WINPEC Working Paper Series No.E2024
March 2021

A Necessary Condition for Semiparametric Efficiency of Experimental Designs

Hisatoshi Tanaka

Waseda INstitute of Political EConomy
Waseda University
Tokyo, Japan

A Necessary Condition for Semiparametric Efficiency of Experimental Designs

Hisatoshi Tanaka

School of Political Science and Economics, Waseda University
Shinjuku, Tokyo 169-8050, Japan
hstnk@waseda.jp

Abstract. Efficiency of estimation depends not only on a method of the estimation, but also on the distribution of data. In statistical experiments, statisticians can at least partially design the data generating process to obtain high performance of the estimation. In this paper, a necessary condition for the semiparametrically efficient experimental design is proposed. A formula to determine the efficient distribution of input variables is derived. An application to the optimal bid design problem of contingent valuation survey experiments is presented.

Keywords: Optimal Design · Semiparametric Efficiency · Binary Response Model · Contingent Valuation Survey Experiments

1 Introduction

In this paper a class of simple statistical experiments described by a 4-tuple,

$$\mathcal{E} = \{(\mu, \nu, \rho, \varphi) : \mu \in \mathcal{M}, \nu \in \mathcal{N}\}, \tag{1}$$

is investigated, where $\mathcal{M}$ is a set of probability measures on $(\mathcal{W}, \mathcal{A})$, $\mathcal{N}$ is a set of probability measures on $(\mathcal{X}, \mathcal{B})$, ρ is a measurable map from $\mathcal{W} \times \mathcal{X}$ to $(\mathcal{Y}, \mathcal{C})$, and φ is a functional on $\mathcal{M}$. In every experiment $(\mu, \nu, \rho, \varphi) \in \mathcal{E}$, input x is drawn from ν, output $y = \rho(\omega, x)$ with $\omega \sim \mu$ is observed, and the value of $\varphi(\mu)$ is estimated from (x, y).

For example, imagine that there exist n lightning bulbs, whose life time hours $\omega_1, \ldots, \omega_n$ are i.i.d. random variables distributed according to μ. In order to estimate the expected life time hours $\varphi(\mu) = E\omega$, the following experiment is conducted. First, all n bulbs are turned on at time 0. Second, one of the bulbs is sampled without replacement at time x and its status is observed. If the sampled bulb is alive, y is set 1. If otherwise, y is set 0. The procedure is repeated n times until all of the bulbs are sampled. Finally, data of n independent pairs (x_1, y_1), $\cdots$, (x_n, y_n) are obtained, and $E\omega$ or any other moments of ω will be consistently estimated by existing efficient estimation methods, such as the nonparametric maximum likelihood estimation.

To be noted here is that efficiency of the estimation depends not only on the estimation method, but also on the distribution ν of $x_1, \cdots, x_n$. In an extreme case where $x_1 = \ldots = x_n = 0$, trivial outcomes $y_1 = \ldots = y_n = 1$ will be obtained unless some of the bulbs are with initial failure. In the opposite extreme case where $x_1 = \ldots = x_n = +\infty$, $y_1 = \ldots = y_n = 0$ will occur with probability one. In both cases, data are so poorly informative that consistent estimation of $E\omega$ is not possible. To find the best distribution ν of x, with which the experiment produces the most informative data, is therefore an interesting problem.

The paper is organized as follows. In Section 2, the problem of the paper is formally stated. For the purpose, geometric theory of semiparametric estimation is introduced. In the theory, every statistical model is considered as a point on an infinite dimensional manifold, and the efficient design is formulated as a minimizer of the Fisher-information norm of the gradient of a functional on the manifold. In Section 3, a necessary condition for the efficient design is proposed. In Section 4, application examples of the main theorem are given. In particular, the optimal bid design problem of contingent valuation survey experiments is solved. In Section 5, results from small Monte Carlo simulations are reported. It is numerically confirmed that the efficiently designed estimations outperform opponents even with small samples.

2 The Model

2.1 The tangent space of a statistical manifold

In this section, geometric theory of semiparametric estimation is introduced to formulate the efficient design problem. Terms and definitions given in the following are according to [12]. Equivalent definitions are also found in [1], [2], [3], and [11].

Let μ be a probability measure on $(\mathcal{W}, \mathcal{A})$. Let $\mathcal{M}$ be a set of probability measures, which are absolutely continuous with respect to μ. A map $t \mapsto \mu_t$ from $(-\epsilon, \epsilon) \subset \mathbb{R}$ to $\mathcal{M}$ such that $\mu_0 = \mu$ is *differentiable in quadratic mean* at $t = 0$ if there exists $\alpha \in L_2(\mu)$ such that

$$\lim_{t \to 0} \int \left(\frac{\sqrt{d\mu_t} - \sqrt{d\mu}}{t} - \frac{1}{2}\alpha\sqrt{d\mu} \right)^2 = 0. \tag{1}$$

Proposition 1. *A map $t \mapsto \mu_t$ is differentiable in quadratic mean at $t = 0$ if (i) a map $t \mapsto \ell_t(\omega) := d\mu_t/d\mu(\omega)$ is continuously differentiable on $(-\epsilon, \epsilon)$ and if (ii) a map $t \mapsto \int (\dot{\ell}_t/\ell_t)^2 d\mu_t$ becomes continuous on $(-\epsilon, \epsilon)$, where $\dot{\ell}_t(\omega) = (d\ell_t/dt)(\omega)$. Under conditions (i) and (ii), $\dot{\ell}_0$ becomes a tangent vector of $\mathcal{M}$ at μ.*

Proof. See *e.g.* Proposition 1 in page 13 of [3]. □

A collection of those differentiable maps $t \mapsto \mu_t$ is denoted by $\mathcal{M}(\mu)$. A *tangent space* $T_\mu\mathcal{M}$ of $\mathcal{M}$ at μ is a set of *tangent vectors* α as in (1). A *tangent*

bundle $T\mathcal{M}$ relates each μ with $T_\mu\mathcal{M}$. A pair $(\mathcal{M}, T\mathcal{M})$ is a *statistical manifold*, which is an infinite dimensional analog of a standard finite-dimensional manifold.

On $(\mathcal{M}, T\mathcal{M})$, the *Fisher-information metric* $\mu \mapsto \langle \cdot, \cdot \rangle_\mu^{1/2}$ is defined by

$$\langle \alpha, \alpha' \rangle_\mu = \int_{\mathcal{W}} \alpha\alpha' \, d\mu \tag{2}$$

for every α and α' in $T_\mu\mathcal{M}$. The *Fisher-information norm* $\| \cdot \|_\mu$ is also given by $\|\alpha\|_\mu = \langle \alpha, \alpha \rangle_\mu^{1/2}$. The following proposition characterizes $T_\mu\mathcal{M}$.

