Supporting Information

S2 Appendix. Multilevel ABC asymptotic performance analysis. This appendix contains a detailed derivation of the asymptotic performance characteristics of the multilevel approximate Bayesian computation method (MLABC). This appendix follows closely the work of Giles et al. [1], but extends the results to multivariate CDFs.

Definitions and starting assumptions

Consider the probability space $(\Omega, \mathcal{F}, \mathbb{P})$, where $\Omega = [0,1]^\kappa$ for $\kappa \in \mathbb{N}$ is the sample space, $\mathcal{F}$ is the Borel σ-field on Ω and $\mathbb{P} : \mathcal{F} \to [0,1]$ is the probability measure.

For $K, L \in \mathbb{N}$, $K > 1$ and $\epsilon_0 \in \mathbb{R}^+$, let $\{\theta_\ell\}_{0 \le \ell \le L}$ be a sequence of random vectors, $\theta_\ell : \Omega \to \mathbb{R}^\kappa$ such that $\theta_\ell \sim p\left(\theta_\ell \,|\, \rho(\mathcal{D}_s, \mathcal{D}) < \epsilon_\ell\right)$ where $\epsilon_\ell = \epsilon_0 K^{-\ell}$. Similarly, let θ be a random vector, $\theta : \Omega \to \mathbb{R}^\kappa$ such that $\theta \sim p\left(\theta \,|\, \mathcal{D}\right)$. That is, θ is the exact posterior random variate and θ_ℓ are successive approximations to it generated using an ABC scheme.

Let $F(\mathbf{s}) : \mathbb{R}^\kappa \to [0,1]$ be the distribution function of θ, that is,

$$F(\mathbf{s}) = \mathbb{P}\left(\{\omega \in \Omega : \theta(\omega) \in A_\mathbf{s}\}\right), \tag{1}$$

where $A_\mathbf{s} = (-\infty, s_1] \times (-\infty, s_2] \times \ldots \times (-\infty, s_\kappa]$. We note that Equation (1) can be expressed as

$$F(\mathbf{s}) = \mathbb{E}\left[\mathbb{1}_{A_\mathbf{s}}(\theta)\right]. \tag{2}$$

Let $\mathbb{S} = [s_1^l, s_1^u] \times [s_2^l, s_2^u] \times \cdots \times [s_\kappa^l, s_\kappa^u] \subset \mathbb{R}^\kappa$ be the support of $p\left(\theta \,|\, \mathcal{D}\right)$ which is assumed to be compact.

As noted by Giles et al. [1], the discontinuity of the indicator functional causes problems for MLMC. We require a function $f : \mathbb{R}^\kappa \to [0,1]$ such that:

- S1 $\exists\, c \in \mathbb{R}, \forall x \in \mathbb{R}^\kappa$, $cost(f(x)) \le c$;
- S2 f is Lipschitz continuous, that is $\exists\, C \in \mathbb{R}$ such that $\forall x, y \in \mathbb{R}^\kappa$, $|f(x) - f(y)| \le C\,|x - y|_\infty$;

S3 $\forall x \in \mathbb{R}^\kappa$, $f(x) = 1$ if $(x_1 < -1) \wedge (x_2 < -1) \wedge \cdots \wedge (x_\kappa < -1)$ and $f(x) = 0$ if $(x_1 > 1) \vee (x_2 > 1) \vee \cdots \vee (x_\kappa > 1)$;

S4 $\forall x \in \mathbb{R}^\kappa$, $\int_{-1}^{1} x_i^j \left(\mathbb{1}_{A_\mathbf{0}}(x) - f(x)\right) \mathrm{d}x_i = 0$ for $j \in [0, 1, \ldots, r-1]$;

S5 $\forall x \in \mathbb{R}^\kappa$, $\int_{-\infty}^{-1} x_i^r \left(\mathbb{1}_{A_\mathbf{0}}(x) - f(x)\right) \mathrm{d}x_i < \infty$.

Now, let $g_\mathbf{s}(\theta) = f(\mathrm{diag}([\delta_1, \delta_2, \ldots, \delta_\kappa]^T)^{-1}(\mathbf{s} - \theta))$. Assuming that $p\left(\theta \,|\, \mathcal{D}\right)$ is r times continuously differentiable over the region $\mathbb{S}_\delta = [s_1^l - \delta_1, s_1^u + \delta_1] \times [s_2^l - \delta_2, s_2^u + \delta_2] \times \cdots \times [s_\kappa^l - \delta_\kappa, s_\kappa^u + \delta_\kappa]$, we arrive at the multivariate version of Lemma 2.2 of Giles et al. [1].

