Efficiency of the u,v method of estimation

Herbert Robbins and Cun-Hui Zhang*

Department of Statistics, Rutgers University, Piscataway, NJ 08854

Contributed by Herbert Robbins, August 24, 2000

Given a pool of motorists, how do we estimate the total intensity of those who had a prespecified number of traffic accidents in the past year? We previously have proposed the u,v method as a solution to estimation problems of this type. In this paper, we prove that the u,v method provides asymptotically efficient estimators in an important special case.

1. The u,v Method

Given a pool of motorists, how do we estimate the total intensity of those in the pool who had a prespecified number of traffic accidents in a given time period? We may also consider patients with a prespecified number of heart attacks, or salesmen with a prespecified number of disgruntled customers, etc. In general, let θ_i be the intensity and X_i the number of occurrences of certain type of events of the ith individual in a pool of size n. Suppose that for $1 \le i \le n$ conditionally on θ_i, X_i has the Poisson distribution with $E[X_i|\theta_i] = \theta_i$. We are interested in estimating the sum

$$S_n \equiv \sum_{i=1}^{n} u(X_i)\theta_i, \qquad [1.1]$$

where $u(x)$ is a known "utility function" dictated by practical considerations. In the examples above, S_n is the sum of the intensity θ_i for those individuals with $X_i = a$ traffic accidents (heart attacks, disgruntled customers, etc.) for a prespecified integer a, if

$$u(x) = u_a(x) \equiv \begin{cases} 1, & x = a \\ 0, & x \neq a. \end{cases} \qquad [1.2]$$

Robbins (1) considered estimation of the sum in **1.1** and certain other related quantities for general, but known, conditional distributions $F(x|y)$ of X_i given $\theta_i = y$. The solution he proposed, called the u,v method, estimates S_n by

$$V_n \equiv \sum_{i=1}^{n} v(X_i), \qquad [1.3]$$

if there exists a function $v(x)$ such that

$$\int \{v(x) - u(x)y\} F(dx|y) = 0, \ \forall y. \qquad [1.4]$$

In the Poisson case, Eq. **1.4** has the unique solution

$$v(x) = xu(x-1), \qquad [1.5]$$

provided that $\sum_{x=0}^{\infty} |u(x)|y^x/x! < \infty$ for all $y > 0$.

In this paper, we consider the asymptotic efficiency of the u,v method. We prove the asymptotic efficiency of **1.3** for the estimation of **1.1** in the special case of Eq. **1.2** in the Poisson setting in Section 2. In Section 3, we discuss related problems and extensions to the estimation of the sums of $u(X_i, \theta_i)$ for general utility functions $u(x, y)$ and general conditional distributions $F(x|y)$.

2. The Poisson Case

Let $f(x|y) \equiv e^{-y}y^x/x!$, $x = 0, 1, 2, \ldots$, be the Poisson probability mass function with intensity $y > 0$ and $\mathcal{G}$ be a known family of probability distributions with support $(0, \infty)$. Suppose (X, θ), (X_i, θ_i), are independent identically distributed random vectors such that

$$X|\theta \sim f(x|\theta), \quad \theta \sim G, \qquad [2.1]$$

where $G \in \mathcal{G}$ is an unknown distribution. We consider in this section estimation of

$$S_n \equiv \sum_{i=1}^{n} u_a(X_i)\theta_i, \qquad [2.2]$$

with the u_a in Eq. **1.2** for a given a. By the u,v method, **2.2** should be estimated by

$$V_n \equiv \sum_{i=1}^{n} v_a(X_i), \quad v_a(x) \equiv \begin{cases} a+1, & x = a+1 \\ 0, & x \neq a+1 \end{cases} \qquad [2.3]$$

as in **1.3** and Eq. **1.5**. For example, according to **2.3**, the total intensity of those motorists with no traffic accidents in the past year is estimated by the total number of motorists with exactly one accident in the past year.

