Efficient Design of Randomized Experiments

Hisatoshi Tanaka

School of Political Science and Economics

Waseda University

hstnk@waseda.jp

1: Motivation

EXAMPLE

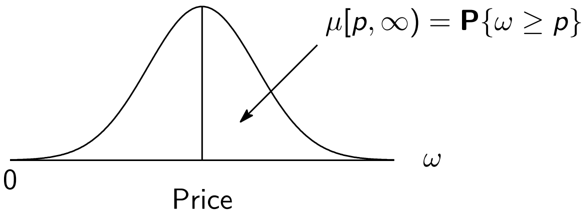
- Imagine that N consumers exist, whose Willingness-To-Pay (WTP) to a certain good are

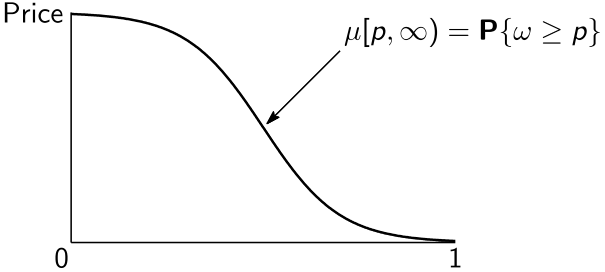
$$\omega_1, \ldots, \omega_N \sim \ i.i.d. \ \mu.$$

- Consumer i would buy the good if and only if

$$\omega_i \geq \text{price}$$

- Function $D(p) := 1 - \mu[0, p)$ is the demand function to the good.


$\mu[p,\infty) = \mathbf{P}\{\omega \geq p\}$
ω
0
Price


Price
$\mu[p,\infty) = \mathbf{P}\{\omega \geq p\}$
0
1

- Experiment to estimate μ:

(Step 1) To consumer i $(= 1, \cdots, N)$, a randomly chosen bidding price x_i is offered.

(Step 2) If the consumer accepts the offer, $y_i = 1$; otherwise, $y_i = 0$, that is,

$$y_i = \mathbb{I}\{\omega_i \geq x_i\}.$$

(Step 3) i.i.d. pairs (x_1, y_1), ..., (x_N, y_N) are obtained.

(Step 4) μ is estimated by the MLE.

PROBLEM

Performance of the estimation depends not only on the estimation method but also on the distribution of x.

- If $x_1 = \cdots = x_N = 0$ is offered, trivial outcomes $y_1 = \cdots = y_N = 1$ will be obtained.
- If $x_1 = \cdots = x_N = +\infty$ is offered, $y_1 = \cdots = y_N = 0$ will occur with probability 1.

What is the 'best distribution' ν of x, with which the experiment produces the most informative data?

Efficient Design Problem

$$\min \quad \text{l.b.}(\theta(\mu)|\nu) \quad \text{with respect to} \quad \nu \in \mathcal{P}_+(\mathcal{X})$$

- l.b.$(\theta(\mu)|\nu)$ = efficiency bound to estimate $\theta(\mu)$ when $x \sim \nu$
- $\theta(\mu)$ = a smooth functional of μ, say $\theta(\mu) = \int \omega \, d\mu(\omega)$
- $\mathcal{P}_+(\mathcal{X})$ = a set of positive probability measures on $\mathcal{X}$

Duffield et al (1991), Cooper (1993)

- $\rho(\omega, x) = \mathbb{I}\{\omega \geq x\}$ (the Contingent Valuation Method)
- $\theta(\mu) = \int \omega \, d\mu(\omega)$

In my work,

- $\rho(\omega, x)$ and $\theta(\mu)$ are not necessary specified.
- Information Geometry is applied.

TOC:

- 1: Motivation ✓
- 2: The Model
- 3: Main Results
- 4: Application examples

2. The Model

Nihat Ay, Jürgen Jost, Hông Vân Lê, Lorenz Schwachhöfer (2018): *Information Geometry*, Springer-Verlag.