Lemma 1 *There exists a constant $c > 0$ such that $\forall\ \mathbf{d} = [\delta_1, \delta_2, \ldots, \delta_\kappa] \in (0, \delta_0]^\kappa$,*

$$\sup_{\mathbf{s} \in \mathbb{S}} \left| F(\mathbf{s}) - \mathbb{E}\left[g_\mathbf{s}(\theta)\right] \right| \leq c \prod_{i=1}^{\kappa} \delta_i^{r+1}.$$

Proof: First we denote the posterior density as $p\left(\theta\right)$ for convenience of notation. By Equation (2), we have

$$\begin{aligned} F(\mathbf{s}) - \mathbb{E}\left[g_\mathbf{s}(\theta)\right] &= \mathbb{E}\left[\mathbb{1}_{A_\mathbf{s}}(\theta)\right] - \mathbb{E}\left[g_\mathbf{s}(\theta)\right] \\ &= \mathbb{E}\left[\mathbb{1}_{A_\mathbf{s}}(\theta) - g_\mathbf{s}(\theta)\right] \\ &= \int_{\mathbb{R}^\kappa} \left(\mathbb{1}_{A_\mathbf{s}}(\theta) - g_\mathbf{s}(\theta)\right) p\left(\theta\right) \mathrm{d}\theta. \end{aligned} \tag{3}$$

Let $\phi = \mathrm{diag}(\mathbf{d})^{-1}(\mathbf{s} - \theta)$ and substitute into Equation (3),

$$F(\mathbf{s}) - \mathbb{E}\left[g_\mathbf{s}(\theta)\right] = \prod_{i=1}^{\kappa} \delta_i \int_{\mathbb{R}^\kappa} \left(\mathbb{1}_{A_\mathbf{0}}(\phi) - f(\phi)\right) p\left(\mathbf{s} + \mathrm{diag}(\mathbf{d})\phi\right) \mathrm{d}\phi. \tag{4}$$

Given the probability density, p, is r times continuously differentiable, we have by Taylor expansion

$$\begin{aligned} p\left(\mathbf{s} + \mathrm{diag}(\mathbf{d})\phi\right) = &\sum_{j=0}^{r-1} \sum_{\mathbf{a} \in N_j} \frac{\prod_{i=1}^{\kappa} (\delta_i \phi_i)^{a_i}}{\prod_{i=1}^{\kappa} a_i!} \left(\frac{\partial^{|\mathbf{a}|_1} p}{\partial\theta_1^{a_1} \partial\theta_2^{a_2} \ldots \partial\theta_\kappa^{a_\kappa}} \right)(\mathbf{s}) \\ &+ \sum_{\mathbf{b} \in N_r} \frac{\prod_{i=1}^{\kappa} (\delta_i \phi_i)^{b_i}}{\prod_{i=1}^{\kappa} b_i!} R_{r-1}\left(\mathbf{s} + \mathrm{diag}(\mathbf{d})\phi\right), \end{aligned} \tag{5}$$

where $N_j = \{\mathbf{a} \in \mathbb{N}^\kappa : |\mathbf{a}|_1 = j\}$ and R_{r-1} is the remainder term, which we note is continuous over $\mathbb{S}$. Substitution of Equation (5) into Equation (4) and application of properties S3 and S4 results in

$$F(\mathbf{s}) - \mathbb{E}\left[g_\mathbf{s}(\theta)\right] = \frac{\prod_{i=1}^{\kappa} \delta_i^{r+1}}{\kappa(r!)} \int_M \prod_{i=1}^{\kappa} \phi_i^r \left(\mathbb{1}_{A_\mathbf{0}}(\phi) - f(\phi)\right) R_{r-1}\left(\mathbf{s} + \mathrm{diag}(\mathbf{d})\phi\right) \mathrm{d}\phi, \tag{6}$$

where $M = \{\mathbf{x} \in \mathbb{R}^\kappa : (\forall i \in [1,\kappa], x_i \leq 1) \wedge (\exists i \in [1,\kappa], x_i \geq -1)\}$. Given properties S5, S2 and continuity of the remainder, we have

$$\sup_{s \in \mathbb{S}} \left| \int_M \prod_{i=1}^{\kappa} \phi_i^r \left(\mathbb{1}_{A_\mathbf{0}}(\phi) - f(\phi) \right) R_{r-1} \left(\mathbf{s} + \mathrm{diag}(\mathbf{d})\phi \right) \mathrm{d}\phi \right| < \infty. \tag{7}$$

