"Improving estimation for asymptotically independent bivariate extremes via global estimators for the angular dependence function"

C. J. R. Murphy-Barltrop, J. L. Wadsworth, E. F. Eastoe

Janos C. R. Füting

December 3, 2023

The Coefficient of Tail Dependence

Recall from the first lecture on multivariate extremes that we can characterize the dependence of two variables by their coefficient of tail dependence η:

$$\mathbb{P}(U > u, V > u) \sim L\left(\frac{1}{1-u}\right)(1-u)^{1/\eta} \qquad 0.5 \leq \eta \leq 1$$

or equivalently on standard exponential margins as

$$\mathbb{P}(X > u, Y > u) = \mathbb{P}(\min(X, Y) > u) = L(e^{u})e^{-u/\eta} \qquad 0 < \eta \leq 1$$

with L our usual slowly varying function.
Both of these approaches imply that we only look at the behavior when the variables jointly grow at the same rate.

The Angular Dependence Function

Recall from the last lecture the pseudo-polar (radial angular) representation of the variables:

$$r = X + Y \qquad w = \frac{X}{X + Y}$$

then we can extend the coefficient of tail dependence characterization to each of those *rays* w:

$$\mathbb{P}(\min\{X/w, Y/(1-w)\} > u) = L(e^u; w)e^{-u\lambda(w)} \qquad \lambda(w) \geq \max(w, 1-w)$$

Not only does the ADF tell us something about the dependence structure of the variables, but we can also invert this definition to get from the ADF to return level curves.

Example ADFs

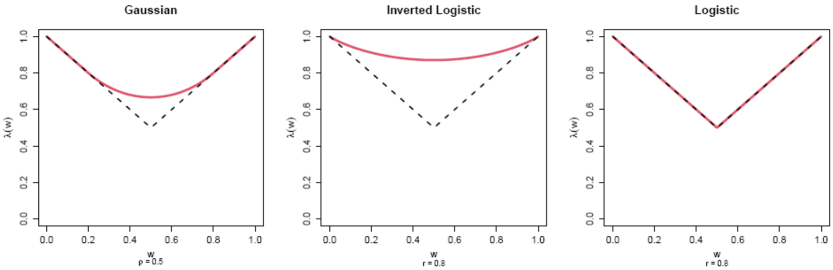


Figure: True ADFs (red) and theoretical lower bound (black). Figure taken from paper.

Previous ADF Estimators 1/2: Hill

Consider the minimum projection along a ray w:

$$T_w := \min\{X/w, Y/(1-w)\}$$

then for this new variable we have

$$\mathbb{P}(T_w > u + t | T_w > u) = \frac{L(e^{u+t}; w)}{L(e^u; w)} e^{-\lambda(w)t} \to e^{-\lambda(w)t} = t_\star^{-\lambda(w)}$$

Which implies that for a fixed w (i.e. point-wise) we can estimate $\lambda(w)$ by Hill estimator.

Previous ADF Estimators 2/2: Simpson and Tawn

The shape of the limiting boundary set of scaled samples of X and Y is linked to a "projection" r_w^{-1}:

$$\lambda(w) = \max(w, 1-w) \times r_w^{-1}$$

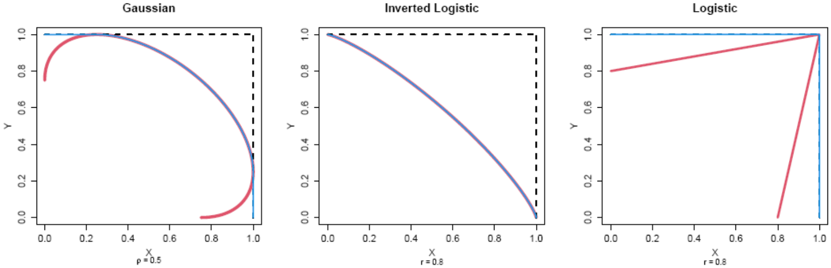


Figure: Boundary sets (red), and our projection r_w (blue). Figure taken from paper.

Approximation: Bernstein-Bézier Polynomials - Standard

If you want to approximate a function you need a sufficiently flexible form, usually splines or polynomials. If you have $f : [0,1] \to [0,1]$ Bernstein-Bézier polynomials are a good choice:

$$\mathscr{B}_k(w) = \left\{ \sum_{i=0}^{k} \beta_i \binom{k}{i} w^i (1-w)^{k-i} \;\middle|\; \beta_i \in [0,1], w \in [0,1] \right\}$$

Approximation: Bernstein-Bézier Polynomials - Adjusted

Regrettably our co-domain is not $[0,1]$, $\lambda(w)$ is bounded below by $\max(w, 1-w)$, especially $\lambda(0) = \lambda(1) = 1$, and unbounded above elsewhere. Adjustment:

$$\mathscr{B}_k^\star(w) = \left\{ (1-w)^k + \sum_{i=1}^{k-1} \beta_i \binom{k}{i} w^i (1-w)^{k-i} + w^k \;\middle|\; w \in [0,1], \beta_i \in \mathbb{R}_+ \right\}$$

with β_i such that $\mathscr{B}_k(w) \geq \max(w, 1-w)$

This indirect restriction of the betas makes the problem not so nice.