2: The Model

- $\mathcal{W} = \{\omega_1, \cdots, \omega_n\}$ (finite measures)
- $\mathcal{F}(\mathcal{W})$ is the set of functions with $e^i(\omega_j) = \begin{cases} 1 & (i = j) \\ 0 & (i \neq j) \end{cases}$
- $\mathcal{S}(\mathcal{W})$ is the set of signed measures with

$$\delta_i(\omega_j) = \begin{cases} 1 & (i = j) \\ 0 & (i \neq j) \end{cases}$$

- Each $\mu \in \mathcal{S}(\mathcal{W})$ is expressed as $\mu = \sum_{i=1}^n \mu^i \delta_i$ so that

$$\mu(f) := \int f \, d\mu := \sum_{i=1}^n \mu^i f_i, \quad f \in \mathcal{F}(\mathcal{W})$$

2: The Model

- Given point $\mu \in \mathcal{S}(\mathcal{W})$, the tangent space of $\mathcal{S}(\mathcal{W})$ at μ is

$$T_\mu \mathcal{S}(\mathcal{W}) = \mathcal{S}(\mathcal{W})$$

2: The Model

- Let $\mathcal{P}_+(\mathcal{W}) = \left\{\mu \in \mathcal{S}(\mathcal{W}) \,:\, \mu \gg 0,\ \sum_{i=1}^n \mu^i = 1\right\}$.
- The tangent space of $\mathcal{P}_+(\mathcal{W})$ at $\mu \in \mathcal{P}_+(\mathcal{W})$ is

$$T_\mu \mathcal{P}_+(\mathcal{W}) = \mathcal{S}_0(\mathcal{W}) := \left\{\mu \in \mathcal{S}(\mathcal{W}) \,:\, \sum_{i=1}^n \mu^i = 0\right\}.$$

2: The Model

- For a and b in $T_\mu \mathcal{P}_+(\mathcal{W}) = \mathcal{S}_0(\mathcal{W})$,

$$\frac{\mathrm{d}a}{\mathrm{d}\mu} = \sum_{i=1}^{n} \frac{a^i}{\mu^i} e^i \quad \text{and} \quad \frac{\mathrm{d}b}{\mathrm{d}\mu} = \sum_{i=1}^{n} \frac{b^i}{\mu^i} e^i.$$

- The Fisher metric on $T_\mu \mathcal{P}_+(\mathcal{W})$ is

$$\mathfrak{g}_\mu(a, b) := \int \left(\frac{\mathrm{d}a}{\mathrm{d}\mu} \cdot \frac{\mathrm{d}b}{\mathrm{d}\mu} \right) d\mu = \sum_{i=1}^{n} \frac{a^i b^i}{\mu^i}, \qquad (1)$$

- The Fisher norm is $\|a\|_\mu := \sqrt{\mathfrak{g}_\mu(a, a)}$.

2: The Model

For two vector fields $A : \mu \mapsto a_\mu$ and $B : \mu \mapsto b_\mu$ on $\mathcal{P}_+(\mathcal{W})$,

- the *m*-connection $\nabla^{(m)}$:

$$\nabla_A^{(m)} B\Big|_\mu := \frac{\partial b}{\partial a_\mu}(\mu)$$

- the *e*-connection $\nabla^{(e)}$:

$$\nabla_A^{(e)} B\Big|_\mu := \frac{\partial b}{\partial a_\mu}(\mu) - \left(\frac{\mathrm{d}a_\mu}{\mathrm{d}\mu} \cdot \frac{\mathrm{d}b_\mu}{\mathrm{d}\mu} - \mathfrak{g}_\mu(a_\mu, b_\mu)\right) \mu$$

2: The Model

For $A : \mu \mapsto a_\mu$, $B : \mu \mapsto b_\mu$, and $C : \mu \mapsto c_\mu$,

$$\frac{\partial}{\partial c_\mu}(\mathfrak{g}(A, B))_\mu = \mathfrak{g}_\mu\Big(\nabla_C^{(m)} A\Big|_\mu, B_\mu\Big) + \mathfrak{g}_\mu\Big(A_\mu, \nabla_C^{(e)} B\Big|_\mu\Big)$$

2: The Model

- $\theta(\mu)$ is the target of estimation, say

$$\theta(\mu) = \int \omega \, d\mu(\omega) = \sum_i \omega_i \mu^i$$

- The differential of $\theta : \mathcal{P}_+(\mathcal{W}) \to \mathbb{R}$ in μ:

$$(d\theta)_\mu : T_\mu \mathcal{P}_+(\mathcal{W}) \to \mathbb{R},$$

$$a \mapsto \frac{\partial \theta}{\partial a}(\mu) := \lim_{t \to 0} \frac{\theta(\mu + ta) - \theta(\mu)}{t}.$$

