Bayesian sequential design for Copula models: a comparison of designs selected under different Copula models

Senarathne SGJ, Drovandi CC, and McGree JM

School of Mathematical Sciences, Science and Engineering Faculty,

Queensland University of Technology, Brisbane, Queensland, Australia

Bayesian experimental design provides rules to allocate resources optimally for the collection of data for experimental goals such as parameter estimation, model discrimination, and/or prediction. Research in this area is mostly restricted to experiments which yield a univariate response or a small class of multivariate responses. Copula models provide a flexible way of constructing multivariate distributions for a wide range of multiple responses. Unfortunately, such models have rarely been considered in the experimental design context due to the lack of developed methodology and approaches to overcome computational challenges in dealing with a large variety of multivariate distributions.

Due to the flexibility of Copula models to describe a wide range of dependence outcomes and the computational efficiencies gained through using the sequential Monte Carlo (SMC) algorithm, we propose the combination of the two can be used to locate Bayesian designs for experiments which yield multiple responses. We demonstrate our approach by deriving designs for the dual objectives of model discrimination and parameter estimation for experiments with mixed outcomes.

Copulas

In a multivariate setting, Sklar's theorem states that there exists a function C:$[0,1]^d \rightarrow [0,1]$ between the multivariate cumulative distribution function (CDF) $G(y_1,y_2,\ldots,y_d)$ and their corresponding marginal CDFs $u_1=F_1(y_1),u_2=F_2(y_2),\ldots,u_d=F_d(y_d)$ such that, $G(y_1,y_2,\ldots,y_d)=C(u_1,u_2,\ldots,u_d)$.

The theorem also states that, if all the marginals are continuous, then C is unique; otherwise, C is uniquely determined on the $Ran(F_1)\times Ran(F_2)\times\ldots\times Ran(F_d)$ which is the Cartesian product of the ranges of marginals.

Copula models for bivariate mixed outcomes

Suppose a random variable Y_1 and another random variable Y_2 are both continuous outcomes having the marginal distributions f_{Y1} and f_{Y2}, respectively, then the Copula representation of the joint density is given by

$$f_{Y_1,Y_2}(y_1,y_2,\alpha)=f_{Y_1}(y_1)f_{Y_2}(y_2)\frac{\partial^2 C(u_1,u_2,\alpha)}{\partial u_1 \partial u_2}.$$

If one of the random variables is discrete (say Y_2), the joint density can be expressed as

$$f_{Y_1,Y_2}(y_1,y_2,\alpha)=f_{Y_1}(y_1)(C_1^*-C_2^*).$$

where $C_1^*=\frac{\partial C(u_1,u_2,\alpha)}{\partial u_1}, C_2^*=\frac{\partial C(u_1,u_2^-,\alpha)}{\partial u_1}$ and u_2^- is the left hand limit of u_2.

If the discrete random variable (Y_2) is a binary outcome, the joint distribution of Y_1 and Y_2 can be expressed as

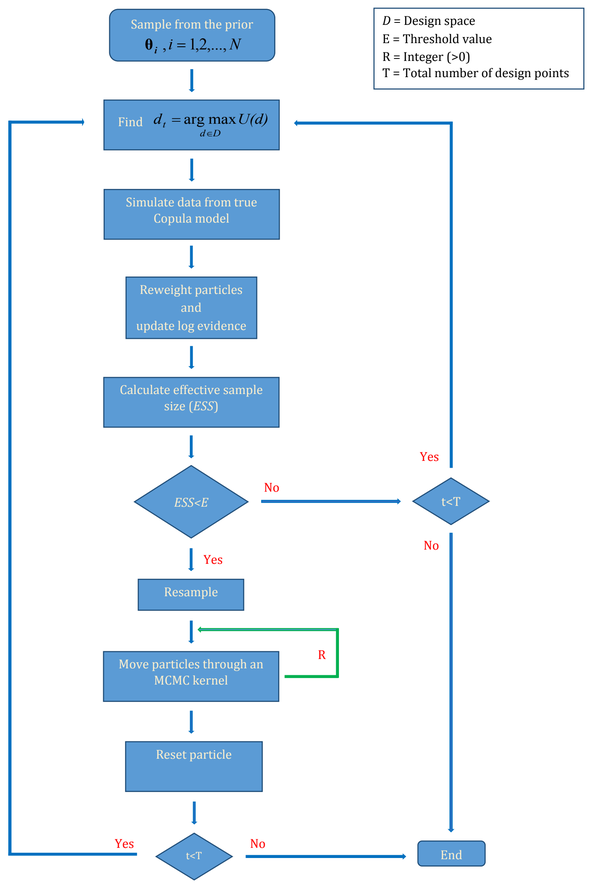
$$f_{Y_1,Y_2}(y_1,y_2,\alpha)=\begin{cases} f_{Y_1}(y_1)(1-C_1^*(u_1,u_2;\alpha)) & , y_2=0 \\ f_{Y_1}(y_1)C_1^*(u_1,u_2;\alpha) & , y_2=1 \end{cases}.$$

