Approximate Bayesian Computation: a non-parametric perspective

Asymptotic bias and variance for ABC

The ZYX of ABC

Michael G.B. Blum

Laboratoire TIMC, CNRS, Grenoble

ABC in Paris, June 2009

Framework

- Two estimators of the posterior distribution
  - Smooth rejection
  - With correction adjustment
- Bias and variance of these different estimators
- Numerical comparisons

Smooth rejection

- Simulate n values θ_i, $i = 1, \ldots, n$ from the prior π
- Simulate n (possibly multivariate) summary statistics $\mathbf{s_i}$ according to $p(\mathbf{s_i}|\theta_i)$
- Consider the weighted sample (θ_i, W_i), $i = 1, \ldots, n$

$$W_i = K(B^{-1}(\mathbf{s}_i - \mathbf{s}_{obs}))$$

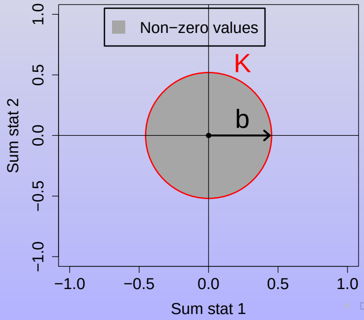
where K is a multivariate kernel and B is a bandwidth matrix.

Smooth rejection

We assume

$$B = bD,$$

where D is a diagonal matrix accounting for the different scales of the summary statistics.
The bandwidth b is assumed to be the same in each direction.



Smooth rejection

Estimator of the posterior

$$\hat{g}_0(\theta|\mathbf{s}_{obs}) = \frac{1}{b'}\sum_{i=1}^{n}\tilde{K}_{b'}(\theta_i-\theta)\frac{W_i}{\sum_{i=1}^{n}W_i},$$

where $\tilde{K}$ denotes an univariate kernel and b' the corresponding bandwidth.

Correction adjustment

- A model of local regression fitted by weighted least squares

$$\theta_i|\mathbf{s}_i = m(\mathbf{s}_i) + \epsilon_i$$

- Adjustment

$$\theta_i^* = \hat{m}(\mathbf{s}_{obs}) + \tilde{\epsilon}_i,$$

where $\tilde{\epsilon}_i$ denotes the empirical residuals

$$\tilde{\epsilon}_i = \theta_i - \hat{m}(\mathbf{s}_i).$$

In the following, we assume that $\hat{m}$ is a linear function (Beaumont et al. 2002) or a quadratic function (B. 2009).
The conditional probability of the residuals is denoted $h(\epsilon|\mathbf{s})$.

Correction adjustment

Estimator of the posterior (Hyndman et al. 1996, Beaumont et al. 2002, Hansen 2004)

$$\hat{g}_j(\theta|\mathbf{s}_{obs}) = \frac{1}{b'} \sum_{i=1}^{n} \tilde{K}_{b'}(\theta_i^* - \theta) \frac{W_i}{\sum_{i=1}^{n} W_i}, \quad j = 1, 2$$

Linear adjustment $j = 1$
Quadratic adjustment $j = 2$

Main theorem B. 2009

Asymptotic bias of $\hat{g}_j(\theta|\mathbf{s}_{obs})$, $j = 0, 1, 2$

$$C_1 b'^2 + C_{2,j} b^2$$

Asymptotic variance of $\hat{g}_j(\theta|\mathbf{s}_{obs})$

$$\frac{C_3}{np(\mathbf{s}_{obs}) b^d b'}$$

where d is the dimension of the summary statistics and n is the number of simulations.

Conseq 1 : The curse of dimensionality

Mean square errors are minimized when

$$b, b' = O(n^{-1/(d+5)})$$

$$\text{Minimals MSE} = O(n^{-4/(d+5)}).$$

The rate at which the minimal MSEs converges to 0 decreases importantly as the dimension d of $\mathbf{s}_{obs}$ increases.

Conseq 2 : Comparison between the estimators with and without adjustment

Asymptotic bias of $\hat{g}_0(\theta|\mathbf{s}_{obs})$ (smooth rejection)

$$\mu_2(K)b^2\left\{\frac{g_{\mathbf{s}}(\theta|\mathbf{s})^t_{|\mathbf{s}=\mathbf{s}_{obs}} D^2 p_{\mathbf{s}}(\mathbf{s}_{obs})}{p(\mathbf{s}_{obs})} + \frac{\mathrm{tr}(D^2 g_{\mathbf{ss}}(\theta|\mathbf{s})_{|\mathbf{s}=\mathbf{s}_{obs}})}{2}\right\}$$

Asymptotic bias of $\hat{g}_1(\theta|\mathbf{s}_{obs})$ (linear correction)

$$\mu_2(K)b^2\left\{\frac{h_{\mathbf{s}}(\epsilon|\mathbf{s})^t_{|\mathbf{s}=\mathbf{s}_{obs}} D^2 p_{\mathbf{s}}(\mathbf{s}_{obs})}{p(\mathbf{s}_{obs})} + \frac{\mathrm{tr}(D^2 h_{\mathbf{ss}}(\epsilon|\mathbf{s})_{|\mathbf{s}=\mathbf{s}_{obs}})}{2} - \frac{h_{\epsilon}(\epsilon|\mathbf{s}_{obs})\mathrm{tr}(D^2 m_{\mathbf{ss}}(\mathbf{s}_{obs}))}{2}\right\}$$

