# Using Second-Order Probabilities to Make Maximum Entropy Approach to Copulas More Reasonable

Hung T. Nguyen[1,2], Vladik Kreinovich[3], and Berlin Wu[4]

[1]Department of Mathematical Sciences
New Mexico State University
Las Cruces, NM 88003, USA, hunguyen@nmsu.edu
[2]Faculty of Economics, Chiang Mai University, Thailand
[3]Department of Computer Science, University of Texas at El Paso
500 W. University, El Paso, TX 79968, USA, vladik@utep.edu
[4]Department of Mathematical Sciences, National Chengchi University
Taipei 116, Taiwan, berlin@nccu.edu.tw

**Abstract**

Copulas are a general way of describing dependence between two or more random variables. When we only have partial information about the dependence, i.e., when several different copulas are consistent with our knowledge, it is often necessary to select one of these copulas. A frequently used method of selecting this copula is the maximum entropy approach, when we select a copula with the largest entropy. However, in some cases, the maximum entropy approach leads to an unreasonable selection – e.g., even if we know that the two random variables are positively correlated, the maximum entropy approach completely ignores this information. In this paper, we show how to properly modify the maximum entropy approach so that it will lead to more reasonable results: by applying this approach not to the probabilities themselves, but to "second order" probabilities – i.e., probabilities of different probability distributions.

## 1 Maximum Entropy Approach to Selecting a Copula: Description, Successes, and Limitations

**Copulas: brief reminder.** In many practical situations, we know the cumulative distribution functions $F_1(x_1) = P(X_1 \le x_1)$ and $F_2(x_2) = P(X_2 \le x_2)$

of the two random variables $X_1$ and $X_2$, and we need to find the cumulative distribution function (cdf) $F(x_1, x_2) = P(X_1 \le x_1 \,\&\, X_2 \le x_2)$ corresponding to their joint distribution.

In general, a joint cdf can be described as $F(x_1, x_2) = C(F_1(x_1), F_2(x_2))$ for some function $C(u, v)$. This function is known as a *copula*; see, e.g., [2, 7].

Similarly, a multi-dimensional copula can be defined as a function $C(u_1, \ldots, u_n)$ for which

$$F(x_1, \ldots, x_n) \stackrel{\text{def}}{=} P(X_1 \le x_1 \,\&\, \ldots \,\&\, X_n \le x_n) = C(F_1(x_1), \ldots, F_n(x_n)).$$

*Comment.* From the mathematical viewpoint, a copula can be viewed as a cdf for a joint distribution of two random variables $U$ and $V$ which are both uniformly distributed on the interval $[0, 1]$. Similarly, a multi-D copula can be viewed as a a joint distribution of $n$ random variables $U_1, \ldots, U_n$ all of which are uniformly distributed on the interval $[0, 1]$.

When the copula is a differentiable function, we can determine the probability density function (pdf) corresponding to this distribution, as $c(u, v) = \dfrac{\partial^2 C(u, v)}{\partial u\, \partial v}$ or, in multi-D case, as $c(u_1, \ldots, u_n) = \dfrac{\partial^n C(u_1, \ldots, u_n)}{\partial u_1 \ldots \partial u_n}$.

**Need to select a copula under uncertainty.** In many practical situations, we only have partial information about the joint probability distribution. In other words, we have several copulas which are consistent with the given knowledge. In this case, it is often desirable to select a single copula.

**Maximum entropy (Laplace) approach.** The general need for selecting probabilities under partial knowledge has been recognized for a long time. The first approach to this problem was Laplace's Principle of Indifference, according to which, if we have no information about which of $n$ alternatives are more probable and which are less probable, it is reasonable to assign the same probability $\dfrac{1}{n}$ to each of these alternatives.

