Lecture 11. Parametric statistics of dependent risks

Enkelejd Hashorva & Pavel Ievlev
Université de Lausanne

1 December, 2025

Leargning objectives

- Statistical models: parametric, semi-parametric, separable, non-separable.
- Maximum likelihood estimation (MLE) and pseudo-MLE (PMLE).
- Hypothesis testing for choosing a copula.

Statistical models

Definition 1 (Statistical model).

A **statistical model** is a **family of distributions** $\mathcal{F} = \{F\}$ used to approximate the mechanism generating the data $\mathscr{D}_n$.

In other words, a statistical model is a set of hypotheses of the form

"the data $\mathscr{D}_n$ are generated from a distribution $F \in \mathcal{F}$."

A model is called **parametric** if the family $\mathcal{F} = \{F_{\boldsymbol{\theta}} : \boldsymbol{\theta} \in \Theta \subset \mathbb{R}^d\}$ is indexed by a **finite-dimensional** parameter $\boldsymbol{\theta}$.

Definition 2 (Statistical estimation).

Statistical estimation aims to find the "best" $F \in \mathcal{F}$ fitting the data $\mathscr{D}_n$. The model does not claim that the true distribution of the data belongs to $\mathcal{F}$, but rather that $\mathcal{F}$ can generate data "sufficiently similar" to $\mathscr{D}_n$.

Parametric and semi-parametric models with pdf

Assume that all dfs F of a given statistical model $\mathcal{F}$ have a pdf f. Then, both marginals and copula have pdfs given by

$$f(\boldsymbol{x}) = c\big(F_1(x_1), \dots, F_d(x_d)\big) \prod_{i=1}^{d} f_i(x_i).$$

We can classify such models as follows:

- **Parametric vs semi-parametric models:** if both marginals and copula are parametric, then the model is called **fully parametric**; if at least one of them is non-parametric, then the model is called **semi-parametric**.
- **Separable vs non-separable models:** if the marginals and the copula can be estimated **separately**, then the model is called **separable**; otherwise, it is called **non-separable**.

Separable models

A **fully parametric** and **separable** model with pdf is of the form

$$f_{\boldsymbol{\theta}_1,\boldsymbol{\theta}_2,\boldsymbol{\theta}_3}(x,y) = f_{1,\boldsymbol{\theta}_1}(x)\, f_{2,\boldsymbol{\theta}_2}(y)\, c_{\boldsymbol{\theta}_3}\big(F_{1,\boldsymbol{\theta}_1}(x), F_{2,\boldsymbol{\theta}_2}(y)\big), \quad \boldsymbol{\theta}_i \in \boldsymbol{\Theta_i}.$$

Note that the parameters of the marginals $\boldsymbol{\theta}_1, \boldsymbol{\theta}_2$ and the copula $\boldsymbol{\theta}_C$ are **distinct**.

A **semi-parametric** and **separable** model with pdf is of the form

$$f(x,y) = f_1(x)\, f_2(y)\, c_{\boldsymbol{\theta}}\big(F_1(x), F_2(y)\big), \quad \boldsymbol{\theta} \in \Theta,\ F_i \text{ non-parametric.}$$

One important example is when we assume a semi-parametric separable model with F_i estimated by the **empirical dfs**.

Maximum likelihood estimation

A **universal** method for estimating parameters of parametric models *with pdfs* is the **maximum likelihood estimation** (MLE). Given a random sample $\mathcal{D}_n \sim F_{\boldsymbol{\theta}}$, we define the **likelihood function** as

$$L(\boldsymbol{\theta}; \mathcal{D}_n) = \prod_{i=1}^{n} f_{\boldsymbol{\theta}}(\boldsymbol{X}_i).$$

The **maximum likelihood estimator** $\hat{\boldsymbol{\theta}}_n$ is defined (**if it is finite!**) as

$$\hat{\boldsymbol{\theta}}_n = \operatorname{argmax}_{\boldsymbol{\theta} \in \Theta} L(\boldsymbol{\theta}; \mathcal{D}_n).$$

It is often more convenient to work with the **log-likelihood function**

$$\ell(\boldsymbol{\theta}; \mathcal{D}_n) = \ln L(\boldsymbol{\theta}; \mathcal{D}_n) = \sum_{i=1}^{n} \ln f_{\boldsymbol{\theta}}(\boldsymbol{X}_i).$$

MLE example: multivariate normal

If $\mathcal{D}_n \sim N(\boldsymbol{\mu}, \Sigma)$ and Σ is invertible, then

$$\ell(\boldsymbol{\mu}, \Sigma; \mathcal{D}_n) = -\frac{nd}{2}\ln(2\pi) - \frac{n}{2}\ln|\Sigma| - \frac{1}{2}\sum_{i=1}^{n}(\boldsymbol{X}_i - \boldsymbol{\mu})^\top \Sigma^{-1}(\boldsymbol{X}_i - \boldsymbol{\mu}).$$

