Hellinger Information and Optimal Design for Nonregular Models

Yi Lin [1] Ryan Martin [2] Min Yang [3]

[1 3] University of Illinois at Chicago [2] North Carolina State University
Contact Email: ylin46@uic.edu

Motivation

Define a measure of information for the purpose of optimal design of experiment for nonregular models, such as

- **Uniform Distribution** $p_\theta(y) = \frac{1}{\theta}\mathbf{1}_{[\mathbf{0},\theta]}(y)$,
- **Nonregular Regression** [Smi94]

  $y_i = g(\theta, \mathbf{x_i}) + \varepsilon_i, i = 1, ..., n, \varepsilon_i \sim P_0, \mathbf{x_i}$ is the vector of ith covariate
  P_0 : gamma, Weibull distributions with shape parameter less than 2, etc,...

for which Fisher information, $\mathcal{I}(\theta)$, does not exist.

Hellinger Distance and its Expansion

Distributions P_θ, $\theta \in \Theta \subseteq R^d, d \geq 1$, on Y having μ-densities $p_\theta = dP_\theta/d\mu$.

- **Squared Hellinger distance** $h(\theta, \vartheta)$,

$$h(\theta, \vartheta) \equiv H^2(P_\theta, P_\vartheta) := \int (p_\theta^{1/2} - p_\vartheta^{1/2})^2 \, d\mu, \theta, \vartheta \in \Theta.$$

- **Local Expansion of** $h(\theta, \vartheta)$

$$h(\theta, \theta + \epsilon u) = J(\theta, u)|\epsilon|^\alpha + o(|\epsilon|^\alpha), \quad \epsilon \to 0, \forall u : \|u\|_2 = 1, \alpha \in (0, 2].$$

  - **Hellinger Information at θ in the direction of u :** $J(\theta, u)$
  - **Hellinger Information at θ in the direction of u: for a sample of independent obserations of size n:** $\mathcal{J}_n(\theta, u) = \sum_{i=1}^n J_i(\theta, u)$
  - **Index of Regularity:** α

- **Regular models are *differentiable in quadratic mean (DQM) at θ*, for which** $\alpha = 2$ **and** $J(\theta, u) = \frac{1}{4}u^\top \mathcal{I}(\theta)u$ [VdV98]
- **Nonregular models refer to cases where** $\alpha < 2$
- **Reparametization Rule:** If there is a differentiable function $g(.) : \Theta \to H$, such that for every $\theta \in \Theta$, there is an $\eta = g(\theta) \in H$, and $p_\theta(y) = q_\eta(y)$, for all y, then, the regularity index of every $\theta_i, i = 1, .., d$ would be the same as regularity index for $\eta : \alpha_\eta$. Then,

$$J(\theta, u) = |\sum_{j=1}^d u_j \frac{\partial g(\theta)}{\partial \theta_j}|^{\alpha_\eta} \tilde{J}(\eta). \tag{1}$$

Let $\psi : \Theta \to R^q$ be a differentiable function with non-singular $q \times d$ derivative matrix $D_\psi(\theta)$, then **Hellinger information of $\psi(\theta)$ at θ in the direction** u is

$$J^\psi(\theta, u) = \left\|D_\psi(\theta)\, u\right\|_2^{-\alpha} J(\theta, u) \tag{2}$$

Examples

- $N(\theta, \sigma)$ is DQM, with regularity index $\alpha = 2$ and $J(\theta, u) = \frac{1}{4}u^\top \mathcal{I}(\theta)u$.
- Uniform distribution $p_\theta(y) = \frac{1}{\theta}\mathbf{1}_{[\mathbf{0},\theta]}(y)$, with $\alpha = 1$, and $J(\theta) = \theta^{-1}$
- Shifted-gamma distribution $p_\eta(z) = \frac{(z-\eta)^{\beta-1}}{\Gamma(\beta)}e^{-(z-\eta)}\mathbf{1}_{[\eta,\infty]}(z), \beta \in [1, 2)$, with regularity index $\alpha_\eta = \beta$ and $J(\eta) = \frac{1}{\beta\Gamma(\beta)}$.
- Nonregular Regression with gamma error : $y_i - g(\theta, x_i) \sim p_\eta$, (shifted-gamma), then by (1), $\alpha_\theta = \beta$ and Hellinger information at θ in the direction u of the sample is

$$J_n(\theta, u) = \sum_{i=1}^n |\sum_{j=1}^d u_j \frac{\partial g(\theta, \mathbf{x_i})}{\partial \theta_j}|^\beta \frac{1}{\beta\Gamma(\beta)}.$$

