Ranking Function Based Parameter Estimation

Benedek Bálint Novák

Supervisor:
Balázs Csanád Csáji

Eötvös Loránd Tudományegyetem

2025 June 6.

The framework

Given:

- A family of probability distributions $\{\mathbb{P}_\vartheta | \vartheta \in \Theta\}$ on the same standard Borel space $\mathcal{X}$ (Θ Polish space).
- $S^{(0)} = (x_1, ..., x_n)$ sample from $\mathbb{P}_{\vartheta^*}$ ($x_1, ..., x_n$ are not necessarily i.i.d.)
- Black box G that can generate new, i.i.d. samples $S_\vartheta^{(1)}, ..., S_\vartheta^{(m)}$ given parameter ϑ.

Black box properties:

- We assume that the seed can be fixed.
- Example for fixed seed: uniform distribution and inverse CDF.

Goal:

Approximate ϑ^*

The framework

The Resampling framework

- 1. Generate $m-1$ alternative samples $S^{(1)}, \ldots, S^{(m-1)}$ from $\mathbb{P}_\vartheta$.
- 2. Assign a real number to each sample based on ϑ and its values called *reference variable*: $Z_\vartheta^{(i)} := T(S_\vartheta^{(i)}, \vartheta)$ $(i = 0, \ldots, m-1)$.
- 3. Rank the samples based on the reference variables.
- 4. Denote the *rank* of the original sample. with $\mathcal{R}_\vartheta^{(m)} \in \{1, \ldots, m\}$

Theorem

$\mathbb{P}(\vartheta^* \in \{\vartheta \in \Theta | \mathcal{R}_\vartheta^{(m)} \le q\}) = \frac{q}{m}$ if there is a strict ordering a.s.

Remark

$\{Z_i\}_{i \neq 0}$ are i.i.d. random variables

Reference Variable

Examples of Reference Variables

- ML based reference variable: $Z_\vartheta^{(i)} = ||\nabla_\vartheta \mathcal{L}(\vartheta, S_\vartheta^{(i)})||^2$
- MMD based reference variable: $Z_\vartheta^{(i)} = \widehat{\mathrm{MMD}}^2[S_\vartheta^{(i)}, S_\vartheta^{(i+m)}]$

where $\{S_\vartheta^{(i+m)}\}$ denotes an extra sample for each of the m samples and $\widehat{\mathrm{MMD}}^2$ is an unbiased estimator for the Maximum Mean Discrepancy of the two probability distributions.

Remark

The MMD is a customisable similarity measure of probability distributions. Note that MMD based reference variable doesn't require any knowledge about the distributions besides the samples.

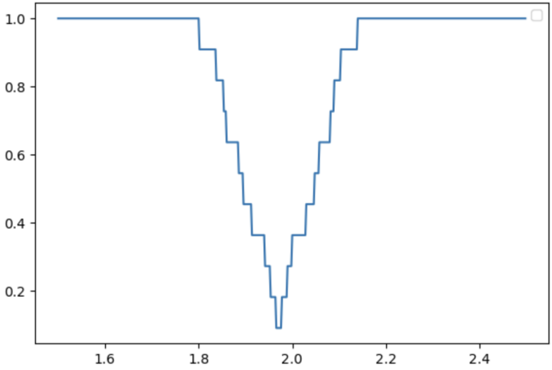
Parameter Estimation

Idea:

$$\widehat{\vartheta} \in \operatorname*{argmin}_{\vartheta \in \Theta} \mathcal{R}_{\vartheta}^{(m)}$$

Problem:

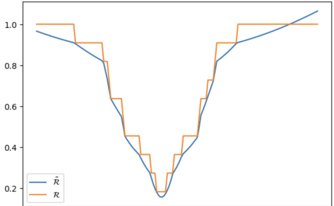
Hard to optimize



Smoothed rank

$$\tilde{\mathcal{R}}^{(m)}_{\vartheta,\xi}(z) = \begin{cases} \frac{1}{m}\left(\frac{z}{Y^{(1)}_{\vartheta}}\right) & \text{if } z < Y^{(1)}_{\vartheta} \\ \frac{1}{m}\left(k + \frac{z - Y^{(k)}_{\vartheta}}{Y^{(k+1)}_{\vartheta} - Y^{(k)}_{\vartheta}}\right) & \text{if } Y^{(k)}_{\vartheta} \leq z < Y^{(k+1)}_{\vartheta} \\ \frac{1}{m}\left(m - 1 + \tau\left(z, Y^{(m-1)}_{\vartheta}\right)\right) & \text{if } Y^{(m-1)}_{\vartheta} \leq z \end{cases}$$

Where $\{Y^{(i)}_{\vartheta}\}$ denotes the ordered version of $\{Z^{(i)}_{\vartheta}\}$ and τ is a continuous function with $\tau(z, y) \geq 0$ and $\tau(z, z) = 0$ assuming $z \geq y$, monotonically increasing in z and decreased in y.