- The gradient $\text{grad}_\mu \theta \in T_\mu \mathcal{P}_+(\mathcal{W})$ of θ at μ is

$$(d\theta)_\mu a \equiv \mathfrak{g}_\mu(a, \text{grad}_\mu \theta), \quad a \in T_\mu \mathcal{P}_+(\mathcal{W}). \tag{2}$$

2: The Model

$\rho : \mathcal{W} \times \mathcal{X} \mapsto \mathcal{Y}$, $\rho(\omega, x) = y$, is a response function, where

$$\mathcal{W} = \{\omega_1, \cdots, \omega_n\},\ \mathcal{X} = \{x_1, \cdots, x_m\},\ \mathcal{Y} = \{y_1, \cdots, y_\ell\}$$

When $\omega \sim \mu$ and $x \sim \nu$, the joint distribution of (x, y) is given by

$$P_{\mu,\nu}(x, y) = \left(\int \mathbb{I}\{\rho(\omega, x) = y\}\, d\mu(\omega) \right) \cdot \nu(x). \qquad (3)$$

Through abuse of notation,

$$\rho : (\mu, \nu) \mapsto P_{\mu,\nu} = \left(\int \mathbb{I}_\rho d\mu \right) \cdot \nu$$

where $\mathbb{I}_\rho(\omega, x, y) := \mathbb{I}\{\rho(\omega, x) = y\}$.

2: The Model

- The joint distribution of (x, y):

$$\rho(\mu, \nu) = \left(\int \mathbb{I}_{\rho} d\mu \right) \cdot \nu$$

- The mappings ρ_{ν} and ρ_{μ} are also provided as

$$\rho_{\nu}(\cdot) = \rho(\cdot, \nu)$$

and

$$\rho_{\mu}(\cdot) = \rho(\mu, \cdot).$$

Assumptions

(A1) $\rho(\mu, \nu) > 0$ for every $\mu \in \mathcal{P}_+(\mathcal{W})$ and $\nu \in \mathcal{P}_+(\mathcal{X})$,

(A2) $\rho(\delta_1, \nu), \cdots, \rho(\delta_n, \nu)$ are linearly independent.

Under **(A1)** and **(A2)**, ρ_ν becomes one to one:

$$\rho_\nu(\mu_1) - \rho_\nu(\mu_2) = \sum_{i=1}^{n-1} (\mu_1^i - \mu_2^i)(\rho(\delta_i, \nu) - \rho(\delta_n, \nu)),$$

which becomes 0 when and only when $\mu_1 = \mu_2$.

2: The Model

The model sets are given by

$$\mathcal{E} := R(\rho) := \{\rho(\mu, \nu) \,:\, \mu \in \mathcal{P}_+(\mathcal{W}), \nu \in \mathcal{P}_+(\mathcal{X})\}$$

$$\mathcal{E}_\nu := R(\rho_\nu) = \{\rho(\mu, \nu) \,:\, \mu \in \mathcal{P}_+(\mathcal{W})\}$$

$$\mathcal{E}_\mu := R(\rho_\mu) = \{\rho(\mu, \nu) \,:\, \nu \in \mathcal{P}_+(\mathcal{X})\}$$

which are submanifolds of $\mathcal{P}_+(\mathcal{X} \times \mathcal{Y})$.

2: The Model

The partial differentials of $\rho(\mu,\nu) = \left(\int \mathbb{I}_\rho d\mu\right) \cdot \nu$ are

$$(d\rho_\nu)_\mu : T_\mu \mathcal{P}_+(\mathcal{W}) \to T_{\rho(\mu,\nu)}\mathcal{E}, \quad \sigma \mapsto \left(\int \mathbb{I}_\rho d\sigma\right) \cdot \nu \tag{4}$$

$$(d\rho_\mu)_\nu : T_\nu \mathcal{P}_+(\mathcal{X}) \to T_{\rho(\mu,\nu)}\mathcal{E}, \quad \eta \mapsto \left(\int \mathbb{I}_\rho d\mu\right) \cdot \eta \tag{5}$$

and the total differential of $\rho(\mu,\nu) = \left(\int \mathbb{I}_\rho d\mu\right) \cdot \nu$ is

$$(d\rho)_{\mu,\nu} = (d\rho_\nu)_\mu + (d\rho_\mu)_\nu, \quad (\sigma,\eta) \mapsto \rho(\sigma,\nu) + \rho(\mu,\eta)$$