The final result follows from Equation (6) and Equation (7). □

Lemma 1 provides a bound on the bias that is introduced by the smoothing if the indicator functional. The introduction of the Lipschitz continuous function $g_\mathbf{s}$ is essential to our analysis of the multilevel estimator. We now suppose the following convergence assumptions of the sequence $\{\theta_\ell\}$ for some $\alpha > 0$,$\beta > 0$ and $\gamma > 0$:

- A1 $\exists\ c \in \mathbb{R}^+, \forall\ \ell \in \mathbb{N}, \sup_{s \in \mathbb{S}} \left| \mathbb{E}\left[g_\mathbf{s}(\theta) \right] - \mathbb{E}\left[g_\mathbf{s}(\theta_\ell) \right] \right| \leq c\epsilon_\ell^\alpha$. That is, $g_\mathbf{s}(\theta_\ell)$ converges weakly to $g_\mathbf{s}(\theta)$ with order α;
- A2 $\exists\ c \in \mathbb{R}^+, \forall\ \ell \in \mathbb{N}, \mathbb{E}\left[\min\left(\left| \mathrm{diag}(\mathbf{d})^{-1}\mathbf{1} \right|_1 \left| \theta_\ell - \theta \right|_1, 1 \right) \right] \leq c\epsilon_\ell^\beta$. That is, θ_ℓ converges strongly to θ with order β. Here $|\cdot|_1$ is the ℓ_1-norm and $\mathbf{1} = [1, 1, \ldots, 1]^\mathrm{T} \in \mathbb{R}^\kappa$.

Next, we assume the following condition on the computational cost of sampling the joint distribution of θ_ℓ and $\theta_{\ell-1}$:

- A3 $\exists\ c \in \mathbb{R}^+, \forall\ \ell \in \mathbb{N}, cost\left(\theta_\ell, \theta_{\ell-1} \right) \leq c\epsilon_\ell^{-\gamma}$.

The following Lemma is crucial to our analysis of the multilevel estimator (equivalent to Lemma 2.4 of Giles et al. [1]). The proof relies upon the Lipschitz continuity of $g_\mathbf{s}$.

Lemma 2 *There exists a constant $c \in \mathbb{R}^+$ such that, for all $\ell \in \mathbb{N}$,*

$$\mathbb{E}\left[\sup_{\mathbf{s} \in \mathbb{S}} \left| g_\mathbf{s}(\theta) - g_\mathbf{s}(\theta_\ell) \right|^2 \right] \leq cK^{-\beta\ell}.$$

Proof: Firstly, by the Lipschitz continuity and boundedness of $g_\mathbf{s}$ implies there exists some constant C such that

$$\mathbb{E}\left[\sup_{\mathbf{s} \in \mathbb{S}} \left| g_\mathbf{s}(\theta) - g_\mathbf{s}(\theta_\ell) \right|^2 \right] \leq C\mathbb{E}\left[\min\left(\left| \mathrm{diag}(\mathbf{d})^{-1}(\theta_\ell - \theta) \right|_\infty^2, 1 \right) \right].$$

By the Cauchy-Schwartz inequality we have

$$\mathbb{E}\left[\min\left(\left|\mathrm{diag}(\mathbf{d})^{-1}(\theta_\ell-\theta)\right|_\infty^2,1\right)\right] \le \mathbb{E}\left[\min\left(\left|\mathrm{diag}(\mathbf{d})^{-1}\mathbf{1}\right|_\infty^2 \left|\theta_\ell-\theta\right|_\infty^2,1\right)\right].$$

If $|\mathrm{diag}(\mathbf{d})^{-1}\mathbf{1}|_\infty^2 \, |\theta_\ell-\theta|_\infty^2 \le 1$, then $|\theta_\ell-\theta|_\infty \le |\mathrm{diag}(\mathbf{d})^{-1}\mathbf{1}|_\infty^{-1}$. It follows that

$$\begin{aligned}\mathbb{E}\left[\min\left|\mathrm{diag}(\mathbf{d})^{-1}\mathbf{1}\right|_\infty^2 \left|\theta_\ell-\theta\right|_\infty^2,1)\right] &\le \mathbb{E}\left[\min\left(\left|\mathrm{diag}(\mathbf{d})^{-1}\mathbf{1}\right|_\infty \left|\theta_\ell-\theta\right|_\infty,1\right)\right]\\ &\le \mathbb{E}\left[\min\left(\left|\mathrm{diag}(\mathbf{d})^{-1}\mathbf{1}\right|_1 \left|\theta_\ell-\theta\right|_1,1\right)\right].\end{aligned}$$

