq l, \\
-\theta_{i} \theta_{k} (1 - 2 \theta_{i} - 2 \theta_{k} + 6 \theta_{i} \theta_{k}), & \text{if } i = j \neq k = l, \\
2 \theta_{i} (1 - 3 \theta_{i}) \theta_{j} \theta_{k}, & \text{if } i = l, j, k \text{ are distinct}, \\
-6 \theta_{i} \theta_{j} \theta_{k} \theta_{l}, & \text{if } i, j, k, l \text{ are distinct}, 
\end{cases}
\]  

(3.5)

where \( \bar{\theta}_{i} = 1 - \theta_{i} \). (The cumulants in (3.5) were given by Wishart (1949), and quoted in Johnson et al. (1997).) The cumulants in (3.5) allow for computation of fourth order delta estimators. For example,

\[
2 t_{1}(\theta) = - \sum_{i,j=1}^{p} (\delta_{ij} \theta_{i} - \theta_{i} \theta_{j}) t_{ij}(\theta) = \sum_{i,j=1}^{p} \theta_{i} \theta_{j} t_{ij}(\theta) - \sum_{i=1}^{p} \theta_{i} t_{ii}(\theta),
\]

and

\[
\hat{t}_{D2} = t\left(\bar{\theta}\right) - n^{-1} t_{1}\left(\bar{\theta}\right)
\]

estimates \( t(\theta) \) with bias \( O(n^{-2}) \). If one wants a delta estimator of order greater than four then higher order cumulants will be needed.

4 Examples of delta estimators

We now give some examples where \( t(\theta) \) or its derivatives become infinite at \( \theta_{1} = 0 \) or 1. In this case, there are several ways in which the method proposed above can be adapted. The first is to choose constants \( p_{1}, p_{2} \) in \((0,1)\) (to be equal to say \( 1/2 \)) and set

\[
\tilde{\theta}_{i} = \begin{cases} 
p_{1}/n, & \text{if } N_{i} = 0, \\
1 - p_{2}/n, & \text{if } N_{i} = n, \\
N_{1}/n, & \text{if } 0 < N_{i} < n.
\end{cases}
\]
One may then show that, for \( r \geq 1 \), \( \tilde{\theta}_i \) has \( r \)th cumulant

\[
\kappa_r (\tilde{\theta}_i) = n^{1-r} \kappa_r (Y_i) + O(\epsilon^n) \quad \text{for } 0 < \epsilon < \min (p_1, 1 - p_2), \quad Y_i \sim \text{Binomial} (1, \theta_i).
\]

So, by Withers (1987), the above method applies with \( \hat{\theta}_i = N_i/n \) replaced by \( \tilde{\theta}_i \). We set

\[
\tilde{t}_{Ds} = t_{Ds} \left( n, \tilde{\theta}_i \right)
\]

for \( t_{Ds}(n, \cdot) \) of (3.2). The estimator, \( \tilde{t}_{Ds} \), has finite bias and may still perform creditably, even though the regularity conditions for the bias to be \( \sim n^{-s} \) have been violated.

A second way of dealing with \( t(\hat{\theta}) \) when \( t(\theta) \) or its derivatives are unbounded, is simply to replace them by zero. This approach again introduces bias which is again exponentially small. A third way is to make our estimators conditional on \( 0 < N_i < 1, \quad i = 1, \ldots, q \). One can show that the difference between conditional and unconditional cumulants is exponentially small, so that the extra bias introduced is exponentially small.

We start with a trinomial example.