Proposition 2 ([10], [13]). *Let $\overline{T_\mu\mathcal{P}(\mathcal{W})}$ be the closure of a tangent space $T_\mu\mathcal{P}(\mathcal{W})$ with respect to $\| \cdot \|_\mu$, then*

$$\overline{T_\mu\mathcal{P}(\mathcal{W})} = L_2^0(\mu) := \left\{ \alpha \in L_2(\mu) \;\middle|\; \int \alpha \, d\mu = 0 \right\}. \tag{3}$$

Proof. Choose an arbitrary $\alpha \in L_2^0(\mu)$ and $M > 0$. Let $\alpha_M^0 = \alpha_M - \int \alpha_M \, d\mu$, where $\alpha_M = \alpha \cdot \{|\alpha| \le M\}$. Define a map $t \mapsto \mu_t$ by

$$\ell_t = \frac{d\mu_t}{d\mu} = \exp\left(t\alpha_M^0 - \gamma_t\right), \quad \gamma_t = \log\left(\int \exp(t\alpha_M^0) \, d\mu\right). \tag{4}$$

Since $|\alpha_M^0| \le M$, (4) is well-defined and $t \mapsto \ell_t(\omega)$ becomes continuously differentiable with derivative

$$\dot{\ell}_t(\omega) = \left(\frac{d}{dt}\right) \frac{d\mu_t}{d\mu}(\omega) = \left(\alpha_M^0(\omega) - \frac{\int \alpha_M^0 \exp(t\alpha_M^0) \, d\mu}{\int \exp(t\alpha_M^0) \, d\mu}\right) \exp\left(t\alpha_M^0(\omega) - \gamma_t\right) \tag{5}$$

at every $\omega \in \mathcal{W}$. A map $t \mapsto \int (\dot{\ell}_t/\ell_t)^2 d\mu_t$ is also well-defined and continuous in $t \in (-\epsilon, \epsilon)$, hence $t \mapsto \mu_t$ is differentiable in quadratic mean at $t = 0$ with derivative $\dot{\ell}_0 = \alpha_M^0$. Let $M \uparrow \infty$, then $\|\alpha - \alpha_M^0\|_\mu \to 0$. Thus, $\alpha \in \overline{T_\mu\mathcal{M}}$ is shown.

On the other hand, for every $(\mu_t)_{t \in (-\epsilon, \epsilon)} \in \mathcal{M}(\mu)$ and $\alpha \in L_2(\mu)$, let $\xi_k = k(\sqrt{d\mu_{1/k}} - \sqrt{d\mu}) - (\alpha/2)\sqrt{d\mu}$ for $k \in \mathbb{N}$. Then, as $k \to \infty$, $\int \xi_k^2 \to 0$ and

$$\begin{aligned}
\left| \int \alpha \, d\mu \right| &= \left| \int \left(\frac{\alpha}{2}\sqrt{d\mu}\right) \left(2\sqrt{d\mu}\right) \right| \\
&\le \left| \int \xi_k^2 \right|^{1/2} \cdot 2 \left| \int d\mu \right|^{1/2} + \frac{1}{k} \left| \int \left(\xi_k + \frac{\alpha}{2}\sqrt{d\mu}\right)^2 \right| \\
&\le o(1) + \frac{1}{k}\left(o(1) + \frac{1}{2}\|\alpha\|_\mu\right)^2 \to 0,
\end{aligned}$$

which implies $T_\mu\mathcal{M} \subset L_2^0(\mu)$. □

2.2 The score operator

Let $\mathcal{N}$ be a class of probability measures on $(\mathcal{X}, \mathcal{B})$, and let $\mathcal{P}$ be a class of probability measures on $(\mathcal{X} \times \mathcal{Y}, \sigma(\mathcal{B} \times \mathcal{C}))$. For every $P \in \mathcal{P}$, let $\mathcal{P}(P)$ be a collection of differentiable maps $t \in (-\epsilon, \epsilon) \mapsto P_t \in \mathcal{P}$ such that $P_0 = P$. Let $T_P\mathcal{P}$ be the tangent space of $\mathcal{P}$ at P. The tangent bundle $T\mathcal{P}$ relates each P with $T_P\mathcal{P}$. The Fisher-information norm on $(\mathcal{P}, T\mathcal{P})$ is $\|\cdot\|_P$ such that $\|\beta\|_P = \left(\int \beta(x,y)\, dP(x,y)\right)^{1/2}$ for every $\beta \in T_P\mathcal{P}$. The closure of $T_P\mathcal{P}$ with respect to $\|\cdot\|_P$ is $L_2^0(P)$ as shown in Proposition 2.

Given a measurable map $\rho : \mathcal{W} \times \mathcal{X} \mapsto \mathcal{Y}$, at every $\nu \in \mathcal{N}$, a map $\rho_\nu : \mathcal{M} \mapsto \mathcal{P}$ defined by

$$\rho_\nu(\mu)(D) = \int \{(x, \rho(\omega, x)) \in D\}\, \mu(d\omega)\nu(dx), \quad D \in \sigma(\mathcal{B} \times \mathcal{C}), \tag{6}$$

is a differentiable map between $(\mathcal{M}, T\mathcal{M})$ and $(\mathcal{P}, T\mathcal{P})$. To see this, note that

$$\frac{d\rho_\nu(\mu_t)}{d\rho_\nu(\mu)}(x,y) = E_{\mu,\nu}\left(\frac{d\mu_t}{d\mu}(\omega)\,\middle|\, x, y\right) \tag{7}$$

because

$$\begin{aligned}\rho_\nu(\mu_t)(D) &= \int \frac{d\mu_t}{d\mu}(\omega)\{(x, \rho(\omega, x)) \in D\}\, \mu(d\omega)\nu(dx) \\ &= E_{\mu,\nu}\left[E_{\mu,\nu}\left(\frac{d\mu_t}{d\mu}(\omega)\,\middle|\, x, y\right)\{(x,y) \in D\}\right].\end{aligned}$$

Particularly when $d\mu_t/d\mu = \exp(t\alpha - \gamma_t)$, where $\alpha \in L_2^0(\mu)$ is bounded and $\gamma_t = \log \int \exp(t\alpha)\, d\mu$, $t \mapsto \ell_t^\rho(x,y) := d\rho_\nu(\mu_t)/d\rho_\nu(\mu)(x,y)$ is continuously differentiable with derivative

$$\begin{aligned}\dot{\ell}_t^\rho(x,y) &:= \left(\frac{d}{dt}\right)\frac{d\rho_\nu(\mu_t)}{d\rho_\nu(\mu)}(x,y) \\ &= E_{\mu,\nu}\left[\left(\alpha - \frac{\int \alpha \exp(t\alpha)\, d\mu}{\int \exp(t\alpha)\, d\mu}\right)\exp\left(t\alpha - \gamma_t\right)\,\middle|\, x, y\right].\end{aligned} \tag{8}$$

Since $t \mapsto \int (\dot{\ell}_t^\rho / \ell_t^\rho)^2\, d\rho_\nu(\mu_t)$ is continuous, $t \mapsto \rho_\nu(\mu_t)$ is a differentiable path on $\mathcal{P}$ with a tangent vector $\dot{\ell}_0^\rho(x,y) = E_{\mu,\nu}(\alpha|x,y)$.