By Assumption A2 we arrive at the desired bound. □

Consider a discrete set of points $S \subset \mathbb{S}$ with $|S| = D^\kappa$ for some $D \in \mathbb{N}$. The task is to construct an estimator of the distribution function Equation (2) at points in S. That is, compute $\mathbb{E}[I(\theta)]$ where $I(\theta) = [g_{\mathbf{s}_1}(\theta), g_{\mathbf{s}_2}(\theta)\ldots, g_{\mathbf{s}_{D^\kappa}}(\theta)]$ for $\mathbf{s}_1, \mathbf{s}_2, \ldots, \mathbf{s}_{D^\kappa} \in S$. The standard biased Monte Carlo estimator is

$$\mu = \frac{1}{n}\sum_{i=1}^{n} I(\theta_L^{(i)}), \tag{8}$$

where $\theta_L^{(i)} \sim p\left(\theta_\ell \,|\, \rho(\mathcal{D}_s,\mathcal{D}) < \epsilon_L\right)$ is generated via ABC rejection. The biased MLMC estimator is given by

$$\mu_{ML} = \frac{1}{n_0}\sum_{i=1}^{n_0} I(\theta_0^{(i)}) + \sum_{\ell=1}^{L}\frac{1}{n_\ell}\sum_{i=1}^{n_\ell}\left[I(\theta_\ell^{(i)}) - I(\theta_{\ell-1}^{(i)})\right]. \tag{9}$$

Now assume that we have a sequence of linear maps $P : \mathbb{S}^{D^\kappa} \times \mathbb{R}^{D^\kappa} \to C^r(\mathbb{S},\mathbb{R})$ that extends a set of point approximations to $F(\mathbf{s})$ to an r times differentiable function over the support $\mathbb{S}$. We assume the following properties of P:

C1 $\exists\, c \in \mathbb{R}^+, \forall\, D \in \mathbb{N}, \forall\, y \in \mathbb{R}^{D^\kappa}, \forall\, x \in \mathbb{S}^{D^\kappa},\ cost(P(x,y)) \le cD^\kappa$;

C2 $\exists\, c \in \mathbb{R}^+, \forall\, D \in \mathbb{N}, \forall\, y \in \mathbb{R}^{D^\kappa}, \forall\, x \in \mathbb{S}^{D^\kappa}, \|P(x,y)\|_\infty \le c\,|y|_\infty$;

C3 $\exists\, c \in \mathbb{R}^+, \forall\, D \in \mathbb{N}, \forall\, x \in \mathbb{S}^{D^\kappa}, \|F - P(x,F(x))\|_\infty \le cD^{-\kappa(r+1)}$.

As noted by Giles et al. [1], a piece-wise polynomial extension can be constructed to achieve these constraints. In particular, a cubic smoothing spline can achieve these with $r = 3$ and an appropriately chosen smoothing parameter. Based on P we have the following definition for the root mean-squared error for an estimator μ taken at points in S;

$$\text{RMSE}(\mu) = \mathbb{E}\left[\|F - P(S,\mu)\|_\infty^2\right]^{\frac{1}{2}}. \tag{10}$$

Given a sequence of posterior CDF estimators $\{\mu_n\}_{n\geq 0}$, such that $\text{RMSE}(\mu_n) \to 0$ as $n \to \infty$, the sequence converges with order (ζ, η) if

$$\text{cost}(\mu_n) \leq c\text{RMSE}(\mu_n)^{-\zeta}\left(-\log\left(\text{RMSE}(\mu_n)\right)\right)^{\eta}. \tag{11}$$

In the next two sections, we will derive the order of convergence for the standard ABC estimator (Equation (8)) and the MLABC estimator (Equation (9)).