Loss Function 1: Composite Likelihood

Recall our minimum projection $T_w := \min\{X/w, Y/(1-w)\}$. Its exceedances over some high enough threshold are exponentially distributed:

$$T_w^\star := \{T_w - u_w | T_w > u_w\} \sim \mathrm{Exp}(\lambda(w))$$

leading to the following composite likelihood across w:

$$\mathcal{L}_C(\beta) = \prod_w \prod_{t_w^\star} \lambda(w;\beta) e^{-\lambda(w;\beta)t_w^\star} = \left(\prod_w \lambda(w;\beta)^{|\mathbf{t}_w^\star|}\right) e^{-\sum_w \sum_{t_w^\star} \lambda(w;\beta) t_w}$$

the negative log of which we can use as our loss function.

Loss Function 2: Probability Ratio

Consider two probabilities $q < p < 1$ close to 1 and the associated quantiles u_w, v_w:

$$\frac{1-p}{1-q} = \mathbb{P}(T_w > v_w | T_w > u_w) \approx e^{-\lambda(w)(v_w - u_w)}$$

We can consider as a loss function the sum of absolute deviations from this relationship:

$$\sum_w \left| \frac{1-p}{1-q} - e^{-\lambda(w)(v_w - u_w)} \right| \approx 0$$

Incorporating Information from the Limit Set

Idea: Use the linear part of the Limit Set estimator only and our other 3 estimation methods inbetween.

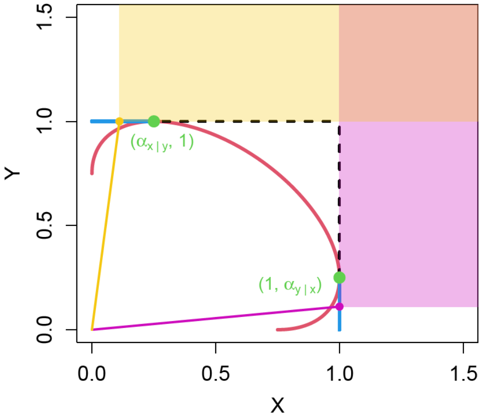


Figure: An illustration of the linear part of our boundary set projection r_w, taken from paper.

Simulation Study: Dependence Structures

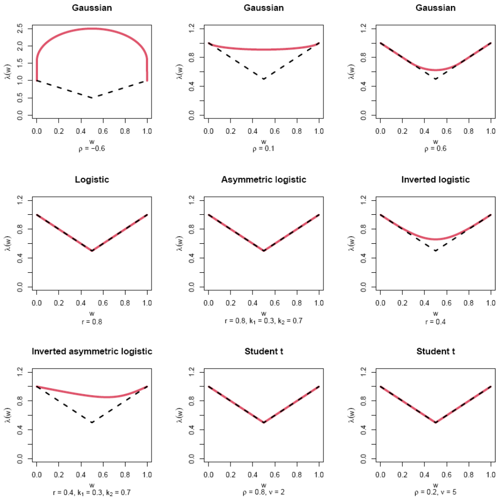


Figure: True ADFs (red) and theoretical lower bound (black). Figure taken from paper.

Simulation Study: Criterion

To asses the goodness of our estimators we use the Root Mean Integrated Squared Error:

$$RMISE(\hat{\lambda}) = \left(\mathbb{E} \left[\int_0^1 \left\{ \hat{\lambda}(w) - \lambda(w) \right\}^2 \mathrm{d}w \right] \right)^{1/2} = [ISB + IV]^{1/2}$$

with ISB the Integrated Squared Bias and IV the Integrated Variance.

Simulation Study: Results

Copula	$\hat{\lambda}_H$	$\hat{\lambda}_{CL}$	$\hat{\lambda}_{PR}$	$\hat{\lambda}_{H2}$	$\hat{\lambda}_{CL2}$	$\hat{\lambda}_{PR2}$	$\hat{\lambda}_{ST}$
Gaussian $\rho = -0.6$	61.1	61.3	66.2	61.4	61.9	66.7	63.7
Gaussian $\rho = 0.1$	3.55	3.33	3.64	3.51	3.33	3.63	2.95
Gaussian $\rho = 0.6$	3.78	3.48	3.84	3.27	3.22	3.57	1.09
Logistic $r = 0.8$	4.9	4.79	6.92	4.28	4.25	6.17	2.77
Asym. Log.	14.1	1.11	17.1	14.1	14.1	17	12.1
Inv. Log.	2.51	1.97	2.15	2	1.74	1.9	2.12
Inv Asym. Log.	2.93	2.64	2.88	2.87	2.66	2.89	3.96
Student t $\rho = 0.8, \nu = 2$	2.49	2.72	2.95	0.66	0.6	0.789	1.87
studen t $\rho = 0.2, \nu = 5$	12.1	12	14.9	12	12	14.9	11.1

Table: $RMISE \times 1000$ for the estimators and given dependence structures, data from paper.