The estimator **2.3** also can be derived from an empirical Bayes point of view. If the distribution G in **2.1** is known, then the Bayes estimator of **2.2** under the squared error loss is the conditional expectation

$$E_G[S_n|X_1, \ldots, X_n] = \sum_{i=1}^{n} u_a(X_i)E_G[\theta_i|X_i],$$

which can be written as

$$S_{n,G} \equiv \sum_{i=1}^{n} u_a(X_i)\tau_a(G) \qquad [2.4]$$

with

$$\tau_a(G) \equiv E_G[\theta_i|X_i = a] = (a+1)\frac{f_G(a+1)}{f_G(a)}, \qquad [2.5]$$

where $f_G(x) \equiv \int f(x|y)dG(y)$ is the marginal probability mass function of X. An empirical Bayes estimator of **2.2** can be obtained by substituting the conditional expectation $\tau_a(G)$ with

*To whom reprint requests should be addressed.

a suitable estimator, say $\tilde{\tau}_{a,n}$, in the Bayes estimator $S_{n,G}$ in **2.4**; i.e.

$$\tilde{S}_n \equiv \tilde{\tau}_{a,n} \sum_{i=1}^{n} u_a(X_i). \qquad \textbf{[2.6]}$$

If G is completely unknown, we may estimate $f_G(x)$ by its empirical version and consequently estimate $\tau_a(G)$ by

$$\hat{\tau}_{a,n} \equiv (a+1)\frac{\hat{f}_n(a+1)}{\hat{f}_n(a)}, \quad \hat{f}_n(a) \equiv \sum_{i=1}^{n} \frac{u_a(X_i)}{n}. \qquad \textbf{[2.7]}$$

This leads to the estimator **2.3** via

$$\sum_{i=1}^{n} u_a(X_i)\hat{\tau}_{a,n} = \sum_{i=1}^{n} u_a(X_i)(a+1)\frac{\sum_{i=1}^{n} u_{a+1}(X_i)/n}{\sum_{i=1}^{n} u_a(X_i)/n}$$

$$= \sum_{i=1}^{n} v_a(X_i).$$

The relationship **2.6** can be reversed to derive estimates of $\tau_a(G)$ from those of **2.2**, say $\tilde{S}_n \equiv \tilde{S}_n(X_1, \ldots, X_n)$; i.e.

$$\tilde{\tau}_{a,n} \equiv \frac{\tilde{S}_n}{\sum_{i=1}^{n} u_a(X_i)}. \qquad \textbf{[2.8]}$$

This provides a vehicle for the investigation of the efficiency of $\tilde{S}_n$ via the efficiency of $\tilde{\tau}_{a,n}$. Let $H_* \equiv H_{*,G}$ be the tangent space of the family $\{f_G : G \in \mathcal{G}\}$ at G,

$$H_* \equiv \text{the closure of the linear span of } \{\rho_\eta : \eta \in C_G\} \text{ in } L_2(f_G), \qquad \textbf{[2.9]}$$

where C_G is the collection of all "differentiable" paths $\eta : [0, 1] \to \mathcal{G}$ *satisfying*

$$\lim_{t \to 0+} t^{-2} \sum_{x=0}^{\infty} \left(\sqrt{f_{\eta(t)}(x)/f_G(x)} - 1 - t\rho_\eta(x)/2 \right)^2 f_G(x) = 0, \qquad \textbf{[2.10]}$$

with the f_G in Eq. **2.5**, and

$$\rho_\eta(x) \equiv \lim_{t \to 0+} t^{-1}\{\log f_{\eta(t)}(x) - \log f_G(x)\}, \quad x = 0, 1, \ldots \qquad \textbf{[2.11]}$$

is the score function for the path η in the parameter space $\mathcal{G}$. See Bickel *et al.* (2). Define

$$\psi \equiv \psi(x; G) \equiv \tau_a(G)\left\{\frac{u_{a+1}(x)}{f_G(a+1)} - \frac{u_a(x)}{f_G(a)}\right\} \qquad \textbf{[2.12]}$$

with the u_a in Eq. **1.2**. It will be shown in the proof of THEOREM 2.1 that at each $G \in \mathcal{G}$ the efficient influence function for the estimation of $\tau_a(\cdot)$ is

$$\psi_* \equiv \psi_*(x; G) \equiv \text{the projection of } \psi \text{ on to } H_*, \qquad \textbf{[2.13]}$$

where H_* is the tangent space given in **2.9**.