This principle can be naturally generalized to the continuous case, in which case, out of all possible probability distributions, we select the one for which the entropy $S \stackrel{\text{def}}{=} -\int f(x) \cdot \ln(f(x))\, dx$ attains its largest possible value, where $f(x)$ is the probability density function; see, e.g., [3]. For the discrete case, this optimization leads exactly to equal probabilities.

In terms of copulas, this means selecting a copula for which the entropy $-\int c(u, v) \cdot \ln(c(u, v))\, du\, dv$ – or, in the multi-D case,

$$-\int c(u_1, \ldots, u_n) \cdot \ln(c(u_1, \ldots, u_n))\, du_1 \ldots du_n,$$

attains the largest possible value.

**In many practical situations, the maximum entropy approach to selecting a copula works well.** In many case, the maximum entropy approach works well. For example, if we have no information about the dependence of the two random variables, then maximizing the above integral under the constraint that the overall probability is equal to 1 (i.e., that $\int c(u,v)\,du\,dv = 1$) leads to the uniform distribution $c(u,v) = 1$. For this uniform distribution, $C(u,v) = u \cdot v$; thus, $F(x_1, x_2) = F_1(x_1) \cdot F_2(x_2)$.

In probabilistic terms, the above equality means that the variables $X_1$ and $X_2$ are independent. In other words, if we have no information about the dependence between the random variables $X_1$ and $X_2$, i.e., we have no reason to believe that these variables are positively correlated – or that they are negatively correlated – then the maximum entropy approach suggests that we assume that these variables are independent. This sounds reasonable.

Similarly, in the multi-D case, for the probability density $c(u_1, \ldots, u_n) = \dfrac{\partial^n C(u_1, \ldots, u_n)}{\partial u_1 \ldots \partial u_n}$, we get a uniform distribution $c(u_1, \ldots, u_n) = 1$ for which $C(u_1, \ldots, u_n) = u_1 \cdot \ldots \cdot u_n$ and thus, $F(x_1, \ldots, x_n) = F_1(x_1) \cdot \ldots \cdot F_n(x_n)$. This too is reasonable.

**What if there is a positive (non-negative) dependence?** Independence of two events $A$ and $B$ means that the probability of $A$ does not change if we know $B$, i.e., that the conditional probability $P(A \mid B)$ of $A$ under the condition $B$ is equal to the probability $P(A)$: $P(A \mid B) = P(A)$.

Since $P(A \mid B) = \dfrac{P(A \,\&\, B)}{P(B)}$, the condition $P(A \mid B) = P(A)$ means that $\dfrac{P(A \,\&\, B)}{P(B)} = P(A)$, i.e., that $P(A \,\&\, B) = P(A) \cdot P(B)$.

In some cases, we know that the presence of $B$ increases the probability of $A$, i.e., that $A$ and $B$ are positively dependent. In precise terms, this means that $P(A \mid B) > P(A)$, i.e., that $\dfrac{P(A \,\&\, B)}{P(B)} > P(A)$ and $P(A \,\&\, B) > P(A) \cdot P(B)$. A more general case is *non-negative* dependence, when $P(A \,\&\, B) \ge P(A) \cdot P(B)$.

Similarly, we can consider the case of *negative dependence*, when $P(A \mid B) < P(A)$ and, thus, $P(A \,\&\, B) < P(A) \cdot P(B)$, or, more generally, *non-positive* dependence, when $P(A \,\&\, B) \le P(A) \cdot P(B)$.

In the case of copulas, non-negative dependence means that $C(u,v) \ge u \cdot v$, and non-positive dependence means that $C(u,v) \le u \cdot v$.

If we only know that the copula corresponds to non-negative dependence, which copula should we select? If we only know that the copula corresponds to non-positive dependence, which copula should we select?

**Limitation of the maximum entropy approach.** To solve both problems, we can use the maximum entropy approach:

- in the first problem, out of all copulas which satisfy the constraint $C(u,v) \ge u \cdot v$ for all $u$ and $v$, we select the copula for which the entropy is the largest (under this constraint);
- in the second problem, out of all copulas which satisfy the constraint $C(u,v) \le u \cdot v$ for all $u$ and $v$, we select the copula for which the entropy

is the largest (under this constraint).