The derivative of $\rho_\nu : \mathcal{M} \mapsto \mathcal{P}$ at μ is the *score operator* $(d\rho_\nu)_\mu : T_\mu\mathcal{M} \mapsto L_2(\rho_\nu(\mu))$, which maps every $\alpha \in T_\mu\mathcal{M}$ to

$$((d\rho_\nu)_\mu\alpha)(x,y) = E_{\mu,\nu}(\alpha|x,y), \quad (x,y) \in \mathcal{X} \times \mathcal{Y}. \tag{9}$$

Then, a tangent space of a submanifold $\rho_\nu(\mathcal{M}) := \{\rho_\nu(\mu) \in \mathcal{P} \,|\, \mu \in \mathcal{M}\}$, which is a set of statistical models to be estimated in experiment $(\mu, \nu, \rho, \varphi)$, is the range of the score operator: that is,

$$T_{\rho_\nu(\mu)}\rho_\nu(\mathcal{M}) = (d\rho_\nu)_\mu(T_\mu\mathcal{M}) = R((d\rho_\nu)_\mu), \tag{10}$$

where $R(\cdot)$ denotes the range of given operators. Note that the score operator is linear and continuous under the Fisher-information metrics. The conjugate operator $(d\rho_\nu)^*_\mu : L_2(\rho_\nu(\mu)) \mapsto L_2(\mu)$, which satisfies $\langle (d\rho_\nu)_\mu \alpha, \beta \rangle_{\rho_\nu(\mu)} = \left\langle \alpha, (d\rho_\nu)^*_\mu \beta \right\rangle_\mu$ for every $\alpha \in L_2(\mu)$ and $\beta \in L_2(\rho_\nu(\mu))$, is given by

$$((d\rho_\nu)^*_\mu \beta)(\omega) = E_{\mu,\nu}(\beta|\omega), \quad \omega \in \mathcal{W}. \tag{11}$$

2.3 The efficiency bound

Assume that $\varphi : \mathcal{M} \mapsto \mathbb{R}$ is a pathwise differentiable functional: that is,

(A1) there exists a linear, continuous operator $\varphi'_\mu : T_\mu \mathcal{M} \mapsto \mathbb{R}$ such that

$$\lim_{t \to 0} \frac{\varphi(\mu_t) - \varphi(\mu)}{t} = \varphi'_\mu \alpha \tag{12}$$

for every $(t \mapsto \mu_t) \in \mathcal{M}(\mu)$ with a tangent vector α.

By Riesz's representation theorem, there uniquely exists the *gradient function* $\partial\varphi_\mu \in L_2^0(\mu)$ such that $\varphi'_\mu \alpha \equiv \left\langle \partial\varphi_\mu, \alpha \right\rangle_\mu$. Assume also that $\varphi(\mu)$ is *identified* in the following sense:

(A2) there exists a functional $\kappa : \mathcal{P} \mapsto \mathbb{R}$ such that $\kappa(\rho_\nu(\mu)) \equiv \varphi(\mu)$ for all $\mu \in \mathcal{M}$.

The functional κ is said *differentiable at $P \in \mathcal{P}$ relative to $\mathcal{P}(P)$*, if there exists a linear, continuous operator $\kappa'_P : T_P \mathcal{P} \mapsto \mathbb{R}$ such that

$$\lim_{t \to 0} \frac{\kappa(P_t) - \kappa(P)}{t} = \kappa'_P \beta \tag{13}$$

for every differentiable path $(t \mapsto P_t) \in \mathcal{P}(P)$ with tangent vector $\beta \in L_2^0(P)$. The *efficient influence function* $\partial\kappa_P$ of κ is the Riesz representation of κ'_P on $\overline{T_P \mathcal{P}} = L_2^0(P)$, that is, $\partial\kappa_P \in L_2^0(P)$ and

$$\kappa'_P \beta \equiv \langle \partial\kappa_P, \beta \rangle_P . \tag{14}$$

Differentiability of κ at $P = \rho_\nu(\mu)$ relative to $\rho_\nu(\mathcal{M}(\mu)) := \{\rho_\nu(\mu_t) \,|\, \mu_t \in \mathcal{M}(\mu)\}$ is a necessary condition for existence of a regular estimator T_n of $\varphi(\mu)$ such that

$$\sqrt{n}(T_n - \varphi(\mu_{h_n/\sqrt{n}})) \quad \Rightarrow_{\rho_\nu(\mu_{h_n/\sqrt{n}})} \quad {}^{\exists}\mathcal{L} \tag{15}$$

for every $\mu_t \in \mathcal{M}(\mu)$ and $h_n \to h \in \mathbb{R}$. By van der Vaart's Differentiability Theorem (Theorem 3.1 of [12]), κ is differentiable at $\rho_\nu(\mu)$ relative to $\rho_\nu(\mathcal{M}(\mu))$ if and only if

$$\partial\varphi_\mu \in R((d\rho_\nu)^*_\mu). \tag{16}$$

When (16) is satisfied, the efficient influence function $\partial\kappa_{\rho_\nu(\mu)}$ is related to the gradient of φ by the *score equation*,

$$\partial\varphi_\mu = (d\rho_\nu)^*_\mu(\partial\kappa_{\rho_\nu(\mu)}), \quad \partial\kappa_{\rho_\nu(\mu)} \in \overline{R((d\rho_\nu)_\mu)}. \tag{17}$$

The limit distribution $\mathcal{L}$ of T_n is the convolution of $N\big(0, \|\partial\kappa_{\rho_\nu(\mu)}\|^2_{\rho_\nu(\mu)}\big)$ and some other probability measures on $\mathbb{R}$ ([3], [12], [13]). In this sense, $\|\partial\kappa_{\rho_\nu(\mu)}\|^2_{\rho_\nu(\mu)}$ gives the lower bound of asymptotic variances of all regular estimators of $\varphi(\mu)$. Let $\mathcal{N}^*$ be a subclass of $\mathcal{N}$ such that

$$\mathcal{N}^* = \{\nu \in \mathcal{N} \,|\, \partial\varphi_\mu \in R((d\rho_\nu)^*_\mu)\}, \tag{18}$$

then an efficiency criterion of experimental designs is given by

$$\text{l.b.}(\varphi(\mu)|\nu) := \begin{cases} \|\partial\kappa_{\rho_\nu(\mu)}\|^2_{\rho_\nu(\mu)} & \text{if } \nu \in \mathcal{N}^* \\ +\infty & \text{if } \nu \notin \mathcal{N}^*. \end{cases} \tag{19}$$

Definition 1. *The probability measure ν^* is efficient for experiment $\mathcal{E}$ at μ if*

$$\text{l.b.}(\varphi(\mu)|\nu^*) \leq \text{l.b.}(\varphi(\mu)|\nu) \tag{20}$$

for every $\nu \in \mathcal{N}$.

3 Main Results

A main theorem of the paper is given as follows.

Theorem 1. *If $\nu^* \in \mathcal{N}^*$ is efficient for $\mathcal{E}$ at μ, then $E_{\mu,\nu^*}(\partial\kappa^2_{\rho_{\nu^*}(\mu)} \,|\, x)$ is ν^*-a.s. constant on $\mathcal{X}$.*

Intuition behind the condition is obtained from the expression,

$$\text{l.b.}(\varphi(\mu)|\nu^*) = \int E_{\mu,\nu^*}(\partial\kappa^2_{\rho_{\nu^*}(\mu)} \,|\, x)\, \nu^*(dx). \tag{1}$$

If the lower bound is minimized at $\nu = \nu^*$, any small perturbations added to ν^* would not significantly change the value of l.b.$(\varphi(\mu)|\nu^*)$. This is possible only if the integrand $E_{\mu,\nu^*}(\partial\kappa^2_{\rho_{\nu^*}(\mu)} \,|\, x)$ of (1) is independent of x.

A proof of the theorem consists of the following lemmas 3.2-3.4.

