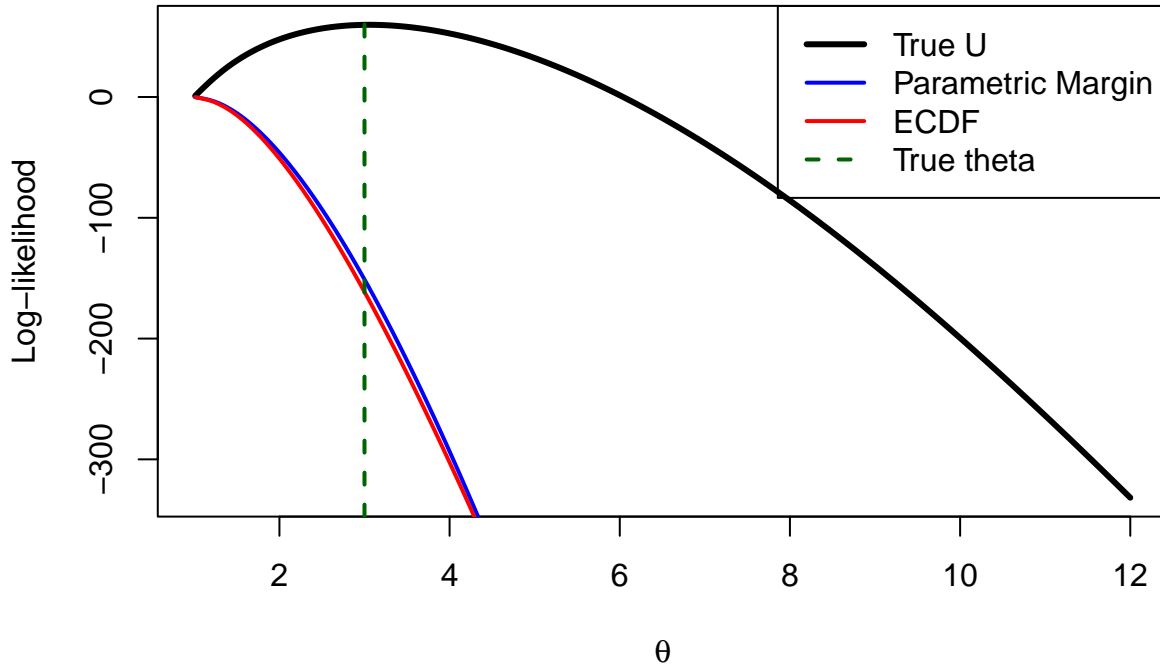# Joint Copula Bayesian Modelling

1. Motivation

We are assuming the following joint distribution for a sequence of random variables $(X_1, \ldots, X_n) \sim C_\theta(F(X_1), \ldots, F(X_n))$, with C a parametric copula (potentially archimedean) and F the marginal of each X_i.

We know from simulation results that the likelihood peaks at the correct copula generating parameter if we take the realisations $U_1, \ldots, U_n \sim C_\theta$. However, estimating the margin F from a realisation of $X_1, \ldots, X_n$, to obtain the pseudo observations $\hat{U}_i = \hat{F}(X_i)$, leads to a likelihood which peaks at $\theta = 1$, the independence case in the Gumbel copula. This phenomenon could be explained by realising that we are using estimators for the margin which work under independence, but might fail under the dependence.

Gumbel log–likelihood

The idea is to avoid the 2 steps procedure by performing estimation on the margin and the copula parameter simultaneously. A fully bayesian model would inherently consider both objects as targets, and might be able to consider the dependence structure while estimating the margin.

