Pseudo-likelihood Information Criteria in Copula Model Selection

Aibat Kossumov

Charles University in Prague
Faculty of Mathematics and Physics
Department of Probability and Mathematical Statistics

July 17, 2025

Sklar's Theorem for Bivariate Distributions

- A function $C : [0,1]^2 \to [0,1]$, is called a **copula** if it is a cumulative distribution function (cdf) with uniform marginals on $[0,1]$.
- **Sklar's Theorem**. Assume that H is the cdf of an **absolutely continuous distribution**. Then there exists a **unique** copula C such that

$$H(x_1, x_2) = C(F_1(x_1), F_2(x_2)),$$

  where F_1 and F_2 are the marginal cdfs of H.
- Sklar's theorem allows us to separate **the dependence structure** from **the structure of the marginal distributions**.
- Since we consider only absolutely continuous distributions, densities can be derived from Sklar's theorem as

$$h(x_1, x_2) = c(F_1(x_1), F_2(x_2)) \prod_{k=1}^{2} f_k(x_k).$$

Fully Parametric Approach

- Assume that the copula C can be parametrized by a real vector $\boldsymbol{\theta}$, and each marginal cdf F_i can be parametrized by a real vector $\gamma(i)$ for $i = 1, 2$. Then the joint density can be written as

$$h_{(\boldsymbol{\theta},\boldsymbol{\gamma})}(x_1, x_2) = c_{\boldsymbol{\theta}}(F_{\gamma(1)}(x_1), F_{\gamma(2)}(x_2)) \prod_{k=1}^{2} f_{\gamma(k)}(x_k),$$

  where $\boldsymbol{\gamma} = (\boldsymbol{\gamma}(1)^T, \boldsymbol{\gamma}(2)^T)^T$.
- In statistics, we observe $\mathcal{X}_n = \{\boldsymbol{x}_i\}_{i=1}^n \overset{\text{i.i.d}}{\sim} h^0$, where h^0 denotes the unknown data-generating density.
- Let $B \in \mathbb{N}$ denote the number of considered parametric families

$$\mathcal{H}_b = \{h_{(\boldsymbol{\theta}_b,\boldsymbol{\gamma}_b)} : (\boldsymbol{\theta}_b, \boldsymbol{\gamma}_b) \in \Theta_b \times \Gamma_b\} \text{ for } b = 1, \ldots, B.$$

- By fitting each family $\mathcal{H}_b$ to the data $\mathcal{X}_n$, one can obtain the MLEs $\left(\widehat{\boldsymbol{\theta}}_b, \widehat{\boldsymbol{\gamma}}_b\right)$.
- Our goal is to rank the fitted models to identify the most suitable one.
- To do so, we compute the Akaike Information Criterion (AIC) for each model:

$$\mathrm{AIC}_b = 2 \cdot \left[\ell_b\left(\widehat{\boldsymbol{\theta}}_b, \widehat{\boldsymbol{\gamma}}_b\right) - \dim(\boldsymbol{\theta}_b, \boldsymbol{\gamma}_b)\right].$$

Challenges in the Parametric Setting

The use of AIC in the parametric setting is theoretically justified, but fully parametric models have limitations:

- Joint estimation of all parameters, $\boldsymbol{\theta}$ and $\boldsymbol{\gamma} = (\boldsymbol{\gamma}(1)^T, \boldsymbol{\gamma}(2)^T)^T$, can be computationally expensive.
- The estimates of the dependence parameters $\boldsymbol{\theta}$ are sensitive to the choice of marginal models.

Semiparametric Approach

- Assume that we are only interested in the reliable estimation of the dependence parameters $\boldsymbol{\theta}$, and we don't want to make any parametric assumptions about marginals.
- Instead of fitting models to the **independent observations** $\mathcal{X}_n \subset \mathbb{R}^2$, we fit only copula models to the **dependent pseudo-observations** ${}^p\mathcal{X}_n \subset [0,1]^2$.
- The transformation $\mathcal{X}_n \longmapsto {}^p\mathcal{X}_n$ is defined by the function

$$\widetilde{\boldsymbol{F}}_n(x_1, x_2) = \left(\widetilde{F}_{n,1}(x_1), \widetilde{F}_{n,2}(x_2)\right),$$

