Hamiltonian ABC

UNIVERSITY OF AMSTERDAM

Edward Meeds and Robert Leenders and Max Welling

Introduction

- Scientists express hypotheses through complex and expensive simulators.
- Posterior distributions give insight into models for both understanding underlying phenomena and improving hypotheses.
- Approximate Bayesian Computation provides a Bayesian framework for posterior analysis, but is very inefficient.
- This work uses a surrogate of the simulator to speed-up ABC.
- Based on the *Metropolis-Hastings Error*, our algorithms determine when a simulation is necessary and provides the user with a "knob" ξ to control it.

Approximate Bayesian Computation

- ABC is a *likelihood-free* method because $\pi(\mathbf{y}|\boldsymbol{\theta})$ is either not computable or very expensive:

$$\pi_\epsilon(\boldsymbol{\theta}|\mathbf{y}) = \frac{\pi(\boldsymbol{\theta})}{\pi(\mathbf{y})}\int \pi_\epsilon(\mathbf{y}|\mathbf{x})\pi(\mathbf{x}|\boldsymbol{\theta})d\mathbf{x}$$

- Kernel functions $\pi_\epsilon(\mathbf{y}|\mathbf{x})$ are proxies for the likelihood, based on draws $\mathbf{x} \overset{\text{sim}}{\sim} \pi(\mathbf{x}|\boldsymbol{\theta})$ from simulator.
- Rejection sampling with ϵ-tube kernel is very inefficient.
- We use ABC MCMC, which approximates the likelihood by Monte Carlo approximation:

images/exponential_proble

ϵ trades-off precision and mixing

$$\pi_\epsilon(\mathbf{y}|\boldsymbol{\theta}') \approx \frac{1}{S}\sum_{s=1}^{S}\pi_\epsilon(\mathbf{y}|\mathbf{x}^{(s)},\boldsymbol{\theta}')$$

$\epsilon = 0.5$	$\epsilon = 0.01$
images/abc_mcmc...	images/abc_mcmc_eps_0p01.png

We accept the proposed parameter $\boldsymbol{\theta}'$ with probability equal to $\alpha(\boldsymbol{\theta}'|\boldsymbol{\theta}) =$

$$\min\left(1, \frac{\pi(\boldsymbol{\theta}')\sum_s \pi_\epsilon(\mathbf{y}|\mathbf{x}'^{(s)},\boldsymbol{\theta}')q(\boldsymbol{\theta}|\boldsymbol{\theta}')}{\pi(\boldsymbol{\theta})\sum_s \pi_\epsilon(\mathbf{y}|\mathbf{x}^{(s)},\boldsymbol{\theta})q(\boldsymbol{\theta}'|\boldsymbol{\theta})}\right)$$

The Synthetic Likelihood

- Introduced by Wood (2010), replace the Monte Carlo approximation with a Gaussian with estimators based on the pseudo-data $\{\mathbf{x}_1, .., \mathbf{x}_S\}$ simulated at $\boldsymbol{\theta}$:

$$\hat{\boldsymbol{\mu}}_{\boldsymbol{\theta}} = \frac{1}{S}\sum_s \mathbf{x}^{(s)} \quad \hat{\boldsymbol{\Sigma}}_{\boldsymbol{\theta}} = \frac{1}{S-1}\sum_s\left(\mathbf{x}^{(s)} - \hat{\boldsymbol{\mu}}_{\boldsymbol{\theta}}\right)\left(\mathbf{x}^{(s)} - \hat{\boldsymbol{\mu}}_{\boldsymbol{\theta}}\right)^T$$

- With a Gaussian kernel:

$$\pi_\epsilon(\mathbf{y}|\mathbf{x}) = K_\epsilon(\mathbf{y},\mathbf{x}) = \frac{1}{(2\pi\epsilon)^{J/2}}e^{-\frac{1}{2\epsilon^2}(\mathbf{x}-\mathbf{y})^T(\mathbf{x}-\mathbf{y})}$$

- The synthetic likelihood can then be computed analytically:

$$\pi(\mathbf{y}|\boldsymbol{\theta}) = \int K_\epsilon(\mathbf{y},\mathbf{x})\mathcal{N}\left(\hat{\boldsymbol{\mu}}_{\boldsymbol{\theta}}, \hat{\boldsymbol{\Sigma}}_{\boldsymbol{\theta}}\right)d\mathbf{x} = \mathcal{N}\left(\hat{\boldsymbol{\mu}}_{\boldsymbol{\theta}}, \hat{\boldsymbol{\Sigma}}_{\boldsymbol{\theta}} + \epsilon^2\boldsymbol{I}\right)$$

MCMC with a Random Acceptance Probability

- The randomness of $\hat{\boldsymbol{\mu}}_{\boldsymbol{\theta}}$ (by re-running S simulations), induces a distribution over the acceptance probabilities.
- Approximate randomness in $\boldsymbol{\mu}_{\boldsymbol{\theta}}$ by drawing M times $\boldsymbol{\mu}_{\boldsymbol{\theta}}^{(m)} \sim \mathcal{N}\left(\hat{\boldsymbol{\mu}}_{\boldsymbol{\theta}}, \hat{\boldsymbol{\Sigma}}_{\boldsymbol{\theta}}/S\right)$
- Estimate $p(\alpha)$ using Monte Carlo approximation based on these M samples:

$p(\alpha)$ and $F(\alpha)$

images/acceptance_distribution_with_fold.png

$$\alpha^{(m)} = \min\left(1, \frac{\pi(\boldsymbol{\theta}')\mathcal{N}\left(\mathbf{y}|\boldsymbol{\mu}_{\boldsymbol{\theta}'}^{(m)}, \hat{\boldsymbol{\Sigma}}_{\boldsymbol{\theta}'} + \epsilon^2\boldsymbol{I}\right)q(\boldsymbol{\theta}|\boldsymbol{\theta}')}{\pi(\boldsymbol{\theta})\mathcal{N}\left(\mathbf{y}|\boldsymbol{\mu}_{\boldsymbol{\theta}}^{(m)}, \hat{\boldsymbol{\Sigma}}_{\boldsymbol{\theta}} + \epsilon^2\boldsymbol{I}\right)q(\boldsymbol{\theta}'|\boldsymbol{\theta})}\right)$$

- Based on $p(\alpha)$, compute the probability of making a M-H Error. This is the area under the folded CDF and is minimized at the median of $p(\alpha)$, denoted τ.