2. Bayesian Model with Parametric Margin

The simplest approach consists in assuming a parametric form for the margin, that is $X_i \sim F_\varphi$, with φ a generic parameter vector. Hierarchically, this yields the following structure:

$$\varphi \sim \pi_\varphi \,, \quad \theta \sim \pi_\theta$$
$$(X_1, \ldots, X_n) | \, \theta, \varphi \sim C_\theta(F_\varphi(X_1), \ldots, F_\varphi(X_n))$$

2.1 Adaptive MCMC

We aim to sample from the joint log posterior:

$$\log \pi(\theta, \varphi \mid x_{1:n}) \; \propto \; \log \pi_\theta(\theta) + \log \pi_\varphi(\varphi) + \log c_\theta(F_\varphi(x_1), \ldots, F_\varphi(x_n)) + \sum_{i=1}^n \log f_\varphi(x_i).$$

We will call the full parameter vector:

$$\boldsymbol{\phi} = (\theta, \varphi) \in \mathbb{R}^d.$$

To sample from this posterior, a simple algorithm is a Random Walk Metropolis–Hastings. Crucial is the proposal covariance Σ_0 set at the beginning. This determines the efficiency of the sampler while exploring the posterior distribution. Higher variance in the proposals implies the sampler will take with higher probability further steps, moving quicker around the posterior space but with a higher probability of ending up in regions of low posterior probability, increasing therefore the risk of rejection. In contrast, smaller proposal variance induces the sampler to take shorter steps, decreasing the risk of being rejected but at the same time exploring the posterior space more slowly. A good balance needs to be found and no simple solution exists.

Algorithm 1: Block Metropolis–Hastings

1. **Initialisation:**
   - Set the initial parameter vector $\boldsymbol{\phi}^{(0)}$.
   - Compute the initial log-posterior $\ell_{\text{curr}} = \ell(\boldsymbol{\phi}^{(0)})$.
   - Set the proposal covariance $\Sigma_0 = \varepsilon_0 I_d$.

2. **For** $t = 1, \ldots, N$:

   (a) **Proposal:**
   Draw
   $$\boldsymbol{\phi}^\star \sim \mathcal{N}_d\Big(\boldsymbol{\phi}^{(t-1)}, \Sigma_0\Big).$$

   (b) **Evaluate proposed log-posterior:**
   Compute $\ell^\star = \ell(\boldsymbol{\phi}^\star)$.

   (c) **Acceptance probability:**
   $$\alpha = \min\{1, \exp(\ell^\star - \ell_{\text{curr}})\}.$$

   (d) **Accept/reject:**
   Draw $u \sim \text{Unif}(0, 1)$.
   - If $u < \alpha$: set $\boldsymbol{\phi}^{(t)} = \boldsymbol{\phi}^\star$ and $\ell_{\text{curr}} = \ell^\star$.
   - Else: set $\boldsymbol{\phi}^{(t)} = \boldsymbol{\phi}^{(t-1)}$.

To overcome the choice of the proposal covariance Σ_0, adaptive methods have been proposed. The idea behind adaptive MCMC is to learn the optimal proposal variance while the chain is running, or to target a specific acceptance rate considered to be optimal. In the case of multidimensional posterior, a standard acceptance rate to target is $\alpha^* = 0.234$.

We considered mainly two adaptive schemes: the Haario scheme and the Robbins-Monroe scheme.

The first one simply starts by an arbitrary covariance and updates it by computing the empirical covariance $\hat{\Sigma}$ of the chains at that moment. Instead, Robbins-Monroe scheme targets a specific acceptance rate α^* by updating the scale of the covariance iteratively.

Care must be taken when using adaptive methods as theoretical guarantees of ergodicity require some assumptions to be met, one of which is for example diminishing adaptation, to ensure that after a while adaptation practically stops and we ensure the Markov Chain will converge to its stationary distribution.

2.2 Simulation Based Inference

This branch of Bayesian inference aims at overcoming the impossibility of evaluating Bayesian models when the likelihood is intractable. The key ingredient is being able to (efficienciently) sample from our model to produce synthetic data that can be

2.2.1 Likelihood Intractability

The MCMC scheme used so far requires being able to evaluate the loglikelihood, and therefore the log density of the copula, to compute the acceptance probability α. However, defining a copula over the entire set of variables leads to numerical issues in evaluating the likelihood. This implies that the likelihood is intractable after a certain number of observations.

```r
library(copula)
set.seed(123)

n <- 1000
theta_true <- 2

cop <- gumbelCopula(param = theta_true, dim = n)
U <- as.numeric(rCopula(1, cop))   # true uniforms

dCopula(U, gumbelCopula(theta_true, dim = n))
```

```
## Error in polyG(lx[i.m[[meth]]], alpha = alpha, d = d, method = meth, log = log): d > 220 not yet sup
```