  where $\widetilde{F}_{n,k}$ is the $\frac{n}{n+1}$-rescaled empirical cdf of the kth marginal, for $k = 1, 2$.
- The corresponding pseudo-observations ${}^p\mathcal{X}_n = \{{}^p\boldsymbol{x}_i\}_{i=1}^n$ are then given by ${}^p\boldsymbol{x}_i = \widetilde{\boldsymbol{F}}_n(\boldsymbol{x}_i)$, for all $i = 1, \ldots, n$.
- The **pseudo-log-likelihood** is then defined as

$${}^p\ell(\boldsymbol{\theta}) = \sum_{i=1}^{n} \log[c_{\boldsymbol{\theta}}({}^p\boldsymbol{x}_i)].$$

- The **maximum pseudo-likelihood estimator (MPLE)** is given by ${}^p\widehat{\boldsymbol{\theta}} = \underset{\boldsymbol{\theta}\in\Theta}{\operatorname{argmax}}\ {}^p\ell(\boldsymbol{\theta})$.

Naive Adaptation of AIC to the Semiparametric Case

- The AIC was originally derived in Akaike [1974] from the ”**loss-function perspective**” in the parametric setting:

$$\text{AIC} = 2 \cdot \left[\ell\left(\widehat{\boldsymbol{\theta}}, \widehat{\gamma}\right) - \dim(\boldsymbol{\theta}, \gamma) \right].$$

Naive Adaptation of AIC to the Semiparametric Case

- The AIC was originally derived in Akaike [1974] from the ”**loss-function perspective**” in the parametric setting:
$$\mathrm{AIC} = 2 \cdot \left[\ell\left(\widehat{\boldsymbol{\theta}}, \widehat{\boldsymbol{\gamma}}\right) - \dim(\boldsymbol{\theta}, \boldsymbol{\gamma})\right].$$

- In Grønneberg and Hjort [2014], the authors attempted to extend the same line of reasoning from Akaike [1974] to the semiparametric case. However, their resulting expression is not applicable in most situations.

Naive Adaptation of AIC to the Semiparametric Case

- The AIC was originally derived in Akaike [1974] from the ”**loss-function perspective**” in the parametric setting:

$$\mathsf{AIC} = 2 \cdot \left[\ell\left(\widehat{\boldsymbol{\theta}}, \widehat{\gamma}\right) - \mathsf{dim}(\boldsymbol{\theta}, \gamma)\right].$$

- In Grønneberg and Hjort [2014], the authors attempted to extend the same line of reasoning from Akaike [1974] to the semiparametric case. However, their resulting expression is not applicable in most situations.
- A **naive adaptation** of AIC to the semiparametric case yields

$${}^{p}\mathsf{AIC} = 2 \cdot \left[{}^{p}\ell\left({}^{p}\widehat{\boldsymbol{\theta}}\right) - \mathsf{dim}(\boldsymbol{\theta})\right].$$

- The use of ^{p}AIC was motivated by the **belief** that, in the limit, a continuous connection between AIC and ^{p}AIC **may** exist-but it turns out that **this is not the case**, see Grønneberg and Hjort [2014].
- **Conclusion**: ^{p}AIC is not formally valid for model selection in the semiparametric case. However, it is still commonly used due to its computational simplicity.

Leave-One-Out Copula Information Criterion

Grønneberg and Hjort [2014] introduced the following information criterion from a ”**prediction perspective**”, based on leave-one-out cross validation:

$$\mathrm{xv}_1 = \frac{1}{n}\sum_{i=1}^{n} \log\left[c_{\boldsymbol{\theta}}\left(\widetilde{\boldsymbol{F}}_{(-i)}(\boldsymbol{x}_i)\right)\right]_{\boldsymbol{\theta}={}^{p}\widehat{\boldsymbol{\theta}}_{(-i)}}, \text{ where}$$

- $\widetilde{\boldsymbol{F}}_{(-i)}(x_1, x_2) = \left(\widetilde{F}_{(-i),1}(x_1), \widetilde{F}_{(-i),2}(x_2)\right)$, where $\widetilde{F}_{(-i),k}$ is the $\frac{n-1}{n}$-rescaled empirical cdf of the kth marginal, computed from the sample $\mathcal{X}_n$ excluding $\boldsymbol{x}_i$, for $k = 1, 2$,
- ${}^{p}\widehat{\boldsymbol{\theta}}_{(-i)} = \underset{\boldsymbol{\theta}\in\Theta}{\operatorname{argmax}} \sum_{j\neq i} \log\left[c_{\boldsymbol{\theta}}\left(\widetilde{\boldsymbol{F}}_{(-i)}(\boldsymbol{x}_j)\right)\right]$.