Therefore, the log-likelihood for a single observation is given by

$$l_i(\boldsymbol{\theta};y)=\log(f_{Y_1}(y_1))+(1-y_2)\log(1-C_1^*(u_1,u_2;\alpha))+y_2\log(C_1^*(u_1,u_2;\alpha)).$$

Bayesian sequential design

- To facilitate efficient Bayesian inference in sequential settings, we adopt the SMC algorithm (Drovandi et al., 2014).



Application

- Motivated by the work of Tao (2010), we assess the performance of this design approach in finding optimal doses for a clinical trial of Angiotensin-converting enzyme (ACE) inhibitors for prevention of hypertension and heart failure.
- Different Copula models were fitted to combine the continuous efficacy outcome with the binary toxicity outcome.
- The efficacy outcome is the change of diastolic blood pressure from baseline which follows a Normal distribution based on an E-max model as follows

$$Y_{1i}=N(f(d_i,\boldsymbol{\beta}),\sigma^2), \quad f(d_i,\boldsymbol{\beta})=\beta_0+\beta_{\max}*d_i/(\beta_{50}+d_i).$$

- The toxicity outcome measures whether the glomerular filtration rate (GFR) decrease from baseline is greater than a threshold value. If so, then this is considered a success otherwise it is a failure.

$$Y_{2i}=Ber(g(d_i,\boldsymbol{\gamma})), \quad g(d_i,\boldsymbol{\gamma})=1/(1+\exp(\gamma_0+\gamma_i d_i))$$

- This study was undertaken in the following design space with only a fixed number of doses being available. That is,

$$d_i=\{0.05n \mid n\in(1,2,\ldots,40)\}.$$

- All Copula parameters (except for the Gaussian Copula) were set to 20. We set the Gaussian Copula parameter to 0.9 to impose a similarly strong positive association between the two outcomes.
- For optimal design selection, the total entropy utility (Borth, 1975; McGree, 2017) was implemented within the SMC algorithm.

Table 1: Prior distributions of the model parameters and the Copula parameters.

Model parameter	Prior distribution	Copula parameter	Prior distribution
$\log(\beta_0)$	$N(\log(2),0.05^2)$	Frank	$U[1,30]$
$\log(\beta_{\max})$	$N(\log(13),0.05^2)$	Gumbel	$U[1,30]$
$\log(\beta_{50})$	$N(\log(0.25),0.05^2)$	Clayton	$U[1,30]$
γ_0	$N(0,3^2)$	Gaussian	$U[0,0.99]$
γ_1	$N(0,3^2)$		

Results

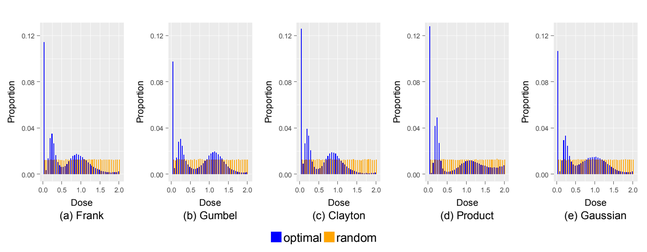


Figure 1: The distributions of the selected designs under each Copula model.