Standard estimator convergence

Theorem 1 *The standard Monte Carlo estimator μ converges with order $(\zeta, \eta) = (2 + \gamma/\alpha, 1)$.*

Proof: First, we define the notation $x \preccurlyeq y$ to mean $x \leq cy$ where c is a constant that does not depend on D, $\mathbf{d}$ or L. We consider the RMSE for the standard Monte Carlo estimator,

$$\text{RMSE}(\mu) = \mathbb{E}\left[\|F - P(S,\mu)\|_\infty^2\right]^{\frac{1}{2}}.$$

After squaring both sides we have

$$\begin{aligned}
\text{RMSE}(\mu)^2 &= \mathbb{E}\left[\|F - P(S,\mu)\|_\infty^2\right], \\
&= \mathbb{E}\left[\|F - P(S, F(S) - F(S) + \mu)\|_\infty^2\right] \\
&= \mathbb{E}\left[\|F - P(S, F(S)) + P(S, F(S) - \mu)\|_\infty^2\right] \text{ (linearity of } P) \\
&\preccurlyeq \mathbb{E}\left[\|F - P(S, F(S))\|_\infty^2 + \|P(S, F(S) - \mu)\|_\infty^2\right] \text{ (Jensen's Inequality)} \\
&= \|F - P(S, F(S))\|_\infty^2 + \mathbb{E}\left[\|P(S, F(S) - \mu)\|_\infty^2\right] \text{ (linearity of expectation)} \\
&\preccurlyeq D^{-2\kappa(r+1)} + \mathbb{E}\left[\|P(S, F(S) - \mu)\|_\infty^2\right] \text{ (Assumption C3)} \\
&\preccurlyeq D^{-2\kappa(r+1)} + \mathbb{E}\left[|F(S) - \mu|_\infty^2\right] \text{ (Assumption C2).}
\end{aligned} \tag{12}$$

For the second term in Equation (12), we note

$$
\begin{aligned}
\mathbb{E}\left[|F(S)-\mu|_\infty^2\right] &= \mathbb{E}\left[|F(S)+(\mathbb{E}[\mu]-\mathbb{E}[\mu])-\mu\|_\infty^2\right] \\
&\le \mathbb{E}\left[2\,|F(S)-\mathbb{E}[\mu]|_\infty^2 + 2\,|\mathbb{E}[\mu]-\mu|_\infty^2\right] \text{ (Jensen's Inequality)} \\
&= 2\mathbb{E}\left[|F(S)-\mathbb{E}[\mu]|_\infty^2\right] + 2\mathbb{E}\left[|\mu-\mathbb{E}[\mu]|_\infty^2\right] \text{ (linearity of expectation)} \\
&= 2\left(|F(S)-\mathbb{E}[\mu]|_\infty^2 + \mathrm{Var}[\mu]\right) \text{ (definition of Var}[\mu]\text{)}. \qquad (13)
\end{aligned}
$$

Substituting Equation (13) into Equation (12) yields

$$
\mathrm{RMSE}(\mu)^2 \preccurlyeq D^{-2\kappa(r+1)} + |F(S)-\mathbb{E}[\mu]|_\infty^2 + \mathrm{Var}[\mu]\,. \qquad (14)
$$

The variance term in Equation (14) can be further refined using the Bienaymé formula for random vectors (see Lemma 1 in Heinrich [2]),

$$
\mathrm{RMSE}(\mu)^2 \preccurlyeq D^{-2\kappa(r+1)} + |F(S)-\mathbb{E}[\mu]|_\infty^2 + (\log_e D)\,\frac{\mathrm{Var}[I(\theta_L)]}{n}. \qquad (15)
$$

Finally, by Assumption A1 and Lemma 1, we have

$$
\mathrm{RMSE}(\mu)^2 \preccurlyeq D^{-2\kappa(r+1)} + \left(\prod_{i=1}^{\kappa} \delta_i^{(r+1)}\right)^2 + K^{-2\alpha L} + \frac{\log_e D}{n}.
$$

We desire $\mathrm{RMSE}(\mu) \preccurlyeq h$, that is,

$$
D^{-2\kappa(r+1)} + \left(\prod_{i=1}^{\kappa} \delta_i^{(r+1)}\right)^2 + K^{-2\alpha L} + \frac{\log_e D}{n} \preccurlyeq h^2, \qquad (16)
$$

which is satisfied for $D \ge h^{-1/\kappa(r+1)}$, $\delta_i \le h^{1/\kappa(r+1)}$, $L \ge (1/\alpha)\log_K h^{-1}$ and $n \ge h^{-2}\log_e D$.