THEOREM 2.1 *(i) A sequence* $\{\tilde{S}_n \equiv \tilde{S}_n(X_1, \ldots, X_n)\}$ *is asymptotically efficient for the estimation of the* $\{S_n\}$ *in* **2.2** *if and only if* $\{\tilde{\tau}_{a,n}\}$ *in* **2.8** *is asymptotically efficient for the estimation of the functional* $\tau_a(G)$ *in* **2.5**. *In this case,*

$$(\tilde{S}_n - S_n)/\sqrt{n} \xrightarrow{D} N(0, f_G^2(a)\sigma_1^2(G) + \sigma_2^2(G)), \qquad \textbf{[2.14]}$$

where $\sigma_1^2(G) \equiv E_G\psi_*^2(X; G)$ *with the* ψ_* *in* **2.13**, $\sigma_2^2(G) \equiv E_G u_a(X) Var_G(\theta|X)$ *with the* u_a *in Eq.* **1.2**, *and* f_G *is the marginal probability mass function of* X. *(ii) If* G *is completely unknown, i.e.,* $\mathcal{G} = \{$*all distributions in* $(0, \infty)\}$, *then* $\{V_n\}$ *in* **2.3** *is asymptotically efficient for the estimation of* **2.2** *and*

$$f_G^2(a)\sigma_1^2(G) + \sigma_2^2(G) = E_G\{v_a(X) - u_a(X)\theta\}^2. \qquad \textbf{[2.15]}$$

Proof: The proof has three parts.

Step 1. Decomposition of $(\tilde{S}_n - S_n)/\sqrt{n}$: By **2.8** and **2.4**

$$\{\tilde{S}_n - S_n\}/\sqrt{n} = \hat{f}_n(a)\xi_{n,1} + \xi_{n,2}, \qquad \textbf{[2.16]}$$

where $\hat{f}_n(a)$ is as in **2.7**, $\xi_{n,1} \equiv \sqrt{n}\{\tilde{\tau}_{a,n} - \tau_a(G)\}$ and $\xi_{n,2} \equiv \{S_{n,G} - S_n\}/\sqrt{n}$. Conditionally on $\{X_i, i \geq 1\}$, $S_{n,G} - S_n$ are sums of independent (not identically distributed) random variables with mean zero, so that by the Lindeberg central limit theorem and the law of large numbers

$$\xi_{n,2} = \sum_{i=1}^{n} \frac{u_a(X_i)(\theta_i - E_G[\theta_i|X_i])}{\sqrt{n}}$$

$$\overset{D}{\approx} N\left(0, \sum_{i=1}^{n} u_a^2(X_i)\mathrm{Var}_G(\theta_i|X_i)/n\right) \xrightarrow{D} N(0, \sigma_2^2(G)) \qquad \textbf{[2.17]}$$

almost surely for all $\{X_i, i \geq 1\}$. The Lindeberg condition can be verified by the law of large numbers, but we shall omit the details. Because the limiting distribution in Eq. **2.17** does not depend on $\{X_i, i \geq 1\}$ and $\hat{f}_n(a) \to f_G(a)$, by Eq. **2.16**

$$\mathcal{L}((\tilde{S}_n - S_n)/\sqrt{n}; P_G) - \mathcal{L}(f_G(a)\xi_{1,n}; P_G) \star N(0, \sigma_2^2(G)) \to 0, \qquad \textbf{[2.18]}$$

provided that either $(\tilde{S}_n - S_n)/\sqrt{n}$ or $\xi_{1,n} \equiv \sqrt{n}\{\tilde{\tau}_{a,n} - \tau_a(G)\}$ are stochastically bounded, where $\mathcal{L}(Z; P)$ is the distribution of Z under probability P and $\star$ stands for convolution. Thus, $\{\tilde{S}_n\}$ is asymptotically efficient for the estimation of S_n if and only if $\{\tilde{\tau}_{a,n}\}$ is asymptotically efficient for the estimation of $\tau_a(G)$.

Step 2. Efficient influence function for the estimation of $\tau_a(G)$: It follows from the information bound in standard semiparametric estimation theory that the limiting distribution of asymptotically efficient $\{\tilde{\tau}_{a,n}\}$ is

$$\sqrt{n}\{\tilde{\tau}_{a,n} - \tau_a(G)\} \to N(0, E_G\psi_*^2(X; G)), \qquad \textbf{[2.19]}$$

provided that ψ_* is the efficient influence function for the estimation of $\tau_a(G)$. By **2.13**, this is the case if for all $\eta \in C_G$

$$\lim_{t \to 0+} t^{-1}\{\tau_a(\eta(t)) - \tau_a(G)\} = E_G\psi(X; G)\rho_\eta(X), \qquad \textbf{[2.20]}$$

where ρ_η is as in Eq. **2.11**. See ref. 2. Thus, it suffices to verify Eq. **2.20** for the proof of THEOREM 2.1 part *i*.