2: The Model

Tangent spaces of the model sets are

- $T_{\rho(\mu,\nu)}\mathcal{E}_\nu = R((d\rho_\nu)_\mu)$
- $T_{\rho(\mu,\nu)}\mathcal{E}_\mu = R((d\rho_\mu)_\nu)$
- $T_{\rho(\mu,\nu)}\mathcal{E} = T_{\rho(\mu,\nu)}\mathcal{E}_\nu \oplus T_{\rho(\mu,\nu)}\mathcal{E}_\mu$

2: The Model

The adjoint operator $(d\rho_\nu)^*_\mu$ plays an important role to compute the efficiency bound:

$$\mathfrak{g}_{\rho(\mu,\nu)}\left((d\rho_\nu)_\mu\sigma, \tau\right) = \mathfrak{g}_\mu(\sigma, (d\rho_\nu)^*_\mu\tau) \tag{6}$$

for every $\sigma \in T_\mu\mathcal{P}_+(\mathcal{W})$ and $\tau \in T_{\rho(\mu,\nu)}\mathcal{E}_\nu$,

which is solved by

$$(d\rho_\nu)^*_\mu\tau = \int \left(\frac{\mathbb{I}_\rho}{\int \mathbb{I}_\rho d\mu}\right) d\tau \cdot \mu, \tag{7}$$

which is independent of ν.

3. Main Results

Proposition (van der Vaart, 1991)

$$\text{l.b.}(\theta(\mu)|\nu) = \|\hat{\tau}\|^2_{\rho(\mu,\nu)}, \tag{8}$$

where $\hat{\tau} = \hat{\tau}(\mu, \nu)$ is a solution to the Score Equation,

$$\text{grad}_\mu \theta = (d\rho_\nu)^*_\mu \hat{\tau}, \quad \hat{\tau} \in T_{\rho(\mu,\nu)}\mathcal{E}_\nu. \tag{9}$$

Proof

Consider a smooth path $t \in (-\epsilon, \epsilon) \mapsto \mu_t \in \mathcal{P}_+(\mathcal{W})$, which passes through μ_0 at $t = 0$ with

$$\left(\frac{d}{dt}\right)_{t=0} \mu_t = \sigma.$$

The efficiency bound of "true" $t = 0$ is the inverse of the Fisher information of $t = 0$:

$$E_{\mu_0,\nu}\left(\left(\frac{d}{dt}\right)_{t=0} \log \rho_\nu(\mu_t)\right)^2 = \|(d\rho_\nu)_{\mu_0}\sigma\|^2_{\rho(\mu_0,\nu)}.$$

Proof

The efficiency bound of $\theta_0 = \theta(\mu_0)$ along with the one-parameter submodel $\{\rho_\nu(\mu_t) : -\epsilon < t < \epsilon\}$ is

$$\begin{aligned} \lambda(\sigma) &:= \left(\frac{\partial\theta}{\partial\sigma}(\mu_0)\right)^2 \|(d\rho_\nu)_{\mu_0}\sigma\|^{-2}_{\rho(\mu_0,\nu)} \\ &= \mathfrak{g}_{\rho(\mu_0,\nu)}\left(\hat{\tau}(\mu_0,\nu), \frac{(d\rho_\nu)_{\mu_0}\sigma}{\|(d\rho_\nu)_{\mu_0}\sigma\|_{\rho(\mu_0,\nu)}}\right)^2. \end{aligned}$$

Since $\hat{\tau}(\mu_0,\nu) \in R((d\rho_\nu)_{\mu_0})$,

$$\lambda(\theta_0|\nu) = \sup_{\sigma\in T_{\mu_0}\mathcal{P}_+(\mathcal{W})} \lambda(\sigma) = \|\hat{\tau}(\mu_0,\nu)\|^2_{\rho(\mu_0,\nu)}$$

is obtained.

[QED]

3: Main Results

Proposition l.b.$(\theta(\mu)|\nu) = \|\hat{\tau}(\mu,\nu)\|^2_{\rho(\mu,\nu)}$ is convex in ν.