Lemma 1. *For every $\nu \in \mathcal{N}^*$, $E_\nu\big[\, E_{\mu,\nu}(\partial\kappa_{\rho_\nu(\mu)}(x,y)^2|x)\,\big]^{1/2} < \infty$.*

(Proof) Let $\nu \in \mathcal{N}^*$. There exists $\beta^* \in L_2(\rho_\nu(\mu))$ such that $\partial\varphi_\mu = (d\rho_\nu)^*_\mu\beta^*$. Let Π_ν be the orthogonal projection from $L_2(\rho_\nu(\mu))$ to $\overline{R((d\rho_\nu)_\mu)}$, then

$$\langle \beta^* - \Pi_\nu\beta^*, (d\rho_\nu)_\mu\alpha\rangle_{\rho_\nu(\mu)} = \big\langle (d\rho_\nu)^*_\mu(\beta^* - \Pi_\nu\beta^*), \alpha\big\rangle_\mu = 0 \tag{2}$$

for every $\alpha \in L_2(\mu)$. Hence, there exists $\delta_\nu \in Ker((d\rho_\nu)^*_\mu)$ such that $\Pi_\nu\beta^* = \beta^* - \delta_\nu$ and that

$$(d\rho_\nu)^*_\mu\Pi_\nu\beta^* = (d\rho_\nu)^*_\mu\beta^* = \partial\varphi_\mu, \quad \Pi_\nu\beta^* \in \overline{R((d\rho_\nu)_\mu)}, \tag{3}$$

which shows $\partial\kappa_{\rho_\nu(\mu)} = \Pi_\nu\beta^*$ is the efficient influence function of κ at $\rho_\nu(\mu)$. By Jensen's inequality and the property of projections,

$$\int \left[E_{\mu,\nu}(\partial\kappa^2_{\rho_\nu(\mu)}|x)\right]^{1/2} \nu(dx) \leq \left(\int E_{\mu,\nu}(\partial\kappa^2_{\rho_\nu(\mu)}|x)\,\nu(dx)\right)^{1/2}$$
$$= \|\Pi_\nu\beta^*\|_{\rho_\nu(\mu)} \leq \|\beta^*\|_{\rho_\nu(\mu)} < \infty.$$

□

Lemma 2. *Define a map $\Gamma : \mathcal{N}^* \mapsto \mathcal{N}$ by*

$$\frac{d(\Gamma\nu)}{d\nu}(x) = C^{-1}\left[E_{\mu,\nu}(\partial\kappa^2_{\rho_\nu(\mu)}|x)\right]^{1/2}$$
$$= C^{-1}\left[\int \partial\kappa_{\rho_\nu(\mu)}(x,\rho(\omega,x))^2\,\mu(d\omega)\right]^{1/2}, \tag{4}$$

where

$$C = E_\nu\left[E_{\mu,\nu}(\partial\kappa^2_{\rho_\nu(\mu)}|x)\right]^{1/2}$$
$$= \int\left[\int \partial\kappa_{\rho_\nu(\mu)}(x,\rho(\omega,x))^2\,\mu(d\omega)\right]^{1/2}\nu(dx). \tag{5}$$

Then, $\Gamma\nu \in \mathcal{N}^*$ for every $\nu \in \mathcal{N}^*$.

Proof. For the simplicity of description, let $\nu' = \Gamma\nu$. For every $\nu \in \mathcal{N}^*$, there exists $\beta^* \in L_2(\rho_\nu(\mu))$ such that $\partial\kappa_{\rho_\nu(\mu)} = \Pi_\nu\beta^*$. Define $B_0 \subset \mathcal{X}$ by

$$B_0 = \left\{x \in \mathcal{X} \,:\, E_{\mu,\nu}(\partial\kappa_{\rho_\nu(\mu)}(x,y)^2|x) = 0\right\}. \tag{6}$$

Let $\nu_0(B) := \nu(B\backslash B_0)$ and $\nu^\perp(B) := \nu(B\cap B_0)$ for every $B \in \mathcal{B}$, so that $\nu' \sim \nu_0$, $\nu' \perp \nu^\perp$, $\nu = \nu_0+\nu^\perp$, and $(d\nu'/d\nu)(x) = (d\nu'/d\nu_0)(x) = C^{-1}\left[E_{\mu,\nu}(\partial\kappa^2_{\rho_\nu(\mu)}|x)\right]^{1/2}$. Note that $\partial\kappa_{\rho_\nu(\mu)}(x,\rho(\omega,x))\{x \in B_0\} \equiv 0$ μ-a.s. because $E_{\mu,\nu}(\partial\kappa^2_{\rho_\nu(\mu)}|x)\{x \in B_0\} \equiv 0$ and that $(d\nu_0/d\nu')\partial\kappa_{\rho_\nu(\mu)} \in L_2(\rho_{\nu'}(\mu))$ because

$$\left\|\frac{d\nu_0}{d\nu'}\partial\kappa_{\rho_\nu(\mu)}\right\|^2_{\rho_{\nu'}(\mu)} = \int\left(\frac{d\nu_0}{d\nu'}(x)\right)^2\partial\kappa_{\rho_\nu(\mu)}(x,\rho(\omega,x))^2\,\mu(d\omega)\nu'(dx)$$
$$= \int\left(\frac{d\nu_0}{d\nu'}(x)\right)\partial\kappa_{\rho_\nu(\mu)}(x,\rho(\omega,x))^2\,\mu(d\omega)\nu_0(dx)$$
$$= C\int\frac{\partial\kappa_{\rho_\nu(\mu)}(x,\rho(\omega,x))^2}{\left[E_{\mu,\nu}(\partial\kappa^2_{\rho_\nu(\mu)}|x)\right]^{1/2}}\,\mu(d\omega)\,(\nu-\nu^\perp)(dx)$$
$$= C^2 < \infty.$$

Let $\Pi_{\nu'} : L_2(\rho_{\nu'}(\mu)) \mapsto \overline{R((d\rho_{\nu'})_\mu)}$ be the orthogonal projection, then

$$(d\rho_{\nu'})^*_\mu\left(\Pi_{\nu'}\frac{d\nu_0}{d\nu'}\partial\kappa_{\rho_\nu(\mu)}\right)(\omega) = \int\left(\frac{d\nu_0}{d\nu'}(x)\right)\partial\kappa_{\rho_\nu(\mu)}(x,\rho(\omega,x))\,\nu'(dx)$$

$$
\begin{aligned}
&= \int \partial\kappa_{\rho_\nu(\mu)}(x, \rho(\omega, x))\,(\nu - \nu^\perp)(dx) \\
&= \left(\int_{\mathcal{X}} - \int_{B_0}\right) \partial\kappa_{\rho_\nu(\mu)}(x, \rho(\omega, x))\,\nu(dx) \\
&= \left((d\rho_\nu)^*_\mu\, \partial\kappa_{\rho_\nu(\mu)}\right)(\omega) = \partial\varphi_\mu(\omega).
\end{aligned}
$$

Therefore, $\partial\varphi_\mu \in R((d\rho_{\nu'})^*_\mu)$ holds at $\nu' = \Gamma\nu$. □

Lemma 3. *For every $\nu \in \mathcal{N}^*$,* $\mathrm{l.b.}(\varphi(\mu)\,|\,\Gamma\nu) \le \mathrm{l.b.}(\varphi(\mu)\,|\,\nu)$.