Hellinger Information Inequality

Theorem 1. *Let $T_n = T(\mathbf{Y^n})$ be any estimator of $\psi(\theta)$. Let $R_\psi(T_n, \theta) = E_\theta\|T_n - \psi(\theta)\|_2^2$ be the risk. If $\epsilon_{n,u} = \epsilon_{n,u,\theta} = \{3\mathcal{J}_n(\theta, u)\}^{-1/\alpha}$, and $\liminf_{n\to\infty} n^{-1}\mathcal{J}_n(\theta, u) > 0$. Let $B_n(\theta)$ be the region whose boundary is determined by the union of $\{\theta + \epsilon_{n,u}u\}$ over all directions u. Then, for all large n,*

$$\inf_{T_n} \sup_{\vartheta \in B_n(\theta)} R_\psi(T_n, \vartheta) \gtrsim \left\{\inf_u \left\|D_\psi(\theta)\, u\right\|_2^{-\alpha} \mathcal{J}_n(\theta, u)\right\}^{-2/\alpha} = \left\{\inf_u \mathcal{J}_n^\psi(\theta, u)\right\}^{-2/\alpha}. \tag{3}$$

Hellinger Information

Definition 1. *The Hellinger information of function $\psi(.)$ at θ based on the sample is defined as*

$$\mathcal{J}_n^\psi(\theta) = \min_{u:\|u\|_2=1} \mathcal{J}_n^\psi(\theta, u). \tag{4}$$

If $\psi(.)$ is identity function, then $\mathcal{J}_n(\theta)$ is denoted as the Hellinger information for θ.

Definition 2. *For nonregular regression models with $x_i \in [-A, A]$, and with design $\xi = \{(w_i, x_i), i = 1, ..., m\}$,* ***Hellinger information of θ based on design ξ is***

$$\mathcal{J}_\xi(\theta) \propto \min_{u:\|u\|_2=1} \sum_{i=1}^m w_i |\sum_{j=1}^d u_i \frac{\partial g(\theta, x_i)}{\partial \theta_j}|^\alpha.$$

Optimal Design Based on Hellinger Information

Optimal Design for $\psi(\theta)$:

$$\xi^* = \arg\max_\xi \mathcal{J}_\xi^\psi(\theta).$$

Optimal Design based on Hellinger information corresponds to E-optimal design under regular models:

$$\mathcal{J}_n^\psi(\theta) = \lambda_{min}\{\frac{1}{4}(D_\psi(\theta)^\top \mathcal{I}_n^{-1}(\theta)D_\psi(\theta))\}^{-1}.$$

Optimal Design for Nonregular Regression Models

Theorem 2. *For nonregular regression models with $g(\theta, x_i) = \theta_1 + \theta_2 x_i$, with $\alpha \in [1, 2)$*

$$\xi^* = \arg\max_{\xi\in\Xi} \min_{u:\|u\|_2=1} \sum_{i=1}^m w_i |u_1 + u_2 x_i|^\alpha = \{(0.5, -A), (0.5, A)\}.$$

The following table contains a simulation result comparing optimal design to 5-, 10-, and 15-point uniform designs. Standard error of mean square error value is recored in parentheses. Sample size is 120 for each run of experiment, and each experiment has 1000 repeats. The estimator used in the simulation is from [Smi94].

$MSE(\theta_0)$	$MSE(\theta_1)$	$Sum(MSE)$	Design
0.00404 (0.00013)	0.00022 (0.00005)	0.00426 (0.00013)	ξ_{opt}
0.00398 (0.00014)	0.00097 (0.00010)	0.00495 (0.00017)	ξ_{5pt}
0.00408 (0.00014)	0.00105 (0.00007)	0.00513 (0.00016)	ξ_{10pt}
0.00376 (0.00013)	0.00124 (0.00010)	0.00500 (0.00016)	ξ_{15pt}
$Gamma(\beta = 1.4)$ distributed error, and $x_i \in [-2, 2]$			

Theorem 3. *Nonregular quadratic regression with $g(\theta, x_i) = \theta_0 + \theta_1 x_i + \theta_2 x_i^2$, with $\alpha = 1$ (i.e., with exponential distributed error)*

$$\xi^* = \arg\max_{\xi\in\Xi} \min_{u:\|u\|_2=1} \sum_{i=1}^m w_i |u_1 + u_2 x_i + u_3 x_i^2|^\alpha$$
$$\in \left\{\xi_w = \{(\frac{1-w}{2}, -A), (w, 0), (\frac{1-w}{2}, A)\}, w \in (0, 1)\right\}.$$

The exact value of w depends on A and can be searched numerically.

In the following simulation result table, the numerically searched optimal weight for ξ^* for each value of A is indicated by w^*.

Design	$A = 1, w^* = 0.5$	$A = 2, w^* = 0.65$	$A = 5, w^* = 0.8$
ξ^*	0.00229 (0.00009)	0.00071 (0.00003)	0.00039 (0.00003)
ξ_{5pt}	0.00484 (0.00020)	0.00176 (0.00011)	0.00107 (0.00007)
ξ_{10pt}	0.00486 (0.00022)	0.00110 (0.00007)	0.00084 (0.00006)
ξ_{15pt}	0.00567 (0.00028)	0.00122 (0.00006)	0.00084 (0.00006)

Further Work

- **Explore Hellinger information as Bayesian prior for multidimensional parameter.**
- **Finding optimal design for other nonregular regression models.**

References

[Smi94] Richard L Smith. Nonregular regression. *Biometrika*, 81(1):173–183, 1994.

[VdV98] Aad W Van der Vaart. *Asymptotic statistics*, volume 3. Cambridge university press, 1998.