**Example 4.1 The survival function.** We suppose that \( X \) is a real random variable with distribution function \( F(x) = P(X \leq x) \). For \( x_1 < x_2 \), the survival function is

\[
t(\theta) = P(X > x_2|X > x_1) = P(X > x_2)/P(X > x_1) = \theta_2/(1 - \theta_1),
\]

where

\[
\begin{align*}
\theta_1 &= P(X \leq x_1) = F(x_1) = 1 - q \text{ say}, \\
\theta_3 &= P(x_1 < X \leq x_2) = F(x_2) - F(x_1), \\
\theta_2 &= P(x_2 < X) = 1 - F(x_2) = p \text{ say}.
\end{align*}
\]

We assume a random sample \( X_1, \ldots, X_n \) with distribution function \( F(x) \). We then have

\[
N \sim \text{Multinomial}_3(n; \theta),
\]

where

\[
N_1 = \sum_{i=1}^{n} I(X_i \leq x_1), \quad N_3 = \sum_{i=1}^{n} I(x_1 < X_i \leq x_2), \quad N_2 = \sum_{i=1}^{n} I(x_2 \leq X_i),
\]

where \( I(A) = 1 \) or 0 for \( A \) true or false; that is, \( I(\cdot) \) is the indicator function. The non-zero derivatives of \( t(\theta) \) are

\[
t_{1\ldots1} = r!pq^{-r-1}, \quad t_{1\ldots12} = r!q^{-r-1}, \quad r \geq 0,
\]

where 1 is repeated \( r \) times. Also it appears that each \( |a_1b_1\ldots| \) of Section 3 is zero, so that \( C_i(\theta, t) = 0 \) and \( t(\theta) = N_2/(n - N_1) \) would be an UE of \( t(\theta) \), except that \( t(\hat{\theta}) = \infty \) if \( N_1 = n \). But \( t(\tilde{\theta}) \) estimates \( t(\theta) \) with exponentially small bias.

An alternative estimator is

\[
\tilde{t} = n^{-1} \sum_{i=1}^{n} Z_i,
\]
where
\[ Z_i = I (X_i > x_2 | X_i > x_1) \sim \text{Binomial} (1, t(\theta)). \]

So,
\[ n\tilde{t} \sim \text{Binomial} (n, t(\theta)) \]
and \( \tilde{t} \) is an UE of \( t(\theta) \). Their relative efficiency is
\[
\lim_{n \to \infty} \frac{\text{var} \left( t(\hat{\theta}) \right)}{\text{var} \left( t(\tilde{\theta}) \right)} = \frac{v_1}{v_2} = 2(1 - q)/q > 1
\]
if and only if \( q < 2/3 \) if and only if \( \theta_1 > 1/3 \) since
\[
\text{var} \left( t(\hat{\theta}) \right) \approx \frac{v_1}{n}, \quad \text{where} \quad v_1 = p(1 - q)q^{-3}(q - p),
\]
\[
\text{var} \left( t(\tilde{\theta}) \right) = \frac{v_2}{n} = pq^{-2}(q - p)/n.
\]

We now give three examples that can be reduced to the univariate case, \( p = 1 \).

To apply (3.4) to \( n\hat{\theta} = N_1 \sim \text{Binomial}(n, \theta) \),
\[
(4.1)
\]
that is, \( r = 2 \), we substitute
\[ v = k_2 = \theta - \theta^2, \quad k_3 = vv_1, \quad k_4 = v - 6v^2, \]
so that
\[
v_1 = 1 - 2\theta, \quad v_1^2 = 1 - 4v, \quad v_2 = -2, \\
k_{3,1} = 1 - 6v, \quad k_{3,2} = -6v_1, \quad A_{12} = A_{22} = A_{32} = -v/2, \quad A_{23} = vv_1/3, \quad A_{24} = v^2/8, \\
A_{33} = 0, \quad A_{34} = v(-1 + 6v)/4, \quad A_{35} = -v^2v_1/6, \quad A_{36} = -v^3/48.
\]

For binomial examples with \( p = 1 \), it is convenient to set \( w = \theta^{-1} \).