Proof. Let $\nu' = \Gamma\nu$. Since the efficient influence function of κ at $\rho_{\nu'}(\mu)$ is given by $\partial\kappa_{\rho_{\nu'}(\mu)} = \Pi_{\nu'}(d\nu_0/d\nu')\partial\kappa_{\rho_\nu(\mu)}$,

$$
\mathrm{l.b.}(\varphi(\mu)|\nu') \le \left\| \left(\frac{d\nu_0}{d\nu'}\right) \partial\kappa_{\rho_\nu(\mu)} \right\|^2_{\rho_{\nu'}(\mu)} = C^2 \le \left\|\partial\kappa_{\rho_\nu(\mu)}\right\|^2_{\rho_\nu(\mu)} = \mathrm{l.b.}(\varphi(\mu)|\nu). \tag{7}
$$

□

(Proof of Theorem 1) Assume that $\nu^* \in \mathcal{N}^*$ is efficient. Let $\zeta^* = E_{\mu,\nu^*}(\partial\kappa^2_{\rho_{\nu^*}(\mu)}|x)$, then

$$
\mathrm{l.b.}(\varphi(\mu)|\Gamma\nu^*) \le (E_{\nu^*}\sqrt{\zeta^*})^2 \le E_{\nu^*}\zeta^* = \mathrm{l.b.}(\varphi(\mu)|\nu^*) \le \mathrm{l.b.}(\varphi(\mu)|\Gamma\nu^*) \tag{8}
$$

by the previous lemma. Therefore, $E_{\nu^*}\zeta^* = (E_{\nu^*}\sqrt{\zeta^*})^2$, which implies $\mathrm{Var}_{\nu^*}\zeta^* = 0$. □

4 Examples

4.1 A model without information loss

Consider an experiment $\mathcal{E}$ with $\mathcal{W} = \mathcal{X} = \mathcal{Y} = \mathbb{R}$ and $\rho(\omega, x) = \omega + x$. Let $\mathcal{M}$ be a set of probability measures on $\mathbb{R}$, and let $\mathcal{N} \subset \mathcal{M}$. The model set is $\rho_\nu(\mathcal{M}) = \{\rho_\nu(\mu)|\mu \in \mathcal{M}\}$, where

$$
\rho_\nu(\mu)(D) = \int \{(x, x+\omega) \in D\}\,\mu(d\omega)\nu(dx) \tag{9}
$$

for every Borel set $D \subset \mathbb{R}^2$. The score operator $(d\rho_\nu)_\mu$ maps each $\alpha \in L^0_2(\mu)$ to $((d\rho_\nu)_\mu\alpha)(x, y) = \alpha(y - x)$. The adjoint operator is $((d\rho_\nu)^*_\mu\beta)(\omega) = \int \beta(x, \omega + x)\,\nu(dx)$. The efficient influence function of $\kappa(\rho_\nu(\mu)) = \varphi(\mu)$ is $\partial\kappa_{\rho_\nu(\mu)}(x, y) = \partial\varphi_\mu(y - x)$, because $\partial\kappa_{\rho_\nu(\mu)} \in R((d\rho_\nu)_\mu)$ and

$$
\left((d\rho_\nu)^*_\mu\partial\kappa_{\rho_\nu(\mu)}\right)(\omega) = \int \partial\varphi_\mu((\omega + x) - x)\,\nu(dx) = \partial\varphi_\mu(\omega)\int \nu(dx) = \partial\varphi_\mu(\omega). \tag{10}
$$

The efficiency bound is given by

$$
\mathrm{l.b.}(\varphi(\mu)|\nu) = \int \partial\varphi_\mu(y - x)^2\rho_\nu(\mu)(dx, dy) = \int \partial\varphi_\mu(\omega)^2\mu(d\omega), \tag{11}
$$

which is independent of ν. Therefore, arbitrary $\nu \in \mathcal{N}$ is efficient for $\mathcal{E}$. In fact, $E_{\mu,\nu}(\partial\kappa^2_{\rho_\nu(\mu)}|x) = \int \partial\varphi_\mu(\omega)^2\mu(d\omega)$ is a constant independent of $\nu \in \mathcal{N}$.

In this example, ω is always observable since $\omega = y - x$. Therefore, the distribution of x has no impact on efficiency of the estimation. The choice of ν becomes significant only when a model with information loss is estimated.

4.2 Parametric Models

Let $\mathcal{M} = \{\mu_\theta : \theta \in \Theta\}$ be a collection of statistical models parametrized by $\theta \in \Theta \subset \mathbb{R}^l$. Assume that every μ_θ is absolutely continuous with respect to a reference measure π. Let $\dot{\ell}_\theta = (\partial/\partial\theta)(d\mu_\theta/d\pi)$, then $T_{\mu_\theta}\mathcal{M} = \{\tau'\dot{\ell}_\theta : \tau \in \mathbb{R}^l\}$, and

$$\overline{R((d\rho_\nu)_{\mu_\theta})} = R((d\rho_\nu)_{\mu_\theta}) = \{\tau'(d\rho_\nu)_{\mu_\theta}(\dot{\ell}_\theta) : \tau \in \mathbb{R}^l\}. \tag{12}$$

Let $\varphi(\mu_\theta)$ be a target of estimation and let $\varphi_0(\theta) := \varphi(\mu_\theta)$, then $\kappa(\rho_\nu(\mu_\theta)) = \varphi_0(\theta)$ is differentiable if and only if there exists $\tau \in \mathbb{R}^l$ such that $(\nabla_\theta\varphi_0)'\dot{\ell}_\theta = \tau'(d\rho_\nu)^*_{\mu_\theta}(d\rho_\nu)_{\mu_\theta}(\dot{\ell}_\theta)$, or

$$(\nabla_\theta\varphi_0)'\dot{\ell}_\theta(\omega) = \tau' E_\theta\left[E_\theta(\dot{\ell}_\theta(\omega)|x,y)\,\middle|\,\omega\right] \tag{13}$$

μ_θ-almost surely. Let $\mathcal{N}^*$ be a collection of ν with which (13) is satisfied by some τ. For every $\nu \in \mathcal{N}^*$, the efficient influence function is solved as $\partial\kappa_{\rho_\nu(\mu_\theta)} = (\tau^*)'\dot{\ell}_\theta$, where

$$\tau^* = \left(E_\theta\left[E_\theta(\dot{\ell}_\theta|x,y)E_\theta(\dot{\ell}_\theta|x,y)'\right]\right)^{-1} E_\theta(\dot{\ell}_\theta\dot{\ell}'_\theta)\nabla_\theta\varphi_0.$$

The efficiency bound of $\varphi_0(\theta)$ is now given by l.b.$(\varphi_0(\theta)|\nu) = (\tau^*)'E_\theta(\dot{\ell}_\theta\dot{\ell}'_\theta)(\tau^*)$, which is minimized with respect to ν only when

$$x \mapsto E_\theta(\partial\kappa^2_{\rho_\nu(\mu_\theta)}|x) = (\tau^*)'E_\theta(\dot{\ell}_\theta\dot{\ell}'_\theta|x)(\tau^*) \tag{14}$$

is a constant map. However, since $\omega \perp\!\!\!\perp x$, $E_\theta(\dot{\ell}_\theta\dot{\ell}'_\theta|x) = E_\theta(\dot{\ell}_\theta\dot{\ell}'_\theta)$ holds. The necessary condition is satisfied by arbitrary $\nu \in \mathcal{N}^*$.

When a parametric model of $\mathcal{M}$ is assumed, choice of ν is relevant for identification and differentiability of the parameter, but irrelevant to estimation efficiency. In other words, the efficient design problem becomes degenerated when the model is parametric.

4.3 The dichotomous choice contingent valuation experiment

Consider an experiment $\mathcal{E}$ with $\mathcal{W} = \mathcal{X} = [0,\infty)$, $\mathcal{Y} = \{0,1\}$, and $\rho(\omega,x) = \{\omega < x\}$. This is the *dichotomous choice contingent valuation* (DC-CV) experiment, which is one of the most widely used experimental methods in environment economics [4]. Cooper (1992) reports results of Monte Carlo simulations of the DC-CV experiments, showing sensitivity of estimates to the choice of ν [5].

An inadequate design of ν may result in bias and/or large standard errors in estimates. The efficient design for the DC-CV experiment to estimate the mean $E\omega$ is proposed by [8] and [6]. In this subsection, their results will be generalized to estimation of arbitrary smooth functionals $\varphi(\mu)$.