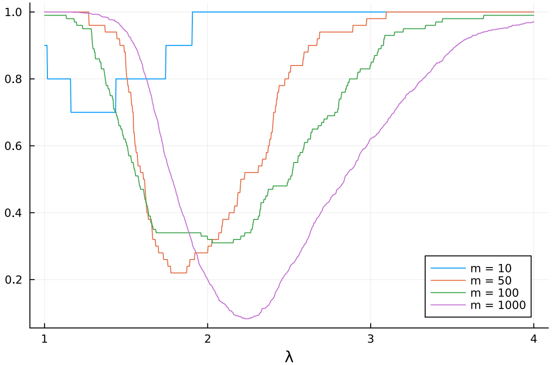
Asymptotic behaviour: $m \to \infty$

Proposition

$\lim_{m\to\infty} \mathcal{R}_\vartheta^{(m)} = F_{Z_\vartheta}(Z_\vartheta^{(0)})$ where F_{Z_ϑ} denotes the CDF of $Z_\vartheta^{(i)}$ for every $i \neq 0$

Proposition

$\lim_{m\to\infty} \tilde{\mathcal{R}}_\vartheta^{(m)} = F_{Z_\vartheta}(Z_\vartheta^{(0)})$



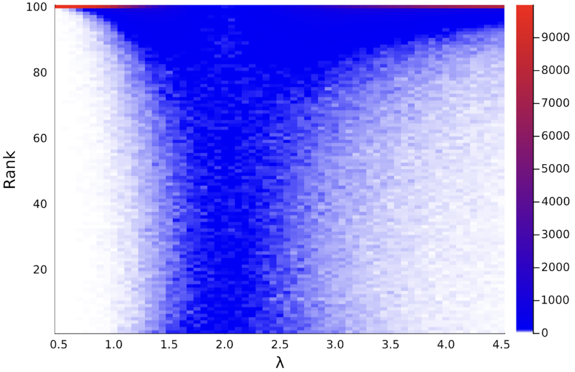
Asymptotic behaviour: $m \to \infty$

Proposition

$\lim_{m\to\infty} \mathcal{R}_\vartheta^{(m)} = F_{Z_\vartheta}(Z_\vartheta^{(0)})$ where F_{Z_ϑ} denotes the CDF of $Z_\vartheta^{(i)}$ for every $i \neq 0$

Proposition

$\lim_{m\to\infty} \tilde{\mathcal{R}}_\vartheta^{(m)} = F_{Z_\vartheta}(Z_\vartheta^{(0)})$



Asymptotic behaviour: $n \to \infty$

Assumption: $S^{(0)} = (x_1, ..., x_n)$ contains i.i.d. instances from $\mathbb{P}_\vartheta$.

Definition

We say that a reference variable is *consistent*, if it holds that

$$\lim_{n\to\infty} Z_\vartheta^{(i)} = \begin{cases} 0 & \text{if } x_j \sim \mathbb{P}_\vartheta \text{ i.i.d.} \\ c \in \mathbb{R}_+ \cup \{\infty\} & \text{else} \end{cases} \tag{1}$$

almost surely for any $\vartheta \in \Theta$ parameter and $i = 0, ..., m$.

Proposition

If $Z_\vartheta^{(i)}$ are *consistent*, then the relative rank $\mathcal{R}_\vartheta^{(m)}$ constructed from it has the following properties:

- I.) $\tilde{\mathcal{R}}_\vartheta^{(m)} \to 1$ a.s. as $n \to \infty$ if $\mathbb{P}_{\vartheta^*} \neq \mathbb{P}_\vartheta$.
- II.) $\tilde{\mathcal{R}}_\vartheta^{(m)} \xrightarrow{d} U_m[0,1]$ as $n \to \infty$ if $\mathbb{P}_{\vartheta^*} = \mathbb{P}_\vartheta$ where $U_m[0,1]$ denotes the discrete uniform distribution over the set $\left\{\frac{1}{m}, ..., \frac{m-1}{m}, 1\right\}$.

Optimization

Definition

Simultaneous Perturbation Stochastic Approximation (SPSA): for finding a minimum of $F : \mathbb{R}^d \to \mathbb{R}$:

$$\vartheta_{n+1,k} = \vartheta_{n,k} + \gamma_n \frac{F(\vartheta_n - \delta_n \Delta_n) - F(\vartheta_n + \delta_n \Delta_n)}{2\delta_n \Delta_{n,k}} \qquad (2)$$

where $\vartheta_{n,k}$ denotes the kth coordinate of ϑ_n, $\{\gamma_n\}$ and δ_n are learning rate hyperparameters, and $\{\Delta_n\}$ are independent, symmetric, zero-mean vectors, for example Bernoulli trials with $\Delta_{n,k} = \pm 1$ with probability $\frac{1}{2}$ each.

Optimization

Different optimizers for a sample from a normal distribution