Proof

Let $G(\nu) := [g_{i,h}(\nu)]$ be an $n \times n$ matrix with the (i,h) element

$$g_{i,h}(\nu) := \mathfrak{g}_{\rho(\mu_0,\nu)}\left(\rho(\delta_i,\nu), \rho(\delta_h,\nu)\right)$$

for $1 \le i \le n$ and $1 \le h \le n$. The matrix is linear in ν and nonsingular according to **(A2)**.

Proof

Because $\hat{\tau}(\mu_0, \nu)$ is in $T_{\mu_0}\mathcal{E}_\nu$, there exists $\hat{\sigma}_\nu \in T_{\mu_0}\mathcal{E}_\nu$ such that $\hat{\tau}(\mu_0, \nu) = \rho(\hat{\sigma}_\nu, \nu)$. Moreover, for $1 \leq i \leq n$,

$$\begin{aligned}
\frac{\mathrm{d}(\mathrm{grad}_{\mu_0}\theta)}{\mathrm{d}\mu_0}(\omega_i) &= \hat{\tau}(\mu_0, \nu)\left(\frac{\mathbb{I}_\rho}{\mu_0(\mathbb{I}_\rho)}\right)(\omega_i) \\
&= \int \frac{\mathbb{I}_\rho(\omega_i, x, y)}{\mu_0(\mathbb{I}_\rho)(x, y)}\,\mathrm{d}\rho(\hat{\sigma}_\nu, \nu)(x, y) \\
&= \int \frac{\mathrm{d}\rho(\delta_i, \nu)}{\mathrm{d}\rho(\mu_0, \nu)} \cdot \frac{\mathrm{d}\rho(\hat{\sigma}_\nu, \nu)}{\mathrm{d}\rho(\mu_0, \nu)}\,\mathrm{d}\rho(\mu_0, \nu) \\
&= \mathfrak{g}_{\rho(\mu_0, \nu)}\left(\rho(\delta_i, \nu), \rho(\hat{\sigma}_\nu, \nu)\right) \\
&= \sum_{h=1}^{n} g_{i,h}(\nu)\hat{\sigma}_\nu^h. \qquad (10)
\end{aligned}$$

Proof

Let $\gamma = (\gamma_1, \cdots, \gamma_n)^\top$ be a vector of coefficients of $\mathrm{d}(\mathrm{grad}_{\mu_0}\theta)/\mathrm{d}\mu_0$, and let $\hat{\sigma}_\nu = (\hat{\sigma}_\nu^1, \cdots, \hat{\sigma}_\nu^n)^\top$. Then, (10) implies that $\hat{\sigma}_\nu = G(\nu)^{-1}\gamma$. From the previous proposition,

$$\lambda(\theta_0|\nu) = \hat{\sigma}_\nu^\top G(\nu)\hat{\sigma}_\nu = \gamma^\top G(\nu)^{-1}\gamma.$$

Therefore,

$$\begin{aligned}
\lambda(\theta_0|\, t\nu_1 + (1-t)\nu_2\,) &= \gamma^\top G(t\nu_1 + (1-t)\nu_2)^{-1}\gamma \\
&= \gamma^\top \Big[tG(\nu_1) + (1-t)G(\nu_2) \Big]^{-1}\gamma \\
&\leq t\lambda(\theta_0|\nu_1) + (1-t)\lambda(\theta_0|\nu_2)
\end{aligned}$$

for arbitrary ν_1 and ν_2 in $\mathcal{P}_+(\mathcal{X})$ and for any $t \in [0,1]$.

[QED]

3: Main Results

Definition

An *efficient design for* estimation of $\theta(\mu)$ is $\nu^* \in \mathcal{P}_+(\mathcal{X})$ such that

$$\nu^* = \textit{arg}\min \ \ \text{l.b.}(\theta(\mu)|\nu) \quad \text{s.t.} \quad \nu \in \mathcal{P}_+(\mathcal{X}) \tag{11}$$

THEOREM

ν is efficient for estimation of $\theta(\mu)$ if and only if

$$\mathfrak{g}_{\rho(\mu,\nu)}\left(\hat{\tau}(\mu,\nu), \nabla^{(e)}_{\eta}\hat{\tau}(\mu,\nu)\right) = 0 \tag{12}$$

is satisfied for any $\eta \in T_\nu\mathcal{P}_+(\mathcal{X})$.