Let $\mathcal{M}$ be a set of probability measures on $[0,\infty)$, where every $\mu \in \mathcal{M}$ is equivalent to the Lebesgue measure $\lambda = \lambda_{[0,\infty)}$ on $[0,\infty)$. Let $\mathcal{N} = \mathcal{M}$, then the model set $\rho_\nu(\mathcal{M}) = \{\rho_\nu(\mu) | \mu \in \mathcal{M}\}$ is given by

$$\rho_\nu(\mu)(D) = \int_D \mu[0,x)^y \mu[x,\infty)^{1-y}\, \nu(dx)\delta_{\mathcal{Y}}(dy) \tag{15}$$

for every Borel set $D \subset \mathbb{R}^2$, where $\delta_{\mathcal{Y}}$ is the Dirac measure on $\mathcal{Y}$. Note that $E_{\mu,\nu}(y|x) = \mu[0,x)$ holds. The score operator $(d\rho_\nu)_\mu$ maps each $\alpha \in L_2^0(\mu)$ to

$$((d\rho_\nu)_\mu \alpha)(x,y) = \frac{y - \mu[0,x)}{\mu[0,x)\mu[x,\infty)} \int_0^x \alpha\, d\mu, \tag{16}$$

and the adjoint operator is given by

$$((d\rho_\nu)_\mu^* \beta)(\omega) = \int_0^\omega \beta(x,0)\, \nu(dx) + \int_\omega^\infty \beta(x,1)\, \nu(dx). \tag{17}$$

Assume that the gradient $\omega \mapsto \partial\varphi_\mu(\omega)$ is differentiable with respect to ω with derivative $(\partial\varphi_\mu)'(\omega) = (d/d\omega)(\partial\varphi_\mu)(\omega)$, and that $\lim_{M\uparrow\infty} \partial\varphi_\mu(M)\mu[M,\infty) = 0$. If $\partial\varphi_\mu \in R((d\rho_\nu)_\mu)$ at some $\nu \sim \lambda$, β given by

$$\beta(x,y) := -\frac{d\lambda}{d\nu}(x)(\partial\varphi_\mu)'(x)(y - \mu[0,x)) \tag{18}$$

solves $\partial\varphi_\nu = (d\rho_\nu)_\mu^* \beta$ because

$$\begin{aligned}((d\rho_\nu)_\mu^* \beta)(\omega) &= \int_0^\omega (\partial\varphi_\mu)'(x)\, dx - \int_0^\infty (\partial\varphi_\mu)'(x)\, \mu[x,\infty)\, dx \\ &= \partial\varphi_\mu(\omega) - \lim_{M\uparrow\infty} \partial\varphi_\mu(M)\mu[M,\infty) \\ &= \partial\varphi_\mu(\omega).\end{aligned}$$

If $\beta \in \overline{R((d\rho_\nu)_\mu)}$ is satisfied, $\partial\kappa_{\rho_\nu(\mu)} = \beta$ is the efficient influence function to estimate $\varphi(\mu)$.

Let ν^* be efficient for $\mathcal{E}$ at μ. Then, there exists a positive constant C such that, for any $x \in [0,\infty)$,

$$\begin{aligned}C &= E_{\mu,\nu^*}\left[\left(-\frac{d\lambda}{d\nu^*}(x)\ (\partial\varphi_\mu)'(x)\ (y - \mu[0,x)) \right)^2 \middle| \, x \right] \\ &= \left(\frac{d\lambda}{d\nu^*}(x)(\partial\varphi_\mu)'(x) \right)^2 \mu[0,x)\mu[x,\infty),\end{aligned}$$

which implies

$$\frac{d\nu^*}{d\lambda}(x) = \frac{|(\partial\varphi_\mu)'(x)|\sqrt{\mu[0,x)\mu[x,\infty)}}{\int_0^\infty |(\partial\varphi_\mu)'(\xi)|\sqrt{\mu[0,\xi)\mu[\xi,\infty)}\,d\xi}\,. \tag{19}$$

The formula (19) is a generalization of the result of [8] and [6]. They find the optimal design for estimation of the mean $E\omega$ by directly minimizing the asymptotic variances of the maximum likelihood estimators under assumptions that ω has a finite support and that $\mathcal{X}$ is a finite set. Our formula (19), on the other hand, does not assume finite supports of ω and x and is applicable to estimations of any differentiable functionals of μ.

Particularly when the m-th moment $\varphi(\mu) = \int \omega^m\,\mu(d\omega)$ is a parameter of interest, the gradient of φ is $\partial\varphi_\mu(\omega) = \omega^m - \int \omega^m\,\mu(d\omega)$. This is because for every differentiable path $d\mu_t = \exp(t\alpha - \gamma_t)d\mu$ with $\alpha \in L_2^0(\mu)$,

$$\lim_{t\to 0}\frac{\varphi(\mu_t)-\varphi(\mu)}{t} = \int \omega^m\alpha(\omega)\,\mu(d\omega) = \langle\partial\varphi_\mu,\alpha\rangle_\mu\,. \tag{20}$$

The gradient is differentiable with respect to ω with derivative $(\partial\varphi_\mu)'(\omega) = m\omega^{m-1}$.

Assume that μ satisfies

(A3) $\lim_{M\uparrow\infty} M^m\mu[M,\infty) = 0$, and
(A4) $\int_0^\infty \xi^{m-1}\sqrt{\mu[0,\xi)\mu[\xi,\infty)}\,d\xi < \infty$.

A broad class of distributions μ satisfies the conditions. Then, an efficient design at such μ is solved as

$$\frac{d\nu^*}{d\lambda}(x) = \frac{x^{m-1}\sqrt{\mu[0,x)\mu[x,\infty)}}{\int_0^\infty \xi^{m-1}\sqrt{\mu[0,\xi)\mu[\xi,\infty)}\,d\xi}. \tag{21}$$

Figure 1 shows graphs of the efficient densities $d\nu^*/d\lambda$ at $\mu = U[0,1]$. It is shown that the efficient density tends to skew to the left as m grows.

The formula (19) is valid only if two technical assumptions are satisfied. The first is differentiability of κ at $\rho_{\nu^*}(\mu)$. Let β be a function given by

$$\beta(x,y) = -\frac{y-\mu[0,x)}{\sqrt{\mu[0,x)\mu[x,\infty)}}\left(\int_0^\infty \xi^{m-1}\sqrt{\mu[0,\xi)\mu[\xi,\infty)}\,d\xi\right). \tag{22}$$

Since $E_{\mu,\nu^*}\beta = 0$ and $E_{\mu,\nu^*}\beta^2 < \infty$, $\beta \in L_2^0(\rho_{\nu^*}(\mu))$. Moreover, since

$$\begin{aligned}((d\rho_{\nu^*})_\mu^*\beta)(\omega) &= \int_0^\omega \left(\frac{\mu[0,x)}{\sqrt{\mu[0,x)\mu[x,\infty]}}\right) x^{m-1}\sqrt{\mu[0,x)\mu[x,\infty)}\,dx \\ &\quad - \int_\omega^\infty \left(\frac{1-\mu[0,x)}{\sqrt{\mu[0,x)\mu[x,\infty]}}\right) x^{m-1}\sqrt{\mu[0,x)\mu[x,\infty)}\,dx \\ &= \omega^m - \int \omega^m\,\mu(d\omega) - \lim_{M\uparrow\infty}(M^m - \int \omega^m)\mu[M,\infty) \\ &= \omega^m - \int \omega^m\,\mu(d\omega) = \partial\varphi_\mu(\omega),\end{aligned}$$

