Lecture 13. Case study: Mixture copula model

Enkelejd Hashorva & Pavel Ievlev
Université de Lausanne

8 December, 2025

Learning objectives

- Derive the copula C of random maxima $(X_{N:N}, Y_{N:N})$
- Sample data from C given N and the df G of (X_1, Y_1)
- Focus on three tractable instances for N and calculate by simulation measures of dependence for C
- Use pseudo-ML to fit C_α to concrete insurance data

Random maximum claims

Let (X_i, Y_i)'s be iid with df G. Define the **random maximum**

$$(M_N, M_N^\star) = \left(\max_{1\le i\le N} X_i, \max_{1\le i\le N} Y_i \right)$$

where the integer-valued rv $N \geq 1$ is independent of (X_i, Y_i)'s.

The df F of $(M_N, M_N^\star)$ is a **mixture df**, i.e.,

$$F(x,y) = \sum_{n=1}^{\infty} \mathbb{P}\left\{N = n\right\} G^n(x,y) = \mathbb{E}\left\{ e^{N \ln G(x,y)} \right\} = L(-\ln G(x,y)),$$

with L the **Laplace transform** of N.

Marginal df's

The marginal df's of F are given from the marginals G_1, G_2 of G by

$$F_1(x) = L(-\ln G_1(x)), \quad F_2(x) = L(-\ln G_2(x)), \quad x \in \mathbb{R}$$

If G has **continuous** marginal df's, then F_i's are continuous implying that the copula Q of G is unique and given by

$$C(u_1, u_2) = L\left(-\ln Q(v_1, v_2)\right), \quad v_i = e^{-L^{-1}(u_i)}, \quad u_1, u_2 \in [0, 1]$$

Simulation from $C \sim (U_1, U_2)$ & known LT L

- **S1:** Simulate n from N with Laplace transform L
- **S2:** Generate a RS (V_{i1}, V_{i2}), $1 \leq i \leq n$ from copula Q
- **S3:** Calculate the component-wise maximum $(\widetilde{M}_1, \widetilde{M}_2)$ by

$$\widetilde{M}_j = \max_{1 \leq i \leq n} V_{ij}, \quad j = 1, 2$$

- **S4:** Return (U_1, U_2) with the representation

$$U_j = L(-\ln \widetilde{M}_j), \quad j = 1, 2$$

Model A: N is shifted Geometric

Suppose that for $\theta \in (0,1)$

$$\mathbb{P}\{N = n\} = (1-\theta)^{n-1}\theta, \quad \text{for } n = 1, 2, \ldots$$

with LT $L(t) = \theta e^{-t}/[1-(1-\theta)e^{-t}]$ and hence

$$F(x,y) = \frac{\theta G(x,y)}{1-(1-\theta)G(x,y)}, \quad x, y \in \mathbb{R}$$

and

$$C(u_1,u_2) = \frac{\theta Q(v_1,v_2)}{1-(1-\theta)Q(v_1,v_2)}, \quad v_j = \frac{u_j}{\theta + (1-\theta)u_j}, \quad j = 1, 2$$

where $v_j = F_j^{-1}(u_j)$ is obtained by inverting the marginal df.[1]

[1] The transformation v_j differs for each model because it comes from inverting $F_j(x) = L(-\ln G_j(x))$.

Model B: N is shifted Poisson

Let $N = 1 + T$ with T a Poisson rv with mean $\theta > 0$ and LT

$$L(t) = \mathbb{E}\left\{e^{-t-tT}\right\} = e^{-t}\mathbb{E}\left\{e^{-tT}\right\} = e^{-t}e^{-\theta(1-e^{-t})}$$

implying

$$F(x,y) = G(x,y)e^{-\theta(1-G(x,y))}, \quad x, y \in \mathbb{R}$$

and

$$C(u_1,u_2) = Q(v_1,v_2)e^{-\theta(1-Q(v_1,v_2))}, \quad u_1, u_2 \in [0,1]$$

where $v_j = u_j e^{\theta(u_j-1)}$, $j = 1, 2$.

Model C: $N = (T|T > 0)$ is truncated Poisson

Suppose that the rv T is Poisson with parameter $\theta > 0$, hence

$$\mathbb{P}\{N = k\} = \frac{e^{-\theta}\theta^k}{k!(1 - e^{-\theta})}, \quad k \geq 1.$$

Its LT is given by

$$L(t) = \sum_{k=1}^{\infty} \mathbb{P}\{N = k\} e^{-kt} = \frac{a}{1-a}\left[e^{\theta e^{-t}} - 1\right], \quad a = e^{-\theta}$$

and hence

$$F(x, y) = \frac{a}{1-a}\left[e^{\theta G(x,y)} - 1\right], \quad x, y \in \mathbb{R}$$

$$C(u_1, u_2) = \frac{a}{1-a}\left[e^{\theta Q(v_1,v_2)} - 1\right], \quad u_1, u_2 \in [0, 1]$$

with $v_j = \frac{1}{\theta}\ln(1 + u_j(1-a)/a)$, $j = 1, 2$.

Simulation from Models A, B and C

Using the simulation algorithm, we can simulate copulas from Models A, B and C for N.

