APPROXIMATE BAYESIAN COMPUTATION FOR ARCHIMEDEAN COPULA MODELS

Roopali Singh

Department of Statistics
Pennsylvania State University

$$\pi(\theta|y) \propto p(y|\theta)\pi(\theta)$$

Problem: How to perform Bayesian inference when the likelihood function $p(y|\theta)$ is computationally intractable?

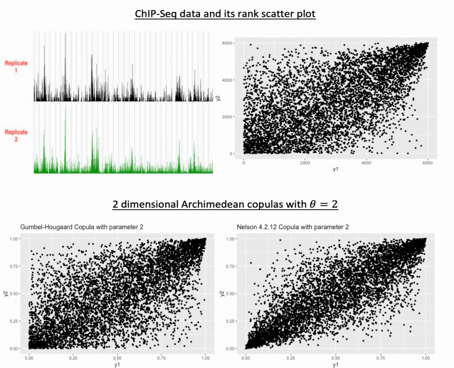
Solution: If we can easily simulate from the likelihood, ABC methods provide a possible way.

$$\pi_{ABC}(\theta|y) \propto \int \mathcal{I}(||y^* - y|| < \epsilon)\pi(\theta|y^*)dy^* \approx \pi(\theta|y)$$

MOTIVATION: ARCHIMEDEAN COPULA MODELS

J-dimensional copula: $C(u_1, \ldots, u_J) = P(U_1 \leq u_1, \ldots, U_J \leq u_J)$ where $U_1, \ldots, U_J$ are uniform random variables and $C : [0,1]^j \rightarrow [0,1]$.

J-dimensional Archimedean copula:
$\psi(C(u_1, \ldots, u_J); \theta) = \psi(u_1; \theta) + \ldots + \psi(u_J; \theta)$ where θ **is an association parameter** describing the strength of the dependence between U_j's, and ψ is a generator function specific to each Archimedean copula such that $\psi : [0,1] \rightarrow [0, \infty)$.



Gumbel Hougaard copula generator function:

$$\psi(u) = (-\log(u))^{\theta}\ ,\ \theta \in [1, \infty)$$

So its distribution function is,

$$C(u_1, ..., u_n) = \psi^{-1}(\psi(u_1) + ... + \psi(u_n)) = \exp\{-((-\log(u_1))^{\theta} + ... + (-\log(u_n))^{\theta})^{1/\theta}\}$$

Using ABC to estimate θ is possible because,

- likelihood is unavailable
- it easy to simulate from the model

Requirements:

- a proposal, $q(.)$
- the observed data, y
- a distance function, $||.||$
- a tolerance level, ϵ

Rejection Sampling (Sisson et al. 2018)

1. Simulate $\theta^* \sim q(.)$
2. Simulate $y \sim p(.|\theta^*)$
3. if $||y - y_{obs}|| < \epsilon$ then accept θ^* with probability $\frac{\pi(\theta^*)}{Kq(\theta^*)}$.
4. Repeat above steps N times.

MCMC (Marjoram et al. 2003)

1. Simulate $\theta^* \sim q(.)$
2. Simulate $y \sim p(.|\theta^*)$
3. if $||y - y_{obs}|| < \epsilon$ then accept θ^* with probability $\min\{1, \frac{\pi(\theta^*)q(\theta|\theta^*)}{\pi(\theta)q(\theta^*|\theta)}\}$ else stay at θ.
4. Repeat above steps N times.

- Should we use the entire dataset y or an appropriate summary statistic, s?
- Choice of distance function and summary statistic.
- Computationally expensive?
- Choice of tolerance level and other tuning parameters.

ABC MCMC: SIMULATION STUDY

- Prior on θ is **Gamma(shape=1, scale=2) truncated at 1**.
- Proposal is a random walk i.e. $N(\theta^{(i-1)}, \sigma^2)$ **truncated at 1**.
- Tolerance, ϵ and σ^2 are chosen such that the **acceptance rate was 20-35%**.
- **Absolute difference** is used as the distance function for two dimensions and for more than 2, **Frobenius norm** is used.
- For each simulation, y_{obs} consisted of 100 data points.
- Number of simulations = 100 and N=10000.

Dimensions:	2		3		5	
Summary statistic	Estimate	MSE	Estimate	MSE	Estimate	MSE
Spearman's rank ρ	3.06	1.32	2.94	1.12	2.19	0.07
Kendall's τ	2.60	0.47	2.55	0.39	2.16	0.05

Table: Results from ABC MCMC algorithm for Gumbel-Hougaard Copula with $\theta = 2$.

ABC MCMC WITH KERNEL FUNCTION

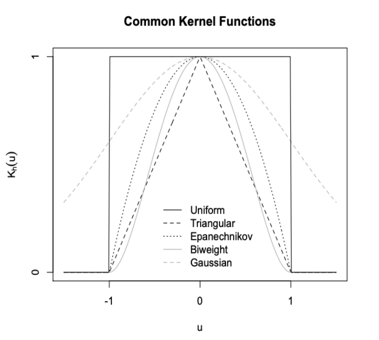
Kernel function: $K_\epsilon(u) = \frac{1}{\epsilon}K(\frac{u}{\epsilon})$ where $K(u) = \frac{1}{\sqrt{2\pi}}e^{-u^2/2}$

Previously: Accept θ^* with probability $\min\{1, \frac{\pi(\theta^*)q(\theta|\theta^*)}{\pi(\theta)q(\theta^*|\theta)}\}$.

Now: Accept θ^* with probability $\min\{1, \frac{K_\epsilon(||s^* - s_{obs}||)\pi(\theta^*)q(\theta|\theta^*)}{K_\epsilon(||s^{(i-1)} - s_{obs}||)\pi(\theta)q(\theta^*|\theta)}\}$.

Dimension	Estimate	MSE
2	2.24	0.12
3	2.14	0.09
5	2.18	0.10

Table: Results from ABC MCMC algorithm (using Kendall's τ) with kernel function for Gumbel Hougaard Copula with $\theta = 2$.



Note: Smoothing functions are discussed in Sisson et al. 2018 for ABC Rejection algorithms and something similar in Fernhead and Prangle 2011.

CONCLUSION

- For large data sets, using y_{obs} instead of summary statistic is not advisable due to computational costs.
- ABC MCMC requires a lot of tuning to run well. If possible, a different ABC algorithm can be preferred.
- Kendall's τ estimates better than Spearman's rank correlation.
- Estimation for higher dimensional copula works better because there is more information for a single parameter.
- ABC MCMC with kernel function seems like a better approach in this case, at least for lower dimensional copula.