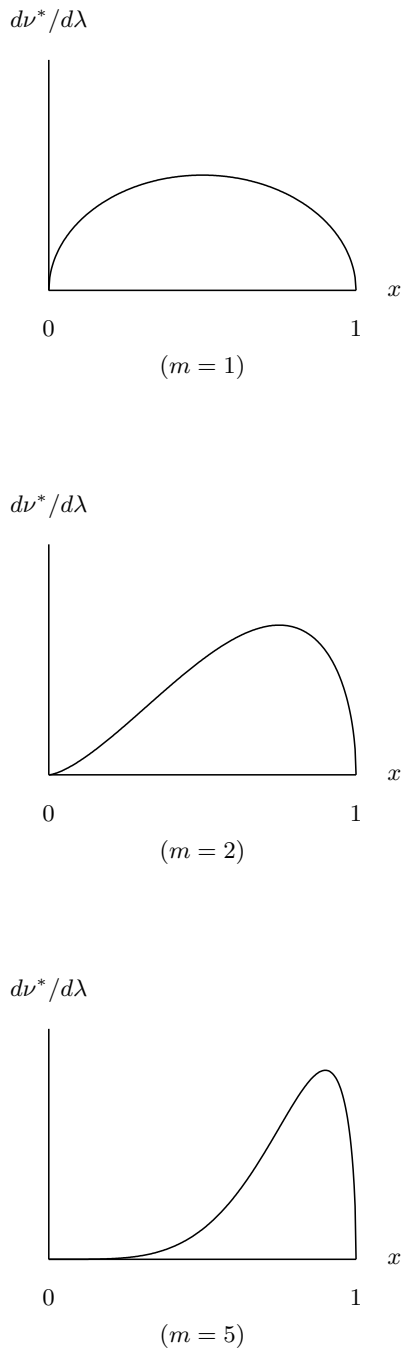

Fig. 1. The efficient densities of x to estimate $E\omega^m$ for the DC-CV experiment at $\mu = U[0, 1]$.

$\partial\varphi_\mu \in R((d\rho_{\nu^*})_\mu)$ is confirmed. Therefore, κ is indeed differentiable at $\rho_{\nu^*}(\mu)$, and $\nu^* \in \mathcal{N}^*$.

The second technical condition is $\beta \in \overline{R((d\rho_{\nu^*})_\mu)}$. Let α_0 be a function given by

$$\alpha_0(\omega) = -\frac{1-2\mu[0,\omega)}{2\sqrt{\mu[0,\omega)\mu[\omega,\infty)}}\left(\int_0^\infty \xi^{m-1}\sqrt{\mu[0,\xi)\mu[\xi,\infty)}\,d\xi\right), \tag{23}$$

then $\int_0^\omega |\alpha_0|\,d\mu < \infty$ and

$$\begin{aligned}((d\rho_{\nu^*})_\mu\alpha_0)(x,y) &= -\frac{y-\mu[0,x)}{\mu[0,x)\mu[x,\infty)}\left[\sqrt{\mu[0,\omega)\mu[\omega,\infty)}\right]_{\omega=0}^{x} \\ &\quad\times\left(\int_0^\infty \xi^{m-1}\sqrt{\mu[0,\xi)\mu[\xi,\infty)}\,d\xi\right) \\ &= -\frac{y-\mu[0,x)}{\sqrt{\mu[0,x)\mu[x,\infty)}}\left(\int_0^\infty \xi^{m-1}\sqrt{\mu[0,\xi)\mu[\xi,\infty)}\,d\xi\right) \\ &= \beta(x,y),\end{aligned}$$

but $\int_0^\omega (\alpha_0)^2\,d\mu < \infty$ fails to hold in general.

Now consider a sequence $\{\alpha_k\}_{k\geq 1}$ such that

$$\alpha_k(\omega) = \alpha_0(\omega)\left\{\frac{1}{k} < \mu[0,\omega) \leq 1-\frac{1}{k}\right\}. \tag{24}$$

Every α_k is bounded, and $E_\mu(\alpha_k)^2 < \infty$. Let $q(t)$ be a solution to $\mu[0,q(t)) = t$ for $0 < t < 1$, then

$$\begin{aligned}\int_0^\infty \alpha_k\,d\mu &= -\left[\sqrt{\mu[0,\omega)\mu[\omega,\infty)}\right]_{\omega=q(1/k)}^{q(1-1/k)} \\ &\quad\times\left(\int_0^\infty \xi^{m-1}\sqrt{\mu[0,\xi)\mu[\xi,\infty)}\,d\xi\right) = 0.\end{aligned} \tag{25}$$

Thus, $\{\alpha_k\} \subset L_2^0(\mu)$ is shown. At $\nu = \nu^*$, where ν^* is given by (21),

$$\begin{aligned}\|\beta - (d\rho_{\nu^*})_\mu\alpha_k\|^2_{\rho_{\nu^*}(\mu)} &= \|(d\rho_{\nu^*})_\mu(\alpha_0-\alpha_k)\|^2_{\rho_{\nu^*}(\mu)} \\ &= E_{\mu,\nu^*}\left(\frac{y-\mu[0,x)}{\mu[0,x)\mu[x,\infty)}\right)^2\left(\int_0^x(\alpha_0-\alpha_k)\,d\mu\right)^2 \\ &\leq 2\nu^*\left\{\mu[0,x) \notin \left[\frac{1}{k},1-\frac{1}{k}\right)\right\} \\ &\qquad\times\left(\int_0^\infty \xi^{m-1}\sqrt{\mu[0,\xi)\mu[\xi,\infty)}\,d\xi\right)^2 \\ &\quad+\frac{4}{k}\left(\int_0^\infty \xi^{m-1}\sqrt{\mu[0,\xi)\mu[\xi,\infty)}\,d\xi\right)^2 \\ &\to 0 \quad (k\to\infty).\end{aligned}$$

Thus, $\beta \in \overline{R((d\rho_{\nu^*})_\mu)}$ is shown. The efficient influence function of κ is indeed given by (22).

When the variance of ω is a parameter of interest,

$$\partial\varphi_\mu(\omega) = (\omega - E_\mu\omega)^2 - \varphi(\mu) \tag{26}$$

and $\frac{d}{d\omega}\partial\varphi_\mu(\omega) = 2(\omega - E_\mu\omega)$. The score equation $\partial\varphi_\mu = (d\rho_\nu)_\mu\beta$ has a solution,

$$\beta(x,y) = -\mathrm{sgn}\,(x - E_\mu\omega)\,\frac{y - \mu[0,x)}{\sqrt{\mu[0,x)\mu[x,\infty)}}\left(\int_0^\infty \xi^{m-1}\sqrt{\mu[0,\xi)\mu[\xi,\infty)}\,d\xi\right), \tag{27}$$

where $\mathrm{sgn}(t) = +1$ if $t \geq 0$, and $\mathrm{sgn}(t) = -1$ if $t < 0$. Therefore, κ is differentiable at $\rho_{\nu^*}(\mu)$. A candidate of the efficient density is therefore given by

$$\frac{d\nu^*}{d\lambda}(x) = \frac{|x - E_\mu\omega|\sqrt{\mu[0,x)\mu[x,\infty)}}{\int |u - E_\mu\omega|\sqrt{\mu[0,u)\mu[u,\infty)}\,du}. \tag{28}$$

Figure 2 depicts a graph of (28), which shows bimodality around the mean, $E\omega = 0.5$.