**Example 4.2 The odds ratio**, \( t(\theta) = \theta_2/\theta_1 = 1 - \theta_1^{-1} \).

In this case, \( t_r = r! \left( -\theta_1 \right)^{-1-r} \) for \( r \geq 1 \), so that
\[
t_1(\theta) = w^2 - w, \quad t_2(\theta) = -2w^2 + 2w, \\
t_3(\theta) = w(w - 1) \left( w^2 - 6w + 12 \right).
\]

**Example 4.3 The log odds ratio**, \( t(\theta) = \log (\theta_2/\theta_1) \).
We first consider $t(\theta) = -\log \theta_1$. We then have $t_r = (r-1)!(-w)^r$, $r \geq 1$ and (4.1) gives

\[ 2t_1(\theta) = 1 - w, \quad 12t_2(\theta) = w^2 - 1, \]
\[ 2t_3(\theta) = (w - 1)(3w^2 + 2w - 4). \]  

(4.3)

We now consider the log odds ratio,

\[ t(\theta) = -\log \theta_1 + \log \theta_2, \]

and set

\[ w_i = \theta_i^{-1}, \quad \gamma_j = \theta_1^{-j} - \theta_2^{-j}, \quad v = \theta_1 \theta_2. \]

We have, by (4.3),

\[ 2t_1(\theta) = -\gamma_1, \quad 12t_2(\theta) = \gamma_2, \quad 2t_3(\theta) = -6\gamma_1 - 2\gamma_2 + 3\gamma_3. \]

Example 4.4 The entropy, or Shannon information for the multinomial distribution is

\[ t(\theta) = -\sum_{i=1}^{q} \theta_i \log \theta_i. \]  

(4.4)

We first consider estimating $t(\theta) = -\theta_1 \log \theta_1$. So, for $r \geq 2$, $t_r = (r-2)!(-\theta_1)^{1-r}$. So, by (4.2),

\[ 2t_1(\theta) = 1 - \theta_1, \quad 12t_2(\theta) = 2\theta_1 - 3 + \theta_1^{-1}, \quad t_3(\theta) = 4\theta_1 - 9 + 6\theta_1^{-1} - \theta_1^{-2}. \]

So, for the Shannon information, (4.4),

\[ 2t_1(\theta) = \sum_{i=1}^{q} (1 - \theta_i) = q - 1, \]
\[ 12t_2(\theta) = 2 - 3q + \sum_{i=1}^{q} \theta_i^{-1}, \]
\[ t_3(\theta) = \sum_{i=1}^{q} (4\theta_i - 9 + 6\theta_i^{-1} - \theta_i^{-2}) = 4 - 9q + 6 \sum_{i=1}^{q} \theta_i^{-1} - \sum_{i=1}^{q} \theta_i^{-2}. \]

The delta estimator of order $s \leq 4$, $\hat{t}_{D_s}$, is now given by (3.2). For example,

\[ \hat{t}_{D2} = t(\hat{\theta}_1) + n^{-1}(q - 1)/2. \]

At $\hat{\theta}_1 = 0$ or 1, $v = 0$ so that $\mathbb{E}[\hat{t}_i(\hat{\theta}_1)] = \pm \infty$ for $i \geq 2$. That is, regularity conditions on $\hat{t}_{D_s}$ break down for $s > 2$, and only $\hat{t}_{D1}$ and $\hat{t}_{D2}$ are sensible estimators.

An alternative estimator is given by Cook et al. (1974).

There are many different ways of expressing $\hat{t}_{D_s}$ to within $O_p(n^{-s})$ as an expression not involving derivatives (for example, by replacing $t_i(\theta)$ by $n[t(\theta + e_{ir}/n) - t(\theta)]$, where as above $e_{ir}$ is the $i$th unit vector in $\mathbb{R}^r$). Section 5 gives one such result, but by a route of jackknifing.
5 Jackknife estimators for the multinomial distribution

We first give a formula for a jackknife estimator of $t(\theta)$ with bias $\sim n^{-s}$ for any given $s \geq 1$ when

$$\hat{\theta} = S_n/n \text{ for } S_n = \frac{1}{n} \sum_{i=1}^{n} X_i,$$

the mean of a random sample from any distribution on $\mathbb{R}^r$ with finite moments. We call this the *jackknife estimator of order $s$*. We then provide an application to the multinomial problem.

The usual jackknife estimator with bias $\sim n^{-2}$ is

$$\hat{t}_{J2} = nt\left(\hat{\theta}\right) - (n - l)\bar{t}_{n-1},$$

where $\bar{t}_{n-j}$ is the mean of all $\binom{n}{j}$ estimators formed by dropping $j$ of the sample values (Gray and Schucany, 1972; Shao and Tu, 1995).

