Instructions for Paper Submissions to AISTATS 2023: Supplementary Materials

1 THORETICAL DERIVATIONS

1.1 Proof that $\hat{\sigma}^2(z_1, z_2)$ is an unbiased estimator of $\sigma_*^2(z_1, z_2)$

In Section .. we use that $\hat{\sigma}^2(z_1, z_2)$ is an unbiased estimator of $\sigma_*^2(z_1, z_2) = \frac{\int_{z_1}^{z_2} \mathbb{E}_{X_c|X_s=z}\left[(f^s(z, X_c) - \mu(z_1, z_2))^2\right]\partial z}{z_2 - z_1}$, under the assumption that the points are uniformly distributed in $[z_1, z_2)$, i.e., $x_s^i \sim \mathcal{U}(z_1, z_2)$.

Proof We have made the hypothesis that the points $\mathbf{x}^i$ are coming from the distibution $\mathcal{S} = p(X_c|X_s)p(X_s) = p(X_c|X_s)\mathcal{U}(x_s; z_1, z_2) = \frac{p(\mathbf{x_c}|z)}{z_2 - z_1}$ uniformly distributed in the interval $[z_1, z_2)$ in terms

$$\mathbb{E}_{\mathcal{S}}[\hat{\sigma}^2(z_1, z_2)] = \mathbb{E}_{\mathcal{S}}[\frac{1}{|\mathcal{S}|}\sum_{i=1}^{|\mathcal{S}|}(f^s(\mathbf{x}^i) - \mu(z_1, z_2))^2] \tag{1}$$

$$= \frac{1}{|\mathcal{S}|}\sum_{i=1}^{|\mathcal{S}|}\mathbb{E}_{\mathcal{S}}[(f^s(\mathbf{x}^i) - \mu(z_1, z_2))^2] \tag{2}$$

$$= \frac{1}{|\mathcal{S}|}\sum_{i=1}^{|\mathcal{S}|}\mathbb{E}_{\mathcal{S}}[(f^s(\mathbf{x}^i) - \mu(z_1, z_2))^2] \tag{3}$$

$$\tag{4}$$

$$\sigma_*^2(z_1, z_2) = \frac{\int_{z_1}^{z_2} \mathbb{E}_{X_c|X_s=z}\left[(f^s(z, X_c) - \mu(z_1, z_2))^2\right]\partial z}{z_2 - z_1} \tag{5}$$

$$= \mathbb{E}_{z\sim\mathcal{U}(z_1,z_2)}\mathbb{E}_{X_c|X_s=z}\left[(f^s(z, X_c) - \mu(z_1, z_2))^2\right] \tag{6}$$

$$= \mathbb{E}_X\left[(f^s(X) - \mu(z_1, z_2))^2\right] \tag{7}$$

1.2 Proof Of Theorem 3.1

If we define (a) the residual $\rho(z)$ as the difference between the expected effect at z and the bin-effect, i.e $\rho(z) = \mu(z) - \mu(z_1, z_2)$ and (b) $\mathcal{E}(z_1, z_2)$ as the mean squared residual of the bin, i.e. $\mathcal{E}(z_1, z_2) = \frac{\int_{z_1}^{z_2}\rho^2(z)\partial z}{z_2 - z_1}$, then it holds that:

$$\sigma_*^2(z_1, z_2) = \sigma^2(z_1, z_2) + \mathcal{E}^2(z_1, z_2) \tag{8}$$

Proof

$$\sigma_*^2(z_1, z_2) = \frac{1}{z_2 - z_1} \int_{z_1}^{z_2} \mathbb{E}_{X_c|z} \left[(f^s(z, X_c) - \mu(z_1, z_2))^2 \right] \partial z \tag{9}$$

$$= \frac{1}{z_2 - z_1} \int_{z_1}^{z_2} \mathbb{E}_{X_c|z} \left[(f^s(z, X_c) - \mu(z) + \rho(z))^2 \right] \partial z \tag{10}$$

$$= \frac{1}{z_2 - z_1} \int_{z_1}^{z_2} \mathbb{E}_{X_c|z} \left[(f^s(z, X_c) - \mu(z))^2 + \rho(z)^2 + 2f^s(z, X_c)\mu(z)) \right] \tag{11}$$