Probability of acceptance error	**Probability of rejection error**
$P(\alpha < u) = \frac{1}{M}\sum_m\left[\alpha^{(m)} < u\right]$	$P(\alpha > u) = \frac{1}{M}\sum_m\left[\alpha^{(m)} \geq u\right]$

- Total *conditional* error: $\mathcal{E}_u(\alpha) = [u \leq \tau]P(\alpha < u) + [u > \tau]P(\alpha \geq u)$
- MHE: $\mathcal{E}(\alpha) = \int \mathcal{E}_u(\alpha)\mathcal{U}(0,1)du$
- The crux of the MH step is to run simulations while $\mathcal{E}(\alpha) > \xi$.

Gaussian Process Surrogate ABC

- Adaptive SL is wasteful for expensive simulations: all results are discarded.
- GPS-ABC follow directly from synthetic likelihood ABC with randomized acceptance.
- Gaussian processes provides uncertainty estimates of the marginal likelihood informing us of the need to conduct additional experiments in order to make confident accept/reject decisions.
- For each statistic j, the surrogate provides the following conditional predictive distribution of the expected value of statistic j based on N training points $\mu_{\boldsymbol{\theta}j} \sim \mathcal{N}\left(\bar{\mu}_{\boldsymbol{\theta}j}, \sigma^2_{\boldsymbol{\theta}j}\right)$

$$\bar{\mu}_{\boldsymbol{\theta}j} = \boldsymbol{k}_{\boldsymbol{\theta}\Theta j}\left[\boldsymbol{K}_{\Theta\Theta j} + \sigma_j^2\boldsymbol{I}\right]^{-1}\mathbf{X}[:,j]$$

$$\sigma^2_{\boldsymbol{\theta}j} = k_{\boldsymbol{\theta}\boldsymbol{\theta}j} - \boldsymbol{k}_{\boldsymbol{\theta}\Theta j}\left[\boldsymbol{K}_{\Theta\Theta j} + \sigma_j^2\boldsymbol{I}\right]^{-1}\boldsymbol{k}_{\boldsymbol{\theta}\Theta j}$$

- Adjusting ξ affects precision and computation:

$\xi = 0.4$, $N = 9$	$\xi = 0.2$, $N = 184$	$\xi = 0.05$, $N = 1297$
high model uncertainty.	model uncertainty drops with ξ	high precision, many simulations
images/toy_gp_s0_5_e_0p40_T_10000_gp_zoom...	images/toy_gp_s0_20_e_0p20_T_10000_gp_zoom_n_184.png	images/toy_gp_s0_20_e_0p05_T_10000_gp_zoom_n_1297.png

- The key advantage of GPS-ABC is that with increasing frequency we will not have to do any expensive simulations whatsoever during a MH step because the GP surrogate is sufficiently confident about the statistics' surface in that region of parameter space.

Toy problem: Inferring parameters of an exponential distribution

- GPS-ABC learns the surface, eventually eliminating any new simulations:

Error per sample	Error per simulation	Simulations per sample
images/exp_paper...	images/exp_paper...	images/exp_paper..._v_bot...

Chaotic Ecological Systems

images/blowfly_probl...

- Adult blowfly populations exhibit dynamic, sometimes chaotic, behavior for which several competing population models exist.
- We use observational data and a simulation model from Wood (2010).
- Population dynamics are modeled using (discretized) differential equations that can produce chaotic behavior for some parameter settings.
- The population dynamics equation generates $N_1, \ldots, N_T$: $N_{t+1} = PN_{t-\tau}\exp(-N_{t-\tau}/N_0)e_t + N_t\exp(-\delta\epsilon_t)$

images/blowfly_theta...

- Injected noise: $e_t \sim \mathcal{G}(1/\sigma_p^2, 1/\sigma_p^2)$, $\epsilon_t \sim \mathcal{G}(1/\sigma_d^2, 1/\sigma_d^2)$
- θ: $\{\log P, \log\delta, \log N_0, \log\sigma_d, \log\sigma_p, \tau\}$
- Statistics $\mathbf{y}$ ($J = 10$): the log of the mean of all 25% quantiles of $N/1000$, the mean of the 25% quantiles of the first-order differences of $N/1000$, and the maximal peaks of smoothed N, with 2 different thresholds.

images/blowfly_conv_...

- Models: rejection sampling ($\epsilon = 0.5$), SL ($S = 10$), GPS-ABC ($\xi = 0.3$)
- Results: generated observations (top), posterior samples $\boldsymbol{\theta}$ (middle), convergence to $\mathbf{y}^\star$ using posterior predictive $p(\mathbf{y}|\mathbf{y}^\star)$ (bottom).

Conclusions

- Adaptive ABC algorithms using Metropolis-Hastings Erro... trols the computational complexity of the inference
- GP surrogate models incorporate both model and pseud... uncertainty into MCMC
- Improvements: Hamiltonian dynamics on the GP surface ... dependent samples; covariance on the outputs; alternative ... gate models; acquisition functions.