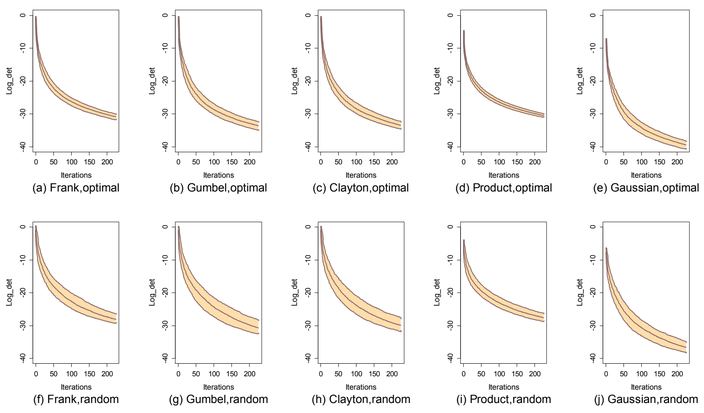


Figure 2: The quartiles of the log-determinant of the posterior variance-covariance matrix for each design point over the 500 simulations.

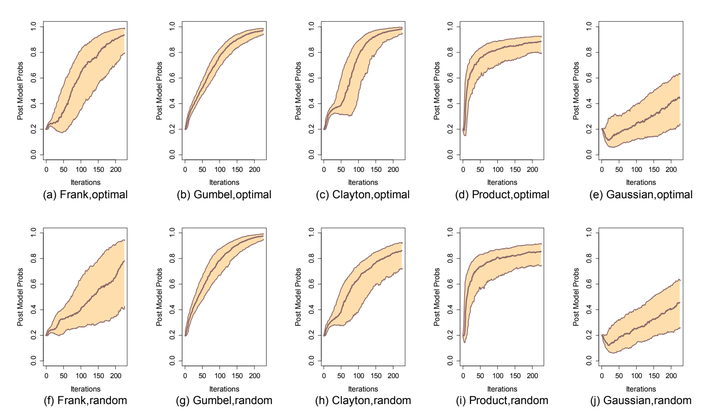


Figure 3: The distribution of the posterior model probabilities of the true Copula model, with optimal designs (row 1) and random designs (row 2), over 500 runs for 225 subjects.

- The distribution of selected optimal doses is similar for all Copula models despite the fact that the dependency structure of one Copula can be significantly different from another Copula.
- The total entropy utility function appears to estimate parameters equally well when compared to the random design for all Copula models.
- The optimal design did not show a significant gain over random design in terms of discriminating between Copula models. Similar results were found in McGree (2017) for a one dimensional design problem.

Conclusion

- The parameter estimation results revealed that it is possible to efficiently estimate model parameters across many different Copula models.
- In cases where Copulas induce similar dependence between the responses, it can be challenging to determine which model is preferred. This was observed in this example when the Frank and Gaussian Copulas were considered.
  - It may be more appropriate to only consider one Copula function to describe a particular form of dependence rather than considering multiple Copulas with only subtle differences in the dependence structure.

References

- Borth, D. M. (1975), 'A total entropy criterion for the dual problem of model discrimination and parameter estimation', *Journal of the Royal Statistical Society. Series B (Methodological)* **37**(1), 77-87.
- Drovandi, C. C., McGree, J. M. and Pettitt, A. N. (2014), 'A sequential Monte Carlo algorithm to incorporate model uncertainty in Bayesian sequential design', *Journal of Computational and Graphical Statistics* **23**(1), 3-24.
- McGree, J. M. (2017), 'Developments of the total entropy utility function for the dual purpose of model discrimination and parameter estimation in Bayesian design', *Computational Statistics and Data Analysis* **113**, 207-225.
- Tao, A. (2010), A dose-finding method in joint modelling of efficacy and safety endpoints in phase II studies: An extension of the MCP-Mod method, PhD thesis, University of Medicine and Dentistry of New Jersey.

Queensland University of Technology QUT