Conseq 2 : Comparison between the two estimators with adjustment

Asymptotic bias of $\hat{g}_1(\theta|\mathbf{s}_{obs})$ (linear correction)

$$\mu_2(K)b^2\left\{\frac{h_{\mathbf{s}}(\epsilon|\mathbf{s})^t_{|\mathbf{s}=\mathbf{s}_{obs}} D^2 p_{\mathbf{s}}(\mathbf{s}_{obs})}{p(\mathbf{s}_{obs})} + \frac{\mathrm{tr}(D^2 h_{\mathbf{ss}}(\epsilon|\mathbf{s})_{|\mathbf{s}=\mathbf{s}_{obs}})}{2} - \frac{h_\epsilon(\epsilon|\mathbf{s}_{obs})\mathrm{tr}(D^2 m_{\mathbf{ss}}(\mathbf{s}_{obs}))}{2}\right\}$$

Asymptotic bias of $\hat{g}_2(\theta|\mathbf{s}_{obs})$ (quadratic correction)

$$\mu_2(K)b^2\left\{\frac{h_{\mathbf{s}}(\epsilon|\mathbf{s})^t_{|\mathbf{s}=\mathbf{s}_{obs}} D^2 p_{\mathbf{s}}(\mathbf{s}_{obs})}{p(\mathbf{s}_{obs})} + \frac{\mathrm{tr}(D^2 h_{\mathbf{ss}}(\epsilon|\mathbf{s})_{|\mathbf{s}=\mathbf{s}_{obs}})}{2}\right\}$$

Conseq 2 : Comparison between the two estimators with adjustment

When the model

$$\theta_i = m(\mathbf{s}_i) + \epsilon_i$$

is homoscedastic in the vicinity of $\mathbf{s}_{obs}$, the bias for the estimator with quadratic adjustment is

$$o(b^2) + C_1 b'^2,$$

Box-Cox transformations (Box and Cox, 1964) are important to make the model as homoscedastic as possible.

How many simulations are required to reach a given level of accuracy in ABC

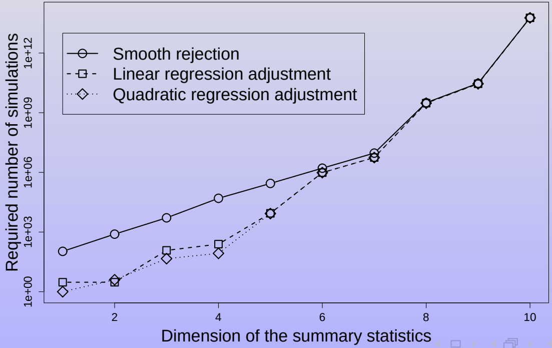
- A standard Gaussian model $(x_1, \ldots, x_d) \rightsquigarrow \mathcal{N}((\mu_1, \ldots, \mu_d), I_d)$.
- Given a sample of $M = 10$ individuals, we can compute the asymptotic mean square error (MSE) arising from the estimation of the partial posterior distribution of e^{μ_1} at 1 $\hat{g}(e^{\mu_1} = 1 | (\bar{x}_1, \ldots, \bar{x}_d))$

$$\mathrm{MSE}(n) = \mathrm{bias}^2 + \mathrm{variance}$$

- How many simulations are required so that the relative mean square error is less than 10%

How many simulations are required to reach a given level of accuracy ... continued

The curse of dimensionality



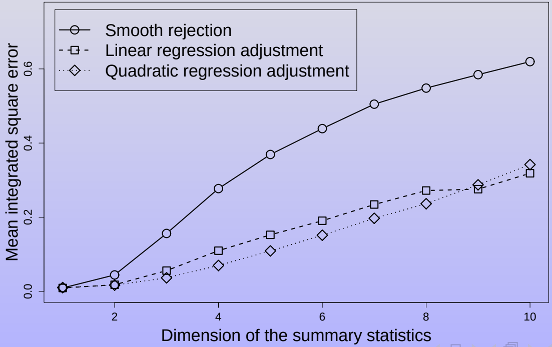
The integrated mean square error as a function of the dimension of the summary statistics d

Monte-Carlo approximation, for $j = 0, 1, 2$, of

$$\text{MISE} = E\left[\int_{e^{\mu_1} \in \mathbb{R}^+} \{g(e^{\mu_1}|(\bar{x}_1, \ldots, \bar{x}_d)) - \hat{g}_j(e^{\mu_1}|(\bar{x}_1, \ldots, \bar{x}_d))\}^2 \, de^{\mu_1}\right].$$

The integrated mean square error as a function of the dimension of the summary statistics d

The curse of dimensionality...continued



Perspectives

The variance is inversely proportional to

$$p(\mathbf{s}_{obs}) = \int_\theta p(\mathbf{s}_{obs}|\theta)\pi(\theta)\, d\theta.$$

Reduction of the variance

- Dimension reduction : PCA (Leuenberger et al. 2009), Neural networks (B. and François 2009)
- Adapting the sampling distribution (Sisson et al. 2007, Toni et al. 2009, Beaumont et al. 2009)