To maximize ℓ with respect to μ and Σ to zero, we set the gradients to zero and solve:

$$\nabla_{\boldsymbol{\mu}}\ell = \Sigma^{-1}\sum_{i=1}^{n}(\boldsymbol{X}_i - \boldsymbol{\mu}) = 0,$$

$$\nabla_{\Sigma}\ell = -\frac{n}{2}\Sigma^{-1} + \frac{1}{2}\Sigma^{-1}\Big(\sum_{i=1}^{n}(\boldsymbol{X}_i - \boldsymbol{\mu})(\boldsymbol{X}_i - \boldsymbol{\mu})^\top\Big)\Sigma^{-1} = 0.$$

MLE example: multivariate normal

Solving the system of equations, we obtain the MLEs

$$\hat{\boldsymbol{\mu}}_n = \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{X}_i, \quad \text{and} \quad \widehat{\Sigma}_n = \frac{1}{n}\sum_{i=1}^{n}(\boldsymbol{X}_i - \hat{\boldsymbol{\mu}}_n)(\boldsymbol{X}_i - \hat{\boldsymbol{\mu}}_n)^\top,$$

which are the sample mean vector and the sample covariance matrix, respectively.

Pseudo-MLE

Assuming a **semi-parametric** and **separable** model, it makes sense to consider the so-called **dependence log-likelihood function**

$$\ell_{\mathrm{dep}}(\boldsymbol{\theta}; \mathcal{D}_n) = \sum_{i=1}^{n} \ln c_{\boldsymbol{\theta}}\big(F_1(X_{i1}), \ldots, F_d(X_{id})\big),$$

Estimating the marginals F_i non-parametrically by the empirical dfs $\hat{F}_{i,n}$, we obtain the **pseudo log-likelihood function**

$$\widehat{\ell}_{\mathrm{dep}}(\boldsymbol{\theta}; \mathcal{D}_n) = \sum_{i=1}^{n} \ln c_{\boldsymbol{\theta}}\big(\hat{F}_{1,n}(X_{i1}), \ldots, \hat{F}_{d,n}(X_{id})\big).$$

The **pseudo maximum likelihood estimator** $\hat{\boldsymbol{\theta}}_n$ is defined (**if is it finite!**) as

$$\hat{\boldsymbol{\theta}}_n = \mathrm{argmax}_{\boldsymbol{\theta} \in \Theta} \, \widehat{\ell}_{\mathrm{dep}}(\boldsymbol{\theta}; \mathcal{D}_n).$$

PMLE for Archimedean copulas

To formulate PMLE for an Archimedean copula model C_ψ, we need to

- assume that ψ comes from some parametric family of generators $\{\psi_{\boldsymbol{\theta}}\}$
- and find its pdf c_ψ.

The formula for the pdf looks easier if we introduce $\phi := \psi^{-1}$. We have:

$$c_\psi(u, v) = \frac{\phi''(\psi(u) + \psi(v))}{\phi'(\psi(u))\,\phi'(\psi(v))}.$$

Mixture copula models

Another popular class of parametric copula models is the class of **mixture copulas**. Recall that mixture copula is defined as

$$C_{\boldsymbol{w}}(u,v) = \sum_{i=1}^{k} w_i\, C(u,v), \quad \boldsymbol{w} \in [0,1]^k : \sum_{i=1}^{k} w_i = 1.$$

To estimate the weights $\boldsymbol{w}$, we can use PMLE, but it is often easier and more efficient to calculate their **moments** or their **dependence measures** such as Kendall's τ or Spearman's ρ_S, and match them against their empirical estimates.

Hypothesis testing for choosing a copula

Choosing an adequate copula model $\mathcal{C}$ (family of copulas) may be formulated as a **hypothesis testing** problem:

$$H_0 : C \in \mathcal{C} \quad \text{vs} \quad H_1 : C \notin \mathcal{C},$$

One popular idea is to compare the **best copula** $C_{\hat{\theta}_n}$ in $\mathcal{C}$ to the **empirical copula**

$$\widehat{C}_n(u,v) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}\{\hat{U}_{i1} \leq u, \hat{U}_{i2} \leq v\}$$

using the **Cramér–von Mises** (CVM) statistic defined as

$$S_n = \sum_{i=1}^{n} \Big(\widehat{C}_n(\hat{U}_{i1}, \hat{U}_{i2}) - C_{\hat{\theta}_n}(\hat{U}_{i1}, \hat{U}_{i2})\Big)^2.$$

Large values of S_n lead to the rejection of H_0.

Questions/exercises

- What's the meaning of likelihood function?
- Why does MLE work only for models with pdfs?
- Why do we require $\mathcal{D}_n$ to be a random sample in the MLE method?
- Can we apply PMLE to non-separable models?
- Why did we assume that Σ is invertible in the MLE example for multivariate normal?