Because $f_G(x) > 0$ for all $x \geq 0$, by (2.11) $t^{-1}\{f_{\eta(t)}(x) - f_G(x)\} \to f_G(x)\rho_\eta(x)$, so that by Eq. **2.5** and **2.12**

STATISTICS

$$t^{-1}\{\tau_a(\eta(t)) - \tau_a(G)\} = \frac{a+1}{t}\left\{\frac{f_{\eta(t)}(a+1)}{f_{\eta(t)}(a)} - \frac{f_G(a+1)}{f_G(a)}\right\}$$
$$\to \tau_a(G)\{\rho_\eta(a+1) - \rho_\eta(a)\}$$
$$= E_G\psi(X, G)\rho_\eta(X).$$

Therefore, Eq. **2.20** holds.

Step 3. Asymptotic efficiency of the u,v method: Let ψ be as in **2.12** and $\hat{\tau}_{a,n}$ be as in **2.7**. By the central limit theorem and the strong law of large numbers, $\sqrt{n}(\hat{\tau}_{a,n} - \tau_a(G))$ converges in distribution to $N(0, E_G\psi^2(X; G))$. Because V_n is the estimator of S_n corresponding to $\hat{\tau}_{a,n}$ by **2.8**, it suffices to show $\psi = \psi_*$ in view of Theorem 2.1 part *i* and its proof.

For $y > 0$ define $\eta(t) \equiv (1-t)G + t\delta_y$, where δ_y puts the whole mass at y. Set $\rho_{(y)}(x) \equiv \{f(x|y) - f_G(x)\}/f_G(x)$. Then, $E_G\rho^2_{(y)}(X) < \infty$ by the Poisson assumption, and the left-hand side of Eq. **2.10** is

$$\sum_{x=0}^{\infty}\left[\frac{\{f_{\eta(t)}(x)/f_G(x) - 1\}/t}{\sqrt{f_{\eta(t)}(x)/f_G(x)} + 1} - \frac{\rho_{(y)}(x)}{2}\right]^2 f_G(x)$$

$$= \sum_{x=0}^{\infty}\frac{\rho^2_{(y)}(x)}{4}f_G(x)\left[\frac{\sqrt{f_{\eta(t)}(x)/f_G(x)} - 1}{\sqrt{f_{\eta(t)}(x)/f_G(x)} + 1}\right]^2 \to 0.$$

Thus, $\rho_{(y)} \equiv f(x|y)/f_G(x) - 1$ is in the tangent space H_* for all $y > 0$ by **2.9**. If h is orthogonal to H_* in $L_2(f_G)$, then

$$0 = E_G h(X)\rho_{(y)}(X) = \sum_{x=0}^{\infty} h(x)f(x|y) - E_G h(X)$$

for all $y > 0$, so that $h(x) = E_G h(X)$ for all $x \geq 0$ by the completeness of the Poisson family. This implies $H_* = L_2(f_G) \cap \{h: E_G h(X) = 0\}$. Hence, $\psi_* = \psi$ by **2.13** and the proof is complete.

3. Discussion

3.1. Related Problems. Let Y_i be random variables such that $E[Y_i|\theta_i, X_i] = \lambda\theta_i$. Suppose Y_i are unobservable and λ is known. Consider the prediction of

$$\sum_{i=1}^{n} u(X_i)Y_i \qquad \textbf{[3.1]}$$

based on observations $X_1, \ldots, X_n$. For example, we may want to predict the total number of accidents in the coming year for the group of motorists with no accidents in the past year, with $\lambda = 1.02$ due to 2% growth of drivers in the region of concern. By the u,v method, **3.1** can be predicted by λV_n if Eq. **1.4** holds, with the V_n in **1.3**. The argument in Section 2 still applies here in the Poisson case with $u(x) = u_a(x)$ in Eq. **1.2**: $\{\lambda V_n\}$, with the V_n in **2.3**, is asymptotically efficient for the prediction of **3.1** with

$$n^{-1/2}\left(\lambda V_n - \sum_{i=1}^{n} u(X_i)Y_i\right)$$
$$\xrightarrow{D} N(0, \lambda^2 f_G^2(a)\sigma_1^2(G) + E_G u_a(X)\mathrm{Var}_G(Y|X)), \qquad \textbf{[3.2]}$$

where $\sigma_1^2(G)$ is as in Theorem 2.1.