Proof

Because $\nu \mapsto I.b.(\theta(\mu)|\nu)$ is convex, it is minimized at ν if and only if the first order condition

$$\left(\frac{\partial}{\partial \eta}\right) I.b.(\theta(\mu)|\nu) = 0 \tag{13}$$

holds for any $\eta \in T_\nu \mathcal{P}_+(\mathcal{X})$.

Proof

Because l.b.$(\theta(\mu)|\nu) = \|\hat{\tau}\|^2_{\rho(\mu,\nu)} = \mathfrak{g}_{\rho(\mu,\nu)}(\hat{\tau}, \hat{\tau})$,

$$\mathfrak{g}_{\rho(\mu,\nu)}\left(\nabla^{(m)}_{\eta}\hat{\tau}, \hat{\tau}\right) + \mathfrak{g}_{\rho(\mu,\nu)}\left(\hat{\tau}, \nabla^{(e)}_{\eta}\hat{\tau}\right) = 0. \qquad (14)$$

By differentiating both the sides of the score equation

$$\mathrm{grad}_{\mu}\theta = (d\rho_{\nu})^*_{\mu}\hat{\tau}$$

with respect to ν,

$$0 = (d\rho_{\nu})^*_{\mu}\left(\frac{\partial\hat{\tau}}{\partial\eta}(\mu, \nu)\right) = (d\rho_{\nu})^*_{\mu}\nabla^{(m)}_{\eta}\hat{\tau},$$

which implies $\mathfrak{g}_{\rho(\mu,\nu)}\left(\nabla^{(m)}_{\eta}\hat{\tau}, \hat{\tau}\right) = 0.$ [QED]

Corollary

ν is efficient if and only if

$$E_{\mu}\left[\left(\frac{\mathrm{d}\hat{\tau}(\mu,\nu)}{\mathrm{d}\rho(\mu,\nu)}(x,y)\right)^{2}\middle| x\right] \equiv Const. \tag{15}$$

for all $x \in \mathcal{X}$.

Proof

Substitute the definition of $\nabla^{(e)}$,

$$\nabla^{(e)}_{\eta}\hat{\tau}(\mu,\nu) = \frac{\partial\hat{\tau}}{\partial\eta}(\mu,\nu) - \left(\frac{\mathrm{d}\eta}{\mathrm{d}\nu}\cdot\frac{\mathrm{d}\hat{\tau}(\mu,\nu)}{\mathrm{d}\rho(\mu,\nu)}\right)\rho(\mu,\nu),$$

into $\mathfrak{g}\left(\hat{\tau},\nabla^{(e)}_{\eta}\hat{\tau}\right) = 0$. Since $\mathfrak{g}(\rho(\mu,\eta),\hat{\tau}(\mu,\nu)) = 0$,

$$E_{\mu,\nu}\left[\frac{\mathrm{d}\eta}{\mathrm{d}\nu}(x)\cdot E_{\mu}\left[\left(\frac{\mathrm{d}\hat{\tau}(\mu,\nu)}{\mathrm{d}\rho(\mu,\nu)}(x,y)\right)^{2}\middle| x\right]\right] \equiv 0$$

holds for $^{\forall}\eta \in \mathcal{S}_0(\mathcal{X})$. [QED]

The intuition behind this condition can be obtained from the following expression:

$$\begin{aligned} l.b.(\theta \mid \nu) &= \|\hat{\tau}\|^2 \\ &= E_{\mu,\nu}\left[\left(\frac{\mathrm{d}\hat{\tau}(\mu,\nu)}{\mathrm{d}\rho(\mu,\nu)}(x,y)\right)^2\right] \\ &= \int E_{\mu}\left[\left(\frac{\mathrm{d}\hat{\tau}(\mu,\nu)}{\mathrm{d}\rho(\mu,\nu)}(x,y)\right)^2 \,\middle|\, x\right] \mathrm{d}\nu(x). \quad (16) \end{aligned}$$

If the lower bound is minimized at ν, any small perturbations η added to ν will not significantly change the value of $l.b.(\theta \mid \nu)$. This is possible if and only if the integrand on the right-hand side is independent of x.