$$= \frac{1}{z_2 - z_1} \int_{z_1}^{z_2} \left(\underbrace{\mathbb{E}_{X_c|z} \left[(f^s(z, X_c) - \mu(z))^2 \right]}_{\sigma^2(z)} + \underbrace{\mathbb{E}_{X_c|z} \left[\rho^2(z) \right]}_{\rho^2(z)} + 2(\underbrace{\mathbb{E}_{X_c|z} \left[(f^s(z, X_c) \right]}_{\mu(z)} - \mu(z))\rho(z)) \right) \partial z \tag{12}$$

$$= \underbrace{\frac{1}{z_2 - z_1} \int_{z_1}^{z_2} \sigma^2(z) \partial z}_{\sigma^2(z_1, z_2)} + \underbrace{\frac{1}{z_2 - z_1} \int_{z_1}^{z_2} \rho^2(z) \partial z}_{\mathcal{E}^2(z_1, z_2)} = \sigma^2(z_1, z_2) + \mathcal{E}^2(z_1, z_2) \tag{13}$$

1.3 Proof Of Corollary

If a bin-splitting $\mathcal{Z}$ minimizes the accumulated error, then it also minimizes $\sum_{k=1}^K \sigma_*^2(z_1, z_2)\Delta z_k$

We want to show that

$$\mathcal{Z}^* = \arg\min_{\mathcal{Z}} \sum_{k=1}^K \sigma_*^2(z_{k-1}, z_k)\Delta z_k \Leftrightarrow \mathcal{Z}^* = \arg\min_{\mathcal{Z}} \sum_{k=1}^K \mathcal{E}^2(z_{k-1}, z_k)\Delta z_k$$

Proof

$$\mathcal{Z}^* = \arg\min_{\mathcal{Z}} \sum_{k=1}^K \sigma_*^2(z_{k-1}, z_k)\Delta z_k \tag{14}$$

$$= \arg\min_{\mathcal{Z}} \left[\sum_{k=1}^K (\sigma^2(z_{k-1}, z_k) + \mathcal{E}^2(z_{k-1}, z_k))\Delta z_k \right] \tag{15}$$

$$= \arg\min_{\mathcal{Z}} \left[\sum_{k=1}^K \left(\frac{\Delta z_k}{\Delta z_k} \int_{z_{k-1}}^{z_k} \sigma^2(z)\partial z) + \mathcal{E}^2(z_{k-1}, z_k)\Delta z_k \right) \right] \tag{16}$$

$$= \arg\min_{\mathcal{Z}} \left[\underbrace{\int_{z_0}^{z_K} \sigma^2(z)\partial z}_{\text{independent of } \mathcal{Z}} + \sum_{k=1}^K \mathcal{E}^2(z_{k-1}, z_k)\Delta z_k) \right] \tag{17}$$

$$= \arg\min_{\mathcal{Z}} \sum_{k=1}^K \mathcal{E}^2(z_{k-1}, z_k)\Delta z_k \tag{18}$$

2 Dynamic Programming

For achieving a computationally-grounded solution we set a threshold K_{max} on the maximum number of bins which also discretizes the solution space. The width of the bin can take discrete values that are multiple of the minimum step $u = \frac{x_{s,max} - x_{s,min}}{K_{max}}$. For defining the solution, we use two indexes. The index $i \in \{0, \ldots, K_{max}\}$ denotes the point (z_i) and the index $j \in \{0, \ldots, K_{max}\}$ denotes the position of the j-th multiple of the minimum step, i.e., $x_j = x_{s,min} + j \cdot u$. The recursive cost function $T(i, j)$ is the cost of setting $z_i = x_j$:

$$\mathcal{T}(i, j) = \min_{l \in \{0, \ldots, K_{max}\}} \left[\mathcal{T}(i-1, l) + \mathcal{B}(x_l, x_j) \right] \tag{19}$$

where $\mathcal{T}(0, j)$ equals zero if $j = 0$ and ∞ in any other case. $\mathcal{B}(x_l, x_j)$ denotes the cost of creating a bin with limits $[x_l, x_j)]$:

$$\mathcal{B}(x_l, x_j) = \begin{cases} \infty, & \text{if } x_j > x_l \text{ or } |\mathcal{S}_{(x_j, x_l)}| < N \\ 0, & \text{if } x_j = x_l \\ \hat{\sigma}^2(x_j, x_l), & \text{if } x_j \leq x_l \end{cases} \tag{20}$$

The optimal solution is given by solving $\mathcal{L} = \mathcal{T}(K_{max}, K_{max})$ and keeping track of the sequence of steps.

3 EXPERIMENTS APPEARED AT THE MAIN PAPER

3.1 Simulation 1

3.2 Simulation 2

3.3 Real World Example

4 Further Experimentation