When the median of ω is a target of estimation,

$$\partial\varphi_\mu(\omega) = -\frac{d\lambda}{d\mu}(\varphi(\mu))\{\omega \in [0,\varphi(\mu))\}, \tag{29}$$

which implies that the score equation is not solvable. This is because $\partial\varphi_\mu(\omega)$ given above is discontinuous at $\omega = \varphi(\mu)$, while

$$((d\rho_\nu)^*_\mu\beta)(\omega) = \int_0^\omega \beta(x,0)\nu(dx) + \int_\omega^\infty \beta(x,1)\nu(dx) \tag{30}$$

is differentiable with respect to ω. Therefore, $\partial\varphi_\mu \notin R((d\rho_\nu)^*_\mu)$ for any $\nu \sim \lambda$. Regular estimation of the median is not possible, and the efficient design for the median estimation does not exist, either.

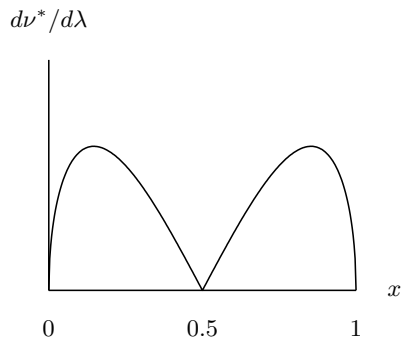

Fig. 2. The efficient density of x to estimate the variance of ω for the DC-CV experiment at $\mu = U[0, 1]$.

5 Monte Carlo Simulations

In this section, small sample properties of the efficient design for the DC-CV experiment are tested through Monte Carlo simulations. In the following, the mean $\theta := E\omega$ is chosen as a target of estimations.

In step 1 of the simulation, $\omega_1, \cdots, \omega_n$ are independently sampled from the true distribution, μ. The total sample size n is varied from 50 to 500. Inputs x_1^*, $\cdots$, x_n^* are sampled from the efficient distribution ν^*, which is determined by the formula (19), while $x_1^o, \cdots, x_n^o$ are sampled from the opponent distribution ν^o, which is arbitrarily selected. Outputs under the efficient and the opponent circumstances are respectively generated by $y_i^* = \{\omega_i < x_i^*\}$ and $y_i^o = \{\omega_i < x_i^o\}$ for $i = 1, \cdots, n$.

In step 2, the distribution function $F(\omega) := \mu[0, \omega)$ of ω is estimated by the Nonparametric Maximum Likelihood Estimation (NPMLE), which maximizes the log-likelihood functional,

$$\log L_n(F|\{x_i, y_i\}_{1\le i\le n}) := \sum_{i=1}^{n} \left[y_i \log F(x_i) + (1 - y_i)\log(1 - F(x_i))\right], \quad (31)$$

with respect to $F \in \mathcal{F}$, where $\mathcal{F}$ is a collection of distribution functions on $\mathbb{R}$. The problem is numerically solved by the isotonic regression technique, and the estimator is known to be semiparametrically efficient ([7], [9], [10], [14]). Let $\hat{F}_n^*$ and $\hat{F}_n^o$ be the NPMLEs respectively obtained from the efficient data, $\{x_i^*, y_i^*\}_{1\le i\le n}$, and the opponent data, $\{x_i^o, y_i^o\}_{1\le i\le n}$.

In step 3, estimates of $E\omega$ are computed by

$$\hat{\theta}_n := \int_0^\infty \omega\, \hat{F}_n(d\omega) \quad (32)$$

for $\hat{F}_n^*$ and $\hat{F}_n^o$. The simulation steps 1-3 are repeated $S = 10,000$ times, and the standard deviations,

$$\text{S.D.} = \sqrt{\frac{1}{S-1}\sum_{s=1}^{S}\left|\theta - \theta_n^{(s)}\right|^2}, \tag{33}$$

where $\theta_n^{(s)}$ is the estimate from the s-th round, are reported respectively under the efficient and the opponent designs.

Table 1 presents the standard deviations for $n = 50$, 250, and 500, where the true distribution μ is the uniform distribution on $(0, 1)$. The efficient scheme is given by

$$\frac{d\nu^*}{dx} = \frac{\sqrt{x(1-x)}}{\int_0^1 \sqrt{x(1-x)}\,dx}, \tag{34}$$

while $\nu^o = N(1/2, 1/7)$ is employed as the opponent. In Table 2, the results of simulations for the inversed setting, where $\mu = N(1/2, 1/7)$ and $\nu^o = U(0, 1)$, are presented. In both cases, the efficient design remarkably reduces the standard deviations even under small-sample environments with $n = 50$ as well as under moderate sample sizes.

Table 1. Standard deviations of the NPMLE for $E\omega$ under the efficient design ν^* and under the suboptimal design $\nu^o = N(1/2, 1/7)$, where the true distribution is $\mu = U(0, 1)$.

Size (n)	S.D. under ν^*	S.D. under ν^o	Ratio
50	0.06380	0.09074	0.7031
250	0.02788	0.05220	0.5343
500	0.01831	0.04121	0.4443

Table 2. Standard deviations of the NPMLE for $E\omega$ under the efficient design ν^* and under the suboptimal design $\nu^o = U(0, 1)$, where the true distribution is $\mu = N(1/2, 1/7)$.

Size (n)	S.D. under ν^*	S.D. under ν^o	Ratio
50	0.03190	0.04294	0.7429
250	0.01374	0.01858	0.7398
500	0.01015	0.01335	0.7598

References

1. Amari, S.: Information Geometry and Its Applications. Springer (2016)
2. Ay, N., Jost, J., Vân Lê, H., Schwachhöfer, L.: Information Geometry. Springer (2017)
3. Bickel, P.J., Klaassen, C.A.J., Ritov, Y., Wellner, J.A.: Efficient and Adaptive Estimation for Semiparametric Models. Johns Hopkins University Press (1993)
4. Carson, R., Hanemann, M.: Contingent valuation. In: Maler, K.G., Vincent, J.R. (eds.) Handbook of Environmental Economics, vol. 2, chap. 17, pp. 821–936. Elsevier, 1 edn. (2006)
5. Cooper, J., Loomis, J.: Sensitivity of willingness-to-pay estimates to bid design in dichotomous choice contingent valuation models. Land economics pp. 211–224 (1992)
6. Cooper, J.C.: Optimal bid selection for dichotomous choice contingent valuation surveys. Journal of Environmental Economics and Management **24**(1), 25–40 (1993)
7. Cosslett, S.R.: Distribution-free maximum likelihood estimator of the binary choice model. Econometrica **51**(3), 765–782 (1983)
8. Duffield, J.W., Patterson, D.A.: Inference and optimal design for a welfare measure in dichotomous choice contingent valuation. Land Economics **67**(2), 225–239 (1991)
9. Groeneboom, P., Jongbloed, G.: Nonparametric Estimation under Shape Constraints. Cambridge University Press (2014)
10. Groeneboom, P., Wellner, J.A.: Information Bounds and Nonparametric Maximum Likelihood Estimation. Birkhäuser Verlag (1992)
11. Pistone, G., Sempi, C., et al.: An infinite-dimensional geometric structure on the space of all the probability measures equivalent to a given one. The annals of statistics **23**(5), 1543–1561 (1995)
12. van der Vaart, A.W.: On differentiable functionals. Annals of Statistics **19**(1), 178–204 (1991)
13. van der Vaart, A.W.: Asymptotic Statistics. Cambridge University Press (1998)
14. van der Vaart, A.W., Wellner, J.A.: Weak Convergence and Empirical Processes: With Applications to Statistics. Springer-Verlag, New York (1996)