Some More Lessons from Simulations

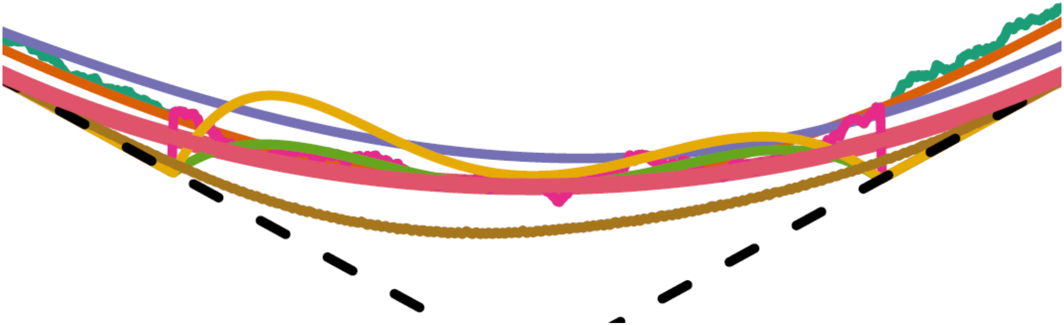
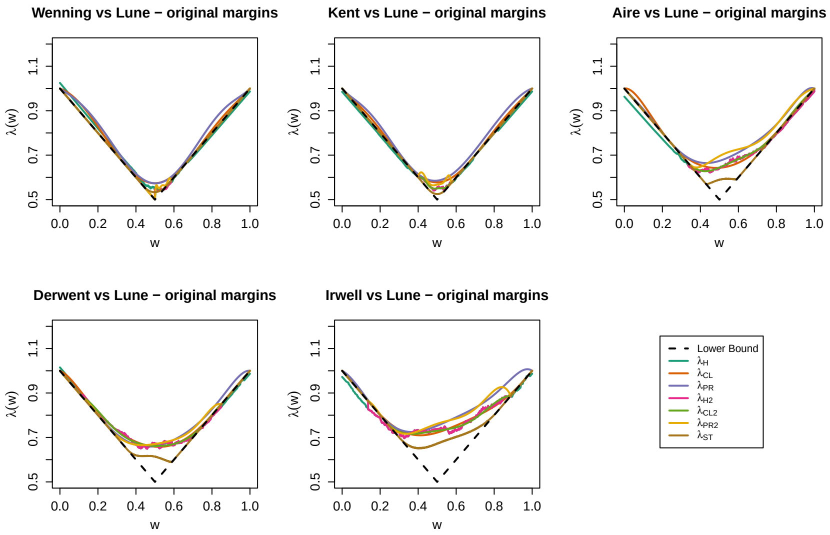
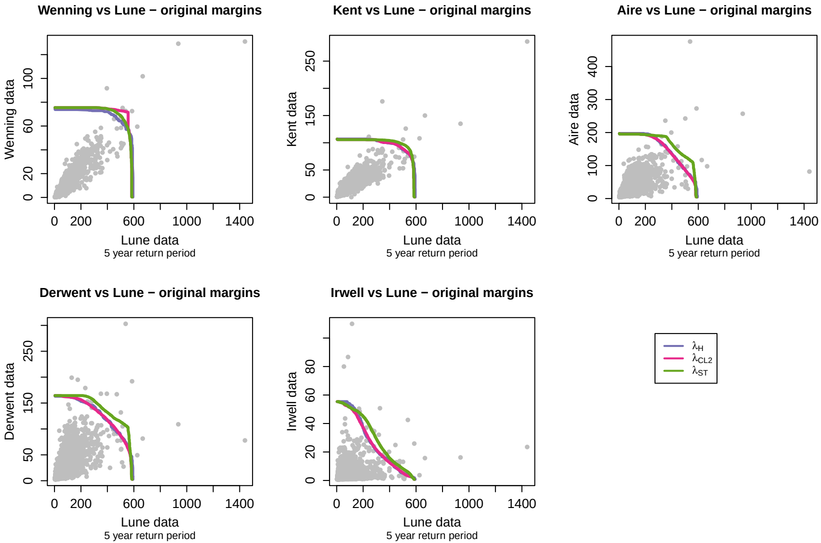


Figure: Middle part of Example Estimates for Gaussian Data with true ADF superimposed. Figure from paper.

Real World Application: River Flows - ADFs



Real World Application: River Flows - Return Curves



The End

Thank you for your attention!

Appendix: Bernstein-Bézier Polynomials - Fitting

They choose to parameterise as $\beta_i = e^{b_i}$ and adjusted loss function

$$\ell_a(\mathbf{b}) = \begin{cases} 10^{10} & \text{if any } \mathscr{B}_k(w) < \max(w, 1-w) - 10^{-10} \\ \ell(\mathbf{b}) & \textit{else} \end{cases}$$

and try to alternate Nelder-Mead and BFGS multiple times to get proper convergence.

Unfortunately this is not entirely reliable, at times we get errors for Nelder-Mead and at the same time an exploding step-size in BFGS that catapults us into in-admissible regions (breaking barrier and/or unable to evaluate function).

So I am currently still working on how to re-frame the problem to make it behave "nice".