We will now express the asymptotic computational cost in terms of the target RMSE. By assumptions A3 and C1 we have,

$$
\begin{aligned}
\mathrm{cost}(Q(S,\mu)) &\preccurlyeq D^{\kappa} + nK^{\gamma L} \\
&\preccurlyeq h^{-1/(r+1)} + h^{-2}(\log_e h^{-1})h^{-\gamma/\alpha} \\
&\preccurlyeq h^{-(2+\gamma/\alpha)}(\log_e h^{-1}). \qquad (17)
\end{aligned}
$$

By Equation (16) and Equation (17), the result follows. □

Multilevel estimator convergence

Theorem 2 *The multilevel Monte Carlo estimator* μ_{ML} *converges with order*

$$(\zeta, \eta) = \begin{cases} (2,1) & \text{if } \beta > \gamma, \\ (2,3) & \text{if } \beta = \gamma, \\ (2 + (\gamma - \beta)/\alpha, 1) & \text{if } \beta < \gamma. \end{cases}$$

.

Proof: First, we note that the analysis up to Equation (14) is general to both the standard and multilevel estimators. Application of the Bienaymé formula for the multilevel case yields

$$\begin{aligned} \text{RMSE}(\mu_{ML})^2 \preccurlyeq\ & D^{-2\kappa(r+1)} + |F(S) - \mathbb{E}\left[\mu_{ML}\right]|_\infty^2 \\ & + (\log_e D) \left[\frac{\text{Var}\left[I(\theta_0)\right]}{n_0} + \sum_{\ell=1}^{L} \frac{\text{Var}\left[I(\theta_\ell) - I(\theta_{\ell-1})\right]}{n_\ell} \right]. \end{aligned} \tag{18}$$

For the variance of the bias correction terms, note that

$$\begin{aligned} \text{Var}\left[I(\theta_\ell) - I(\theta_{\ell-1})\right] &= \mathbb{E}\left[|I(\theta_\ell) - I(\theta_{\ell-1}) - \mathbb{E}\left[I(\theta_\ell) - I(\theta_{\ell-1})\right]|_\infty^2 \right] \\ &\le 2\mathbb{E}\left[|I(\theta_\ell) - I(\theta_{\ell-1})|_\infty^2 + |\mathbb{E}\left[I(\theta_\ell) - I(\theta_{\ell-1})\right]|_\infty^2 \right] \\ &= 2\mathbb{E}\left[|I(\theta_\ell) - I(\theta_{\ell-1})|_\infty^2 \right] + 2\left|\mathbb{E}\left[I(\theta_\ell) - I(\theta_{\ell-1})\right]\right|_\infty^2 \\ &\le 4\mathbb{E}\left[|I(\theta_\ell) - I(\theta_{\ell-1})|_\infty^2 \right]. \end{aligned} \tag{19}$$

Substitution of Equation (19) into Equation (18) yields

$$\begin{aligned} \text{RMSE}(\mu_{ML})^2 \preccurlyeq\ & D^{-2\kappa(r+1)} + |F(S) - \mathbb{E}\left[\mu_{ML}\right]|_\infty^2 \\ & + (\log_e D) \left[\frac{1}{n_0} + \sum_{\ell=1}^{L} \frac{\mathbb{E}\left[|I(\theta_\ell) - I(\theta_{\ell-1})|_\infty^2 \right]}{n_\ell} \right]. \end{aligned}$$

By Lemma 2, we have,

$$\begin{aligned} \text{RMSE}(\mu_{ML})^2 &\preccurlyeq D^{-2\kappa(r+1)} + |F(S) - \mathbb{E}\left[\mu_{ML}\right]|_\infty^2 + (\log_e D) \left[\sum_{\ell=0}^{L} \frac{K^{-\beta\ell}}{n_\ell} \right] \\ &\preccurlyeq D^{-2\kappa(r+1)} + \left(\prod_{i=1}^{\kappa} \delta_i^{(r+1)} \right)^2 + K^{-2\alpha L} + (\log_e D) \sum_{\ell=0}^{L} \frac{K^{-\beta\ell}}{n_\ell}. \end{aligned}$$