4: Application Example

Consider an experiment using the binary response function

$$y = \mathbb{I}\{\omega \geq x\},$$

with

- $\mathcal{W} = \{\xi_1, \cdots, \xi_{n-1}, \xi_n\}$
- $\mathcal{X} = \{\xi_1, \cdots, \xi_{n-1}\}$
- $\mathcal{Y} = \{0, 1\}$

where $\xi_1, \cdots, \xi_n$ are n real numbers such that $0 \leq \xi_1 < \cdots < \xi_n$.

- This experiment is also known as the *dichotomous choice contingent valuation* (DC-CV) experiment in the literature on environmental economics (see *e.g.* Carson and Hanemann, 2006).

- An efficient design for the experiment, in which the mean $E_\mu \omega$ is the target of estimation, was determined by Duffield et al (1991) and Cooper (1993).

In the experiment, the joint distribution of (x, y) is obtained by

$$\rho(\mu, \nu)(x, y) = (y \cdot \mu[0, x] + (1 - y) \cdot \mu(x, \infty))\, \nu(x) \tag{17}$$

The differential of ρ_ν is obtained by

$$((d\rho_\nu)_\mu \sigma)(\xi_j, y) = \big(y \cdot \sigma[0, \xi_j] + (1 - y) \cdot \sigma(\xi_j, \infty)\big)\nu^j$$

for every $\sigma \in T_\mu \mathcal{P}_+(\mathcal{W}) = \mathcal{S}_0(\{\xi_1, \cdots, \xi_n\})$. The adjoint operator is determined by

$$((d\rho_\nu)^*_\mu \tau)(\xi_i) = \left(\sum_{j=i}^{n-1} \nu_j \frac{\sigma[0, \xi_j]}{\mu[0, \xi_j]} + \sum_{j=1}^{i-1} \nu_j \frac{\sigma(\xi_j, \infty)}{\mu(\xi_i, \infty)} \right) \mu^i$$

for each $\tau = \rho(\sigma, \nu) \in T_{\rho(\mu,\nu)} \mathcal{E}_\nu$.

Let $\gamma := \mathrm{d}(\mathsf{grad}_\mu\, \theta)/\mathrm{d}\mu = \sum_{i=1}^{n} \gamma_i e^i$ with

$$\gamma_i = \begin{cases} \dfrac{\partial\theta}{\partial\mu^i}(\mu) - \displaystyle\sum_{h=1}^{n-1} \mu^h \dfrac{\partial\theta}{\partial\mu^h}(\mu) & (1 \le i \le n-1) \\ \\ -\displaystyle\sum_{h=1}^{n-1} \mu^h \dfrac{\partial\theta}{\partial\mu^h}(\mu) & (i = n) \end{cases} .$$

Let $\hat{\tau}(\mu, \nu) = \rho(\hat{\sigma}_\nu, \nu)$, where $\hat{\sigma}_\nu$ satisfies

$$\gamma_i = \sum_{j=i}^{n-1} \nu_j \frac{\hat{\sigma}_\nu[0, \xi_j]}{\mu[0, \xi_j]} + \sum_{j=1}^{i-1} \nu_j \frac{\hat{\sigma}_\nu(\xi_j, \infty)}{\mu(\xi_i, \infty)} \tag{18}$$

for $1 \leq i \leq n$. This equation is solved by

$$\hat{\sigma}_\nu[0, \xi_j] = -\frac{\gamma_{j+1} - \gamma_j}{\nu_j} \mu[0, \xi_j] \mu(\xi_j, \infty)$$

for $1 \leq j \leq n-1$ and $\hat{\sigma}_\nu[0, \xi_n] = 0$.

Thus, we obtain

$$\frac{\mathrm{d}\hat{\tau}(\mu,\nu)}{\mathrm{d}\rho(\mu,\nu)}(\xi_j, y) = -\frac{\gamma_{j+1}-\gamma_j}{\nu_j}\Big(y\mu(\xi_j,\infty) - (1-y)\mu[0,\xi_j]\Big)$$

and

$$l.b.(\theta(\mu)\mid\nu) = \sum_{j=1}^{n-1}\frac{(\gamma_{j+1}-\gamma_j)^2}{\nu_j}\mu[0,\xi_j]\mu(\xi_j,\infty), \qquad (19)$$

where

$$\gamma = \frac{\mathrm{d}(\mathrm{grad}_\mu\theta)}{\mathrm{d}\mu} = \sum_{j=1}^{n}\gamma_j\delta^j.$$