However, since computing xv_1 is computationally expensive, the authors recommend using its approximation, $\mathrm{xv}_{\mathrm{CIC}}$, defined as:

$$\mathrm{xv}_{\mathrm{CIC}} = 2 \cdot \left[{}^{p}\ell\left({}^{p}\widehat{\boldsymbol{\theta}}\right) - \widehat{p} - \widehat{q} - \widehat{r}\right],$$

where $\widehat{p}$, $\widehat{q}$ and $\widehat{r}$ are bias-correction terms (see Section 4 in Grønneberg and Hjort [2014] for details).

Leave-n_v-Out Copula Information Criterion

In the context of **linear model selection**, Shao [1993] showed that the optimal selection procedure is leave-n_v-out cross-validation, where the **the validation set size** n_v is of the same order as the full sample size n, that is, $n_v/n \to 1$ as $n \to \infty$.

$$\mathrm{xv}_{n_v} = \frac{1}{n_v b_n} \sum_{s_v \in \mathcal{T}_n} \sum_{i \in s_v} \log \left[c_{\boldsymbol{\theta}} \left(\widetilde{\boldsymbol{F}}_{(-s_v)}(\boldsymbol{x}_i) \right) \right]_{\boldsymbol{\theta} = {}^p\widehat{\boldsymbol{\theta}}_{(-s_v)}}, \text{ where}$$

- $\mathcal{T}_n$ is a collection of $b_n = O(n)$ subsets of $\{1, \ldots, n\}$, each of size n_v, randomly drawn without replacement. For example, one could set $b_n = \lfloor 0.8n \rfloor$ and $n_v = n - n^{0.9}$.
- $s_v \in \mathcal{T}_n$ is the set of indices for the n_v validation observations.
- $\widetilde{\boldsymbol{F}}_{(-s_v)}(x_1, x_2) = \left(\widetilde{F}_{(-s_v),1}(x_1), \widetilde{F}_{(-s_v),2}(x_2) \right)$, where $\widetilde{F}_{(-s_v),k}$ is the $\frac{(n-n_v)}{(n-n_v)+1}$-rescaled empirical cdf of the kth marginal, computed from the sample $\mathcal{X}_n$ excluding $\{\boldsymbol{x}_i : i \in s_v\}$, for $k = 1, 2$,
- ${}^p\widehat{\boldsymbol{\theta}}_{(-s_v)} = \underset{\boldsymbol{\theta} \in \Theta}{\operatorname{argmax}} \sum_{j \notin s_v} \log \left[c_{\boldsymbol{\theta}} \left(\widetilde{\boldsymbol{F}}_{(-s_v)}(\boldsymbol{x}_j) \right) \right]$.

Summary of Information Criteria

In the semiparametric setting of copula model selection, we consider the following four information criteria:

1. Naive Akaike Information Criterion: ${}^{p}\mathrm{AIC}$ **(not valid, easy to compute)**
2. Leave-One-Out Cop. Information Criterion: xv_1 **(valid, computationally expensive)**
3. Approximate Leave-One-Out Criterion: $\mathrm{xv}_{\mathrm{CIC}}$ **(valid, moderately expensive to compute)**
4. Leave-n_v-Out Copula Information Criterion: xv_{n_v} **(?, computationally expensive)**

In the study Jordanger and Tjøstheim [2014], the authors compared $\mathrm{xv}_{\mathrm{CIC}}$ and ${}^{p}\mathrm{AIC}$.