So,

$$\bar{t}_{n-1} = n^{-1} \sum_{i=1}^{n} t\left((S_n - X_i) / (n - 1)\right).$$

For $s \geq 1$, equation (4.17) of Schucany et al. (1971) introduced a jackknife estimator with bias $\sim n^{-s}$, $\hat{t}_{Js}$, as the ratio of two determinants. It can be shown that

$$\hat{t}_{Js} = \left\{ \sum_{j=0}^{s-1} (-1)^j (n - j)^{s-j} \bar{t}_{n-j} \right\} / (s - 1)!.$$

(5.1)

In general, $\bar{t}_{n-j}$ requires $j$ summations and so $\sim n^j$ calculations, so that $\hat{t}_{Js}$ requires $\sim n^{s-1}$ calculations. For $s \geq 3$, this is a severe computational disadvantage compared to $\hat{t}_{Ds}$, as $\hat{t}_{Ds}$ only requires $\sim n$ calculations, given $s$.

We now consider the multinomial problem of (1.1), (1.3). So, the above results hold with $X_1 \sim \text{Multinomial}_r(1; \theta)$ and $S_n = N$. Also for $1 \leq i \leq r$, $N_i$ of $\{X_1, \ldots, X_n\}$ is equal
to \(e_{ir}\), the \(i\)th unit vector in \(\mathbb{R}^r\). So,

\[
\tilde{\tau}_n = t(\hat{\theta}) = t(N/n),
\]

\[
\tilde{\tau}_{n-1} = n^{-1}\sum_j N_j t((N - e_{jr})/(n - 1)),
\]

\[
\tilde{\tau}_{n-2} = \left(\frac{n}{2}\right)^{-1} \left\{ \sum_j \left(\frac{N_j}{2}\right) t ((N - 2e_{jr})/(n - 2))
\right. \\
\left. + \sum_{i<j} N_iN_j t((N - e_{ir} - e_{jr})/(n - 2)) \right\},
\]

\[
\tilde{\tau}_{n-3} = \left(\frac{n}{3}\right)^{-1} \left\{ \sum_j \left(\frac{N_j}{3}\right) t ((N - 3e_{jr})/(n - 3))
\right. \\
\left. + \sum_{i\neq j} \left(\frac{N_i}{2}\right) N_j t ((N - 2e_{ir} - e_{jr})/(n - 3)) 
\right. \\
\left. + \sum_{i<j<k} N_iN_jN_k t ((N - e_{ir} - e_{jr} - e_{kr})/(n - 3)) \right\},
\]

and so on, where \(i, j, k\) range over \(1, \ldots, r\). So, (5.1)-(5.5) give us an estimator of bias \(\sim n^{-4}\), which we now compare with the delta estimator of order \(s\) given by (3.2)-(3.5). In particular, if \(N_1 \sim \text{Binomial}(n, \theta_1)\), we have the simple formula

\[
\tilde{\tau}_{n-j} = \left(\frac{n}{j}\right)^{-1} \sum_{k=0}^{j-1} \left(\frac{N_i}{k}\right) \left(\frac{n-N_1}{j-k}\right) t((N-k)/(n-j)).
\]

So, for the binomial problem with \(s > 1\), the jackknife estimator \(\hat{\tau}_{js}\) of (5.1)-(5.5) has a simpler form than the delta estimator \(\hat{\tau}_{ds}\) of (3.2), (3.4). On the other hand, this jackknife estimator, like the bootstrap, requires \(\sim n^{s-1}\) calculations while the delta estimator requires only \(\sim n\) calculations (Efron, 1981).

If an argument of \(t(\cdot)\) in (5.3)-(5.6) does not lie in \([0, 1]\), then \(t(\cdot)\) is not defined. If such a formula multiplies \(t(\cdot)\) by zero, then the product is to be interpreted as zero. For example, if \(\hat{\theta} = 0\) in (5.3), then \(\hat{\theta} \cdot t((n\hat{\theta} - 1)/(n - 1))\) is to be interpreted as zero.

