# Rejection ABC with Linear Regression Adjustment

**Rejection ABC (top-$k$).** Given an observation $x_{\text{obs}}$, draw $\theta_i \sim p(\theta)$ and simulate $x_i \sim p(x \mid \theta_i)$ for $i = 1, \dots, N$. Let $\phi(\cdot)$ denote the summary/feature map used for comparison (in our implementation: flattening of simulator output, optionally followed by standardization). Define distances

$$d_i \;=\; \|\phi(x_i) - \phi(x_{\text{obs}})\|_2\,,$$

and retain the $k$ accepted pairs $\{(\theta_i, x_i)\}_{i \in \mathcal{A}}$ corresponding to the smallest $d_i$ (top-$k$ acceptance).

**Kernel weights (Epanechnikov).** Let $\varepsilon = \max_{i \in \mathcal{A}} d_i$ and define weights

$$w_i \;=\; \max\bigl(0,\; 1 - (d_i/\varepsilon)^2\bigr)\,, \qquad i \in \mathcal{A}.$$

**Feature vectors and (optional) standardization.** Let $Z_i = \phi(x_i) \in \mathbb{R}^p$ and $Z_{\text{obs}} = \phi(x_{\text{obs}}) \in \mathbb{R}^p$. Optionally, we standardize using the accepted set (and apply the same transform to $Z_{\text{obs}}$):

$$X_i = \frac{Z_i - \mu}{\sigma}, \qquad X_{\text{obs}} = \frac{Z_{\text{obs}} - \mu}{\sigma},$$

where $(\mu, \sigma)$ are computed from $\{Z_i\}_{i \in \mathcal{A}}$ (optionally using weights $w_i$).

**(Optional) parameter transform.** For constrained parameters we may apply a bijective transform $t$ (e.g. log / logit):

$$\tilde{\theta}_i = t(\theta_i),$$

and map back after adjustment via $t^{-1}$.

**Local linear regression (LRA).** Define centered predictors

$$\Delta X_i \;=\; X_i - X_{\text{obs}}, \qquad i \in \mathcal{A}.$$

For each parameter component $j = 1, \dots, d_\theta$, fit a weighted linear model with intercept:

$$\tilde{\theta}_{i,j} \approx \alpha_j + \beta_j^\top \Delta X_i, \quad \text{by minimizing} \quad \sum_{i \in \mathcal{A}} w_i \left(\tilde{\theta}_{i,j} - \alpha_j - \beta_j^\top \Delta X_i\right)^2.$$

(Equivalently, one may fit a single multi-output linear regression for all components at once.)

**Regression adjustment step.** Let $\widehat{B}$ collect the fitted coefficients (stacked across $j$). Then each accepted draw is adjusted as

$$\tilde{\theta}_i^{\text{adj}} \;=\; \tilde{\theta}_i - \widehat{B}\,\Delta X_i \;=\; \tilde{\theta}_i - \widehat{B}\,(X_i - X_{\text{obs}}), \qquad i \in \mathcal{A}.$$

This is equivalent to the common "predicted shift" form $\tilde{\theta}_{i,j}^{\text{adj}} = \tilde{\theta}_{i,j} + \widehat{f}_j(X_{\text{obs}}) - \widehat{f}_j(X_i)$.

**Return adjusted draws and resample.** Finally, map back if needed: $\theta_i^{\mathrm{adj}} = t^{-1}(\tilde{\theta}_i^{\mathrm{adj}})$. We approximate the posterior by KDE-resampling from $\{\theta_i^{\mathrm{adj}}\}_{i\in\mathcal{A}}$, choosing the KDE bandwidth by cross-validation over a positive bandwidth grid.