An application of the feed-forward neural network in Approximate Bayesian Computation

Miranda Fix

Statistical Machine Learning

December 4, 2014

The Basic ABC Algorithm

Goal: posterior distribution of a multidimensional parameter of interest ϕ

1. Calculate observed value $\mathbf{s}$ of a set of summary statistics $\mathbf{S}$
2. Choose a tolerance δ
3. Simulate parameter ϕ_i from the prior distribution
4. Simulate data using ϕ_i
5. Calculate $\mathbf{s_i}$ from the simulated data
6. If $||\mathbf{s_i} - \mathbf{s}|| < \delta$, accept ϕ_i. Otherwise reject.

Repeat this M times. The accepted ϕ_i form the simulated posterior.

The Basic ABC Algorithm

We are approximating the posterior distribution

$$p(\phi|\mathbf{s}) \propto p(\mathbf{s}|\phi)p(\phi)$$

with

$$p_\delta(\phi|\mathbf{s}) \propto p(||\mathbf{s} - \mathbf{s_i}|| < \delta|\phi)p(\phi)$$

As $\delta \to \infty$, the approximate posterior distribution returns the prior distribution. If the summary statistics are sufficient with respect to ϕ, then as $\delta \to 0$ the approximate posterior distribution will converge to the true posterior distribution.

Sources of Error

- Non-zero tolerance δ
- Non-sufficient summary statistics
- Curse of dimensionality (high-dimensional $\mathbf{s}$)

Beaumont et al. 2002 - LocL model

Proposed improvement to achieve insensitivity of the approximation to δ, thus permitting an increasing number of summary statistics.

The parameters ϕ_i are weighted by the values $K_\delta(||\mathbf{s_i} - \mathbf{s}||)$, where K_δ is the Epanechnikov kernel,

$$K_\delta(t) = \begin{cases} c\delta^{-1}(1 - (t/\delta)^2 & \text{if } t \leq \delta \\ 0 & \text{if } t > \delta \end{cases}$$

Beaumont et al. 2002 - LocL model

The posterior density is described as a local-linear model of the form

$$\phi_i = \alpha + (\mathbf{s_i} - \mathbf{s})^\top \beta + \zeta_i, \qquad \zeta_i \sim (0, \sigma^2)$$

Find (α, β) by minimizing the weighted least-squares criterion

$$\sum_{i=1}^{M} \left\{ \phi_i - (\alpha + (\mathbf{s_i} - \mathbf{s})^\top \beta) \right\}^2 K_\delta(||\mathbf{s_i} - \mathbf{s}||)$$

Calculate $\phi_i^* = \phi_i - (\mathbf{s_i} - \mathbf{s})^\top \hat{\beta}$. Weighted by $K_\delta(||\mathbf{s_i} - \mathbf{s}||)$, the ϕ_i^*'s provide an approximate sample from the posterior distribution.

Blum and Francois 2010 - NCH model

To minimize departures from linearity and homoscedasticity, model both the location and scale of the response parameter using a nonlinear conditional heteroscedastic (NCH) model:

$$\phi_i = m(\mathbf{s_i}) + \sigma(\mathbf{s_i}) \times \zeta_i$$

where $m(\mathbf{s_i})$ is the conditional expectation $\mathrm{E}[\phi|\mathbf{S} = \mathbf{s_i}]$,
and $\sigma^2(\mathbf{s_i})$ is the conditional variance $\mathrm{Var}[\phi|\mathbf{S} = \mathbf{s_i}]$.

Estimate the conditional expectation $\hat{m}(\mathbf{s_i})$ by fitting a flexible nonlinear regression model. Then estimate the variance term using a second regression model: $\log(\phi_i - \hat{m}(\mathbf{s_i}))^2 = \log \sigma^2(\mathbf{s_i}) + \xi_i$.

Blum and Francois 2010 - NCH model

The authors use feed-forward neural network (FFNN) regression models to fit both $m(\mathbf{s_i})$ and $\log \sigma^2(\mathbf{s_i})$ because FFNNs have the potential to reduce the dimensionality of the set of summary statistics via internal projection on lower dimensional subspaces.

Blum and Francois 2010 - NCH model

Linear combinations of the inputs (summary statistics) are transformed to compute the values z_j at the H hidden units:

$$z_j = h(\sum_{k=1}^{D} w_{jk}^{(1)} \mathbf{s}^k + w_{j0}^{(1)}), \quad j = 1, \ldots, H$$

Since H is typically smaller than D, this reduces the initial dimension of the vector of summary statistics. The values of the hidden units are then linearly combined to produce the output of the FFNN:

$$g_{\mathbf{w}}(\mathbf{s}) = \sum_{j=1}^{H} (w_j^{(2)} z_j + w_0^{(2)})$$

Blum and Francois 2010 - NCH model

The weights **w** of the first FFNN are found by minimizing the regularized least-squares criterion

$$\sum_{i=1}^{M} \{\phi_i - g_{\mathbf{w}}(\mathbf{s})\}^2 K_\delta(||\mathbf{s_i} - \mathbf{s}||) + \lambda||\mathbf{w}||^2$$

The weights of the second FFNN estimating the conditional variance are found by minimizing

$$\sum_{i=1}^{M} \left\{\log(\phi_i - \hat{m}(\mathbf{s_i}))^2 - g_{\mathbf{w'}}(\mathbf{s})\right\}^2 K_\delta(||\mathbf{s_i} - \mathbf{s}||) + \lambda||\mathbf{w'}||^2$$

Blum and Francois 2010 - NCH model

Lastly, perform parameter adjustment:

$$\phi_i^* = \hat{m}(\mathbf{s}) + (\phi_i - \hat{m}(\mathbf{s_i})) \times \frac{\hat{\sigma}(\mathbf{s})}{\hat{\sigma}(\mathbf{s_i})}$$

Assuming the NCH model corresponds to the true relationship between ϕ_i and $\mathbf{s_i}$, then the ϕ_i^*'s form a random sample from the distribution $p(\phi|\mathbf{s})$ provided that $\hat{m}$ could be considered equal to m and $\hat{\sigma}$ equal to σ.
As before, a tolerance error δ is allowed, and the adjusted parameters ϕ_i^* are weighted by $K_\delta(||\mathbf{s_i} - \mathbf{s}||)$.

References

Beaumont, M.A., W. Zhang and D.J. Balding (2002)
Approximate Bayesian Computation in Population Genetics
Genetics 162: 2025-2035

Blum, M.G.B. and O. Francois (2010)
Non-linear regression models for Approximate Bayesian Computation
Stat Comput 20: 63-73