The task now is to choose n_ℓ such that the computational cost is minimized while maintaining $\text{RMSE}(\mu_{ML}) \preccurlyeq h$. We denote $f(n_0, n_1, \ldots, n_L)$ as the computational cost of the estimator excluding the extension using P, and we denote $g(n_0, n_1, \ldots, n_L)$ as the portion of the error controllable by changing n_ℓ. That is,

$$f(n_0, n_1, \ldots, n_L) = \sum_{\ell=0}^{L} n_\ell K^{\gamma\ell},$$
$$g(n_0, n_1, \ldots, n_L) = (\log_e D) \sum_{\ell=0}^{L} \frac{K^{-\beta\ell}}{n_\ell}.$$

We apply the method of Lagrange multipliers to minimize f subject to $g = h^2$. The Lagrangian is

$$\mathcal{L}(n_0.n_1, \ldots, n_L, \lambda) = f(n_0, n_1, \ldots, n_L) + \lambda(g(n_0, n_1, \ldots, n_L) - h^2).$$

We proceed by solving $\nabla \mathcal{L} = \mathbf{0}$. First consider,

$$\frac{\partial \mathcal{L}}{\partial n_\ell} = 0,$$
$$K^{\gamma\ell} - \lambda n_\ell^{-2} K^{-\beta\ell} \log_e D = 0,$$
$$n_\ell^2 = \lambda K^{-(\gamma+\beta)\ell} \log_e D. \tag{20}$$

Solve for λ using

$$\frac{\partial \mathcal{L}}{\partial \lambda} = 0,$$
$$h^2 = (\log_e D) \sum_{\ell=0}^{L} \frac{K^{-\beta\ell}}{n_\ell},$$
$$h^2 = (\log_e D) \sum_{\ell=0}^{L} \left[\lambda K^{-(\gamma+\beta)\ell} \log_e D\right]^{-\frac{1}{2}} K^{-\beta\ell},$$
$$\sqrt{\lambda} = h^{-2} \sqrt{\log_e D} \sum_{\ell=0}^{L} K^{((\gamma-\beta)/2)\ell}. \tag{21}$$

Using Equation (20) and Equation (21) we obtain

$$n_\ell = h^{-2} (\log_e D) K^{-((\gamma+\beta)/2)\ell} \sum_{\ell=0}^{L} K^{((\gamma-\beta)/2)\ell}. \tag{22}$$

Now, we consider the full cost function of the multilevel estimator,

$$\text{cost}(P(S, \mu_{ML})) \preccurlyeq D^\kappa + \sum_{\ell=0}^{L} n_\ell K^{\gamma\ell}. \tag{23}$$

Substitution of Equation (22) into Equation (23) yields

$$\begin{aligned}
\text{cost}(P(S,\mu_{ML})) &\preccurlyeq D^{\kappa} + \sum_{\ell=0}^{L}\left[h^{-2}(\log_e D)K^{-((\gamma+\beta)/2)\ell}\sum_{m=0}^{L}K^{((\gamma-\beta)/2)m}\right]K^{\gamma\ell}, \\
&= D^{\kappa} + h^{-2}(\log_e D)\left[\sum_{\ell=0}^{L}K^{((\gamma-\beta)/2)\ell}\right]^2.
\end{aligned}$$

Using standard geometric series results, we have

$$\left[\sum_{\ell=0}^{L}K^{((\gamma-\beta)/2)\ell}\right]^2 \preccurlyeq \begin{cases} 1 & \text{if } \beta > \gamma, \\ L^2 & \text{if } \beta = \gamma, \\ K^{(\gamma-\beta)(L+1)} & \text{if } \beta < \gamma. \end{cases}$$

Therefore, for $D \geq h^{-1/\kappa(r+1)}$, $L \geq (1/\alpha)\log_K h^{-1}$, we have

$$\text{cost}(P(S,\mu_{ML})) \preccurlyeq \begin{cases} h^{-2}(\log_e h^{-1}) & \text{if } \beta > \gamma, \\ h^{-2}(\log_e h^{-1})^3 & \text{if } \beta = \gamma, \\ h^{-(2+(\gamma-\beta)/\alpha)}(\log_e h^{-1}) & \text{if } \beta < \gamma. \end{cases} \tag{24}$$

The result follows from Equation (24). □

References

1. Giles MB, Nagapetyan T, Ritter K. Multilevel Monte Carlo approximation of cumulative distribution function and probability densities. SIAM/ASA Journal on Uncertainty Quantification. 2015;3:267–295. doi:10.1137/140960086.

2. Heinrich S. Monte Carlo complexity of global solution of integral equations. Journal of Complexity. 1998;14(2):151–175. doi:10.1006/jcom.1998.0471.