Because

$$E_\mu \left[\left(\frac{\mathrm{d}\hat{\tau}(\mu,\nu)}{\mathrm{d}\rho(\mu,\nu)}(x,y) \right)^2 \middle| \; x = \xi_j \right] = \left(\frac{\gamma_{j+1} - \gamma_j}{\nu_j} \right)^2 \mu[0,\xi_j]\mu(\xi_j,\infty),$$

$$E_\mu \left[\left(\frac{\mathrm{d}\hat{\tau}(\mu,\nu)}{\mathrm{d}\rho(\mu,\nu)}(x,y) \right)^2 \middle| \; x = \xi_j \right] \equiv Const.$$

holds if and only if

$$\nu(\xi_j) \propto |\gamma_{j+1} - \gamma_j| \sqrt{\mu[0,\xi_j] \cdot \mu(\xi_j,\infty)} \tag{20}$$

for $j = 1, \cdots, n-1$, where

$$\gamma = \frac{\mathrm{d}(\mathrm{grad}_\mu \theta)}{\mathrm{d}\mu} = \sum_{j=1}^{n} \gamma_j \delta^j.$$

In particular, when $\theta(\mu) = \int f \mathrm{d}\mu$ with $f \in \mathcal{F}(\mathcal{W})$, the efficient design for $\theta(\mu)$ is obtained by

$$\nu(\xi_j) \propto |f(\xi_{j+1}) - f(\xi_j)| \sqrt{\mu[0, \xi_j] \cdot \mu(\xi_j, \infty)}. \tag{21}$$

However, the efficient design is infeasible because it contains an unknown μ. A feasible alternative is the *min-max design* that is defined by

$$\nu_{\text{Min-Max}} := \mathit{Arg} \min_{\nu \in \mathcal{P}_+(\mathcal{X})} \left[\max_{\mu \in \mathcal{P}_+(\mathcal{W})} \mathit{l.b.}(\theta(\mu)|\nu) \right].$$

In the binary experiment, the maximal risk to estimate $\theta(\mu) = \int f \mathrm{d}\mu$ is equal to

$$\max_{\mu \in \mathcal{P}_+(\mathcal{W})} l.b.(\theta(\mu) \mid \nu) = \sum_{j=1}^{n-1} \frac{(f(\xi_{j+1}) - f(\xi_j))^2}{4\nu^j},$$

where the maximum is obtained by $\mu = \delta_1/2 + \delta_n/2$. The risk is minimized by

$$\nu_{\text{Min-Max}}(\xi_j) \propto \left| f(\xi_{j+1}) - f(\xi_j) \right|. \tag{22}$$

In particular, when $\mathcal{W}$ is equally spaced so that $\xi_2 - \xi_1 = \cdots = \xi_n - \xi_{n-1}$, the min-max design for estimating the mean $E_\mu \omega$ becomes a uniform distribution on $\mathcal{X}$.

References

- Amari, S. (2016): Information Geometry and Its Applications, Springer Japan.
- Amari,S.and H.Nagaoka (2000) : Methods of information geometry, vol.191, American Mathematical Soc.
- Ay, N., Jost, J., Le, H.V., Schwachhofer, L.: Information Geometry, Ergebnisse der Mathematik und ihrer Grenzgebiete 3. Forge, 64. Springer (2017)
- Carson, R. and M. Hanemann (2006): Contingent Valuation in Handbook of Environmental Economics, ed. by K. G. Maler and J. R. Vincent, Elsevier, vol. 2, chap. 17, 821—936, 1 ed.
- Cooper, J. and J. Loomis (1992): Sensitivity of Willingness-To-Pay Estimates to Bid Design in Dichotomous Choice Contingent Valuation Models, Land economics, 68, 211—224.

References

- Cooper, J. C. (1993): Optimal Bid Selection for Dichotomous Choice Contingent Valuation Surveys, Journal of Environmental Economics and Management, 24, 25—40.
- Duffield, J. W. and D. A. Patterson (1991): Inference and Optimal Design for a Welfare Measure in Dichotomous Choice Contingent Valuation, Land Economics, 67, 225—239.
- van der Vaart, A. W. (1991): On Differentiable Functionals, Annals of Statistics, 19, 178—204.