\section{A simulation study}

In this section, we assess the performance of the bias corrected estimators given by (3.2) and (3.3). In particular we ask the question: do these estimators also reduce the mean squared errors in addition to reducing the biases? This question will be the subject of future research. Here, we answer the question by a simulation study.

We consider Example 4.2 concerning the odds ratio. This example gives four estimators for the odds ratio: \(\hat{\tau}_{D1} = 1 - \hat{\omega}\), the maximum likelihood estimator; \(\hat{\tau}_{D2} = \hat{\tau}_{D1} + (1/n)(\hat{\omega}^2 - \hat{\omega})\), the delta estimator of order one; \(\hat{\tau}_{D3} = \hat{\tau}_{D2} + (1/n^2)(-2\hat{\omega}^2 + 2\hat{\omega})\), the delta estimator of order two; \(\hat{\tau}_{D4} = \hat{\tau}_{D3} + (1/n^3)\hat{\omega} (\hat{\omega} - 1)(\hat{\omega}^2 - 6\hat{\omega} + 12)\), the delta estimator of order three. Here, \(\hat{\omega} = 1/\hat{\theta}\) and \(\hat{\theta}\) is the maximum likelihood estimator of \(\theta\).
We compute the bias and mean squared error of the four estimators by means of the following simulation scheme:

1. generate ten thousand variates of Binomial \((n, \theta)\);
2. compute the four estimators for the ten thousand variates, yielding the estimates \(\hat{t}_{Ds,i}\) for \(i = 1, 2, \ldots, 10000\) and \(s = 1, 2, 3, 4\);
3. compute the bias and mean squared error given by
   \[
   \text{bias}_s(n) = \frac{1}{10000} \sum_{i=1}^{10000} (\hat{t}_{Ds,i} - 1 + 1/\theta)
   \]
   and
   \[
   \text{MSE}_s(n) = \frac{1}{10000} \sum_{i=1}^{10000} (\hat{t}_{Ds,i} - 1 + 1/\theta)^2
   \]
   for \(s = 1, 2, 3, 4\).

We repeat these steps for \(n = 1, 2, \ldots, 100\) and \(\theta = 0.3, 0.5, 0.7, 0.9\), so computing bias\(_s(n)\) and MSE\(_s(n)\) for \(s = 1, 2, 3, 4\) and \(n = 1, 2, \ldots, 100\).

Figures 1 and 2 show how the bias and mean squared error of the four estimators vary with respect to \(n\) and \(\theta\). The following observations can be made:

1. the biases generally approach zero with increasing \(n\);
2. the biases generally decrease with increasing \(\theta\);
3. the biases are generally negative for \(s = 1\);
4. the biases are generally positive for \(s = 2, 3, 4\);
5. the biases appear largest for \(s = 1\);
6. the biases appear second largest for \(s = 2\);
7. the biases appear smallest for \(s = 3, 4\);
8. the mean squared errors generally decrease to zero with increasing \(n\);
9. the mean squared errors generally decrease with increasing \(\theta\);
10. the mean squared errors appear largest for \(s = 1\);
11. the mean squared errors appear second largest for \(s = 4\);
12. the mean squared errors appear smallest for \(s = 2, 3\).

We have presented results only for \(\theta = 0.3, 0.5, 0.7, 0.9\) for reasons of space. But the results were similar for other choices for \(\theta\).

Hence, we see that the delta estimators reduce both bias and mean squared error. Some theoretical work will be required to see if this holds in general.
Acknowledgments

The authors would like to thank the Editor and the referee for careful reading and for their comments which greatly improved the paper.

References


Figure 1. Biases versus $n$ for $\theta = 0.3$ (top left), $\theta = 0.5$ (top right), $\theta = 0.7$ (bottom left) and $\theta = 0.9$ (bottom right).
Figure 2. Mean squared errors versus $n$ for $\theta = 0.3$ (top left), $\theta = 0.5$ (top right), $\theta = 0.7$ (bottom left) and $\theta = 0.9$ (bottom right).