Setup of the Simulation Study

The simulation study is based on the following settings:

- One-dimensional parametric copula families: Clayton, Gumbel, Joe, Frank, Gaussian.
- Each copula was parametrized using different values of Kendall's tau τ.
- In each simulation scenario, we conducted 5000 replications.

Hit rates

IC	Clayton	Gumbel	Joe	Frank	Gaussian
AIC	62.5 ± 1.3	16.2 ± 1.0	61.9 ± 1.3	33.6 ± 1.3	18.2 ± 1.1
xv_1	62.6 ± 1.3	16.2 ± 1.0	61.9 ± 1.3	33.6 ± 1.3	18.1 ± 1.1
xv_{CIC}	60.7 ± 1.3	12.4 ± 0.9	64.9 ± 1.3	30.8 ± 1.3	20.8 ± 1.1
xv_{n_v}	43.9 ± 1.4	26.9 ± 1.2	59.6 ± 1.4	28.1 ± 1.2	16.8 ± 1.0

Table: Hit rates ($n = 100$, $\tau = 0.10$) are shown with 95% confidence intervals, and all values are expressed as percentages.

IC	Clayton	Gumbel	Joe	Frank	Gaussian
AIC	90.6 ± 0.8	49.5 ± 1.4	79.4 ± 1.1	60.6 ± 1.3	49.2 ± 1.4
xv_1	90.6 ± 0.8	49.5 ± 1.4	79.4 ± 1.1	60.6 ± 1.3	49.2 ± 1.4
xv_{CIC}	87.2 ± 0.9	46.4 ± 1.4	83.4 ± 1.0	63.2 ± 1.3	49.2 ± 1.4
xv_{n_v}	90.2 ± 0.8	51.5 ± 1.4	77.7 ± 1.1	61.9 ± 1.3	47.0 ± 1.4

Table: Hit rates ($n = 200$, $\tau = 0.20$) are shown with 95% confidence intervals, and all values are expressed as percentages.

Coincidence Percentages for Weak Dependence

	n	$\tau = 0.05$	$\tau = 0.1$	$\tau = 0.15$	$\tau = 0.2$	All
AIC & xv_1	100	99.77	99.90	99.92	99.94	99.88
AIC & xv_1	200	99.93	99.99	99.99	100.00	99.97
AIC & $\mathrm{xv}_{\mathrm{CIC}}$	100	79.85	85.63	88.52	89.74	85.93
AIC & $\mathrm{xv}_{\mathrm{CIC}}$	200	86.10	91.58	93.09	93.75	91.13
AIC & xv_{n_v}	100	47.86	67.71	80.67	87.44	70.91
AIC & xv_{n_v}	200	59.30	83.14	91.77	94.65	82.21

Table: Coincidence of AIC with cross-validation based information criteria, with all values expressed as percentages.

Conclusion

- The proposed method xv_{n_v} was still unable to beat the well-known AIC.
- For larger sample sizes or stronger dependence, all considered criteria are able to select the true copula model reliably.
- All criteria perform poorly under small sample sizes and weak dependence.
- Regardless of the sample size and the value of Kendall's tau, the most challenging copulas to identify for all criteria are Gaussian and Gumbel.
- As an interesting secondary finding, it was shown that for all considered values τ, the closest method to AIC is xv_1.
- Under weaker dependence $\tau \in \{0.05, 0.1, 0.15\}$, xv_{CIC} is much closer to AIC than xv_{n_v}.
- In the specific case when the true copula model is Gumbel, the proposed xv_{n_v} outperformed the other criteria (in terms of hit rates and their confidence intervals) for all considered combinations of τ and n.

References

H. Akaike. A new look at the statistical model identification. *IEEE Transactions on automatic control*, 19:716–723, 1974.

S. Grønneberg and N. L. Hjort. The copula information criteria. *Scandinavian Journal of Statistics*, 41:436–459, 2014.

L. A. Jordanger and D. Tjøstheim. Model selection of copulas: AIC versus a cross validation copula information criterion. *Statistics and Probability Letters*, 92:249–255, 2014.

J. Shao. Linear model selection by cross-validation. *Journal of the American Statistical Association*, 88:486–494, 1993.

FULL SIMULATION STUDY

kossumov@karlin.mff.cuni.cz