In many applications, Y_i are observable and the problem is to estimate λ. In this case, the u,v methodology provides the estimator

$$\hat{\lambda}_n \equiv V_n^{-1}\sum_{i=1}^{n} u(X_i)Y_i. \qquad \textbf{[3.3]}$$

The u,v method also produces estimates of variances. For example, if Eq. **1.4** holds, the variance $E_G(V_n - S_n)^2 = nE_G\{v(X) - u(X)\theta\}^2$ can be estimated by

$$\sum_{i=1}^{n}\{v^2(X_i) + v_2(X_i)\}, \qquad \textbf{[3.4]}$$

with two applications of the u,v method, first to $u_1 \equiv u^2$ and then to $u_2 \equiv v_1 - 2uv$, where v_j satisfy $\int (v_j(x) - u_j(x)y)F(dx|y) = 0, \forall y$.

The u,v method can further be extended to obtain unbiased estimation of

$$S_n \equiv \sum_{i=1}^{n} \theta_i u_i(X_1, \ldots, X_n). \qquad \textbf{[3.5]}$$

If there exist functions $v_i(x_1, \ldots, x_n)$ satisfying

$$y\int u_i(x_1, \ldots, x_n)F(dx_i|y) = \int v_i(x_1, \ldots, x_n)F(dx_i|y)$$

for all y, $i \leq n$ and $\{x_j, j \neq i\}$, then we may estimate **3.5** by

$$V_n \equiv \sum_{i=1}^{n} v_i(X_1, \ldots, X_n). \qquad \textbf{[3.6]}$$

For example, in the exponential case $f(x|y) \equiv y^{-1}e^{-x/y}1\{x > 0\}$, the θ_i associated with the largest observation can be written as **3.5**,

$$\theta_{R_n} = S_n \equiv \sum_{i=1}^{n}\theta_i u_i(X_1, \ldots, X_n),$$

$$u_i(x_1, \ldots, x_n) = 1\left\{x_i = \max_{1\leq j\leq n} x_j\right\},$$

and its unbiased estimation **3.6** is $V_n = X_{R_n} - X_{R_{n-1}}$ with

$$v_i(x_1, \ldots, x_n) = \int_0^{x_i} u_i(x_1, \ldots, x_{i-1}, t, x_{i+1}, \ldots, x_n)dt,$$

where R_i are the antiranks of the observations defined by $X_{R_1} < \cdots < X_{R_n}$.

The related problems mentioned here and their applications were considered in refs. 1 and 3–5.

3.2. Extensions. The applicability of our methodology is not limited to the sum of $u(X_i)\theta_i$ in **1.1**. In general, **1.3** can be used to estimate

$$S_n \equiv \sum_{i=1}^{n} u(X_i, \theta_i) \quad \textbf{[3.7]}$$

for any integrable functions $u(x, y)$, as long as

$$\int \{v(x) - u(x, y)\} F(dx|y) = 0, \quad \forall y. \quad \textbf{[3.8]}$$

In fact, for the estimation of variance in **3.4**, Eq. **3.8** holds for the pair $\tilde{u}(x, y)$ and $\tilde{v}(x)$, with $\tilde{u}(x, y) \equiv \{v(x) - u(x)y\}^2$ and $\tilde{v}(x) \equiv v^2(x) + v_2(x)$.

The asymptotic theory for the estimation of **3.7** is more complicated and will be studied elsewhere. Define

$$S_{n,G} \equiv \sum_{i=1}^{n} u_G(X_i), \ u_G(X_i) \equiv E_G[u(X_i, \theta_i)|X_i]. \quad \textbf{[3.9]}$$

The asymptotic independence of $(\tilde{S}_n - S_{n,G})/\sqrt{n}$ and $(S_{n,G} - S_n)/\sqrt{n}$ can still be derived from the Lindeberg central limit theorem and the strong law of large numbers as in Section 2, but the rest of the argument there does not directly apply without the one-to-one linear mappings between estimates of S_n and $\tau_a(G)$ in **2.6** and **2.8**.

This research was partially supported by the National Science Foundation and National Security Agency.

1. Robbins, H. (1988) in *Statistical Decision Theory and Related Topics IV*, eds. Gupta, S. S. & Berger, J. O. (Springer, New York), Vol. 1, pp. 265–270.
2. Bickel, P. J., Klaassen, C. A. J., Ritov, Y. & Wellner, J. A. (1992) *Efficient and Adaptive Estimation for Semiparametric Models* (Johns Hopkins Univ. Press, Baltimore).
3. Robbins, H. & Zhang, C.-H. (1988) *Proc. Nat. Acad. Sci. USA* **85,** 3670–3672.
4. Robbins, H. & Zhang, C.-H. (1989) *Proc. Nat. Acad. Sci. USA* **86,** 3003–3005.
5. Robbins, H. & Zhang, C.-H. (1991) *Biometrika* **78,** 349–354.

STATISTICS