Discussion of "Bayesian Restricted Likelihood and model diagnostics"

Radu Craiu

Department of Statistical Sciences
University of Toronto

Athens - June 9, 2025

Initial thoughts

- Goodness-of-fit / Model criticism tends to be more difficult than model choice.
- Bayesian analysis' vulnerability to model misspecification must front-load our concerns
- Big data require increasingly complex models, but perhaps of interest are only certain aspects: nuisance parameters, nuisance modules
- Some of these ideas were discussed at length (Lewis et al., 2021 - *Bayesian Analysis*, with discussion).
- My focus will be tinted by my own research interests: dependence modeling via copulas.

One step further: partial updates with dependence

- Consider a copula model

$$f(y_1, y_2|\theta_1, \theta_2, \eta)) = f_1(y_1|\theta_1)f_2(y_2|\theta_2)c(F_1(y_1|\theta_1), F_2(y_2|\theta_2)|\eta)$$

  where the copula family is parametric, indexed by η

- In Steve's example f_i is the density of $N(\theta_i, \sigma^2)$ and the copula is the independent one.
- The density $c(u, v)$ of the bivariate Clayton copula with parameter $\eta > 0$ is:

$$c(u, v|\eta) = (\eta + 1)(uv)^{-(\eta+1)} \left(u^{-\eta} + v^{-\eta} - 1\right)^{-\frac{2\eta+1}{\eta}} .$$

- Marginals in the same family (not necessarily Gaussian), add the Clayton copula and still be interested in $H_0 : \theta_1 = \theta_2$

One (small) step further: partial updates with dependence

- What is the predictive distribution of $Y_2|Y_1$ under H_0?
- The fly in the ointment is η
- Under $\theta_1 = \theta_2 = \theta$

$$
\begin{aligned}
p(y_2|y_1) &= \int p(y_2|y_1, \theta, \eta) p(\eta) p(\theta|y_1) d\eta d\theta \\
&= \int f_2(y_2|\theta) c(F_1(y_1|\theta), F_2(y_2|\theta)|\eta) p(\eta) p(\theta|y_1) d\eta d\theta
\end{aligned}
$$

One (small) step further: partial updates with dependence

- Integrate out η:

$$\begin{aligned} p(y_1, y_2|\theta_1, \theta_2) &= \int f(y_1, y_2|\eta, \theta_1, \theta_2) p(\eta) d\eta \\ &= f_1(y_1|\theta_1) f_2(y_2|\theta_2) \int c((F_1(y_1|\theta_1), F_2(y_2|\theta_2)|\eta) p(\eta) d\eta \\ &= K(\theta_1, \theta_2) f_1(y_1|\theta_1) f_2(y_2|\theta_2) \end{aligned}$$

- Even for relatively simple models, we may deal with doubly intractable likelihoods/targets
- What's the effect of getting the copula wrong?

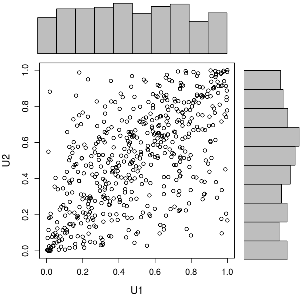
Picking up deficiencies

- Sometimes an important challenge is to identify the conditioning set/statistic.
- Consider same model as before but in a regression context:

$$f(y_1, y_2|\theta_1(x), \theta_2(x), \eta(x)) = f_1(y_1|\theta_1(x))f_2(y_2|\theta_2(x)) \\ \times c(F_1(y_1|\theta_1(x)), F_2(y_2|\theta_2(x))|\eta(x))$$

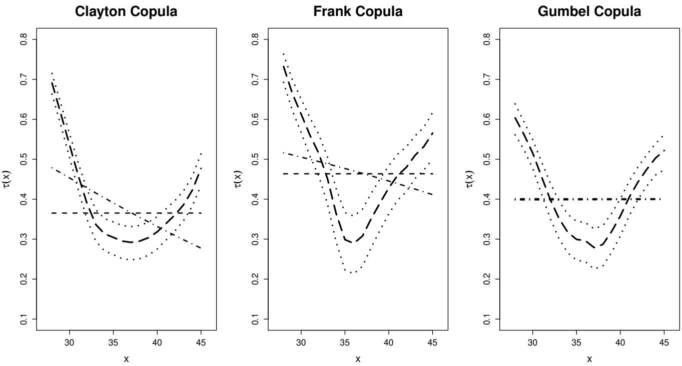
- Have confidence in the specification of marginals but not the copula.
- Specifically, in question is whether $\eta(x) = \eta \ \forall x$, aka simplifying assumption Levi and Craiu (2020).
- Significant impact in copula modeling (Haff et al., 2010; Acar et al., 2012; Hasler et al., 2018)

Picking up deficiencies



- Transformed body weights for twins delivered after different gestational ages (Acar et al., 2011). Does the dependence change with GA?
- Sometimes it is not possible to "see" what's off.
- Can we identify the conditioning statistics when we are "blind"?

Picking up deficiencies



Partial updating for picking up deficiencies

- Steve: "purposive split, tied to potential model deficiency"
- Divide the paired data into two groups, say G_1 and G_2 (ordered by GA)
- If η is independent of GA then dependence in the two groups should be the same.
- Apply the partial updating and compute p-val based on the partial predictive distribution for G_2 given G_1.
- Statistics for discrepancy could be the empirical Kendall's tau

$$\tau = \frac{2}{n_2(n_2-1)} \sum_{1 \leq i < j \leq n_2} \text{sign}(u_{1i} - u_{1j}) \cdot \text{sign}(u_{2i} - u_{2j})$$

computed in the second group.

Final thoughts

- Coherence: can two conditioning statistics pertaining to the same questionable aspect of the model lead to different conclusions?
- Robust inference: Differences and similarities with cut posteriors.
- Computation challenges quickly escalate as models increase in complexity.
- Thank you, Steve, for a stimulating talk!

References

Acar, E. F., Craiu, R. V. and Yao, F. (2011). Dependence calibration in conditional copulas: A nonparametric approach. *Biometrics* **67** 445–453.

Acar, E. F., Genest, C. and Nešlehová, J. (2012). Beyond simplified pair-copula constructions. *Journal of Multivariate Analysis* **110** 74–90.

Haff, I. H., Aas, K. and Frigessi, A. (2010). On the simplified pair-copula construction—simply useful or too simplistic? *Journal of Multivariate Analysis* **101** 1296–1310.

Hasler, C., Craiu, R. V. and Rivest, L.-P. (2018). Vine copulas for imputation of monotone non-response. *International Statistical Review* **86** 488–511.

Levi, E. and Craiu, R. V. (2020). Assessing data support for the simplifying assumption in bivariate conditional copulas. In *Nonparametric Statistics: 4th ISNPS, Salerno, Italy, June 2018 4* (M. La Rocca, B. Liseo and L. Salmaso, eds.). Springer.