- Q belongs to the Gumbel or Clayton family with parameter $\alpha = 10$
- N follows the shifted Poisson df
- S1–S4 of simulation algorithm are repeated 10'000 times
- Computed empirical Kendall's τ for both C and Q for different values of $\mathbb{E}\{N\}$ is given in the following table

Simulation results

$\mathbb{E}\{N\}$	Q: Gumbel copula with $\alpha = 10$		Q: Clayton copula with $\alpha = 10$	
	$\tau(C)$	$\tau(Q)$	$\tau(C)$	$\tau(Q)$
10	0.9059	0.9022	0.3533	0.8343
100	0.8980	0.9002	0.0518	0.8348
1'000	0.9007	0.9004	0.0043	0.8334
10'000	0.9016	0.9018	0.0019	0.8324
100'000	0.8997	0.8996	-0.0104	0.8316

Interpretation of simulations

- For Q the Gumbel copula, the level of dependence governed by C is approximately equal to that of Q, even when $\mathbb{E}\{N\}$ increases.
- If Q is the Clayton copula, the bigger $\mathbb{E}\{N\}$ the weaker the dependence associated with C.
- The results indicate (and this turns out to be true) that, if $\lambda(C_U) = 0$, i.e., **Clayton copula**, when $\mathbb{E}\{N\}$ increases, $C \to C_I$ since $\tau(C) \simeq 0$.
- If Q is an EVC, i.e., **Gumbel copula** in our case, the dependence is preserved because EVCs have positive upper tail dependence $\lambda(C_U) > 0$.

Using C for modelling insurance data

Suppose that $Q = Q_\alpha$ depends on an unknown parameter $\alpha > 0$. Then the copula C for Model A, B, C is parametrised by α and θ, where θ is the parameter of the df of N.

Estimation of α and θ can be done using the **pseudo-likelihood approach**.

Loss-ALAE from medical insurance

SOA Medical Group Insurance data sets describing the medical claims observed over the years 1991–1992.

	Loss	ALAE[a]
Min	25’003	5
Q1	30’859	7’775
Q2	40’985	14’111
Q3	64’067	23’547
Max	1’404’432	409’586
No. Obs.	5’106	5’106
Mean	62’589	20’001
Std. Dev.	69’539	24’130

[a] Allocated loss adjustment expenses ≈ paid expenses to a given loss

Dependence measures	Values
Pearson’s Correlation	0.44
Spearman’s Rho	0.44
Kendall’s Tau	0.30
Upper tail dependence	0.38

Loss-ALAE from medical insurance: Fitting

- Q is either Gumbel, Frank, Student or Joe copula
- Criteria for the goodness of fit:
  - AIC criteria: $AIC = -2l(\hat{\Theta}) + 2p$, where p corresponds to the number of parameters to estimate
  - Cramér–von Mises statistic: computation of the p-values based on a bootstrap procedure
  - Root Mean Square Error

Loss-ALAE data set: Results

Original copula Q	Distribution for N	P-value	RMSE	AIC
Gumbel	None	0.755	0.0039	-1,371.21
	Geometric	0.740	0.0039	-1,369.28
	Truncated Poisson	0.745	0.0039	-1,369.27
	Shifted Poisson	0.527	0.0047	-1,328.14
Frank	None	0.021	0.0096	-1,137.12
	Geometric	0.026	0.0096	-1,135.09
	Truncated Poisson	0.021	0.0096	-1,135.10
	Shifted Poisson	0.017	0.0096	-1,135.12
Student	None	0.046	0.0090	-1,195.83
	Geometric	0.063	0.0090	-1,193.82
	Truncated Poisson	0.024	0.0090	-1,193.82
	Shifted Poisson	0.045	0.0090	-1,193.82
Joe	None	0.055	0.0039	-1,371.21
	Geometric	**0.986**	**0.0027**	**-1,393.23**
	Truncated Poisson	0.919	0.0032	-1,386.87
	Shifted Poisson	0.892	0.0034	-1,384.34

Loss-ALAE from general liability insurance

This data set describes the general liability claims associated with their ALAE retrieved from the Insurance Services Office available in the R package.

It consists of 1'466 uncensored data points and 34 censored observations.

Let X_i be the i-th loss observed and Y_i the ALAE associated to the settlement of X_i.

Each loss is associated with a maximum insured claim amount (policy limit) M. Thus, the loss variable X_i is censored when it exceeds the policy limit M. We define the censored indicator of the loss variable by

$$\delta_i = \begin{cases} 1 & \text{if } X_i \leq M, \\ 0 & \text{if } X_i > M, \end{cases} \qquad i = 1, \ldots, 1'500$$

Loss-ALAE from general liability insurance (contd.)

Due to censoring, a modified estimator (**Kaplan–Meier estimator**) $\hat{G}_X$ is used to estimate G_1. The corresponding pseudo log-likelihood function is also adapted accordingly.

The analysis is more technical, but similar to the first application, see R code.

Danish fire insurance data

This dataset consists of three components.

We shall model the dependence between the loss amount on the building and the loss amount for the content inside the building.

The total number of observations is 1'501.

We shall consider only observations where both components are positive.

The analysis is similar to the first application, see R code.

Loss-ALAE from accident insurance

- We consider an **insurance data** from a large insurance company operating in Switzerland
- The dataset consists of 33'258 accident insurance losses and their corresponding **allocated loss adjustment expenses** (ALAE) which include the cost of medical consultancy and legal fees
- The observation period encompasses the claims occurring during 1986–2014

Loss-ALAE from accident insurance: Summary

	Loss	ALAE
Min	10	1
Q1	13'637	263
Q2	32'477	563
Q3	95'880	1'509
Max	133'578'900	2'733'282
No. Obs.	33'258	33'258
Mean	292'715	5'990
Std. Dev.	2'188'622	42'186

Dependence measures	Values
Pearson's Correlation	0.74
Spearman's Rho	0.74
Kendall's Tau	0.60
Upper tail dependence	0.68

References

- M. Tamraz, “Mixture copulas and insurance applications,” *Annals of Actuarial Science*, vol. 12, no. 2, pp. 391–411, 2018.
- E. Hashorva, G. Ratovomirija, and M. Tamraz, “On some new dependence models derived from multivariate collective models in insurance applications,” *Scandinavian Actuarial Journal*, vol. 2017, no. 8, pp. 730–750, 2017.