One can easily see in both situations, we select the same product copula $C(u,v) = u \cdot v$ corresponding to independence – since the product copula satisfies both constraints and has the largest entropy among *all* copulas – and thus, has the largest entropy among all copulas which satisfy each of the two constraints.

This selection is counter-intuitive: we assumed, crudely speaking, that the correlation between $X_1$ and $X_2$ is non-negative, and the maximum entropy approach ignored this information altogether.

**What we do in this paper.** In this paper, we show how the maximum entropy approach can be modified, so that we will get more reasonable conclusions.

# 2 How to Modify the Maximum Entropy Approach: Main Idea and Resulting Applications

**Analysis of the problem.** Let us recall that we are analyzing the probabilistic models of real-world events. Before probabilistic models, scientists used deterministic models. To come up a deterministic model, they had to select, from all possible trajectories, the most probable one. In comparison with the resulting limited deterministic model, a probabilistic model is more adequate – it allows us to take into account that for the same initial condition, several different trajectories are possible.

**Resulting idea.** Let us use a similar idea to overcome the above-described limitation of the maximum entropy approach. This limitation comes from the fact that we select a single probability distribution – or, in the copula case, a single copula. A more adequate idea is to select the class of possible copulas – and to assign probabilities to different copulas. To assign probabilities to different copulas, we can again use the maximum entropy approach.

As a result, instead of a single copula, we get different copulas with different probabilities. To find the resulting probability $C(u,v)$, we can then use the formula of complete probability, i.e., take the average over all possible copulas.

Let us apply this idea of using "second order" probabilities – i.e., probabilities of different probability distributions – to our problems.

**Let us apply this idea to the case of non-negative dependence.** For each $u$ and $v$, we have $\max(u+v-1, 0) \le C(u,v) \le \min(u,v)$. If we additionally know that the variables are non-negatively dependent, then we also get $C(u,v) \ge u \cdot v$, so $u \cdot v \le C(u,v) \le \min(u,v)$.

In this case, the set of possible values of $C(u,v)$ is the interval $[u \cdot v, \min(u,v)]$. To select a probability distribution on this interval, we will use the Maximum Entropy approach. Since we have no information about which values from this interval are more probable and which values are less probable, the maximum entropy approach results in a uniform distribution on this interval.

For a uniform distribution on an interval $[a, b]$, its expected value is equal to the interval's midpoint $\dfrac{a+b}{2}$. Thus, in our case, we get the following copula:

$$C(u, v) = \frac{u \cdot v + \min(u, v)}{2}.$$

*Comment.* This formula indeed defines a copula, since it is a convex combination of two copulas, and it is known that a convex combination of copulas is always a copula; see, e.g., [2, 7].

**Discussion.** One can see that, in contrast to the usual maximum entropy approach, the resulting copula is different from the independent case – and reflects the fact that we have non-negative dependence.

**Non-negative dependence: multi-D case.** In case of several random events, non-negative dependence means that $P(A_1 \,\&\, \ldots \,\&\, A_n) \geq P(A_1) \cdot \ldots \cdot P(A_n)$. In particular, for a multi-D copula, this means that $C(u_1, \ldots, u_n) \geq u_1 \cdot \ldots \cdot u_n$.

In general, a multi-D copula satisfies the inequality

$$\max(u_1 + \ldots + u_n - (n-1), 0) \leq C(u_1, \ldots, u_n) \leq \min(u_1, \ldots, u_n).$$

Thus, copulas corresponding to non-negative dependence satisfy the inequality

$$u_1 \cdot \ldots \cdot u_n \leq C(u_1, \ldots, u_n) \leq \min(u_1, \ldots, u_n).$$

The set of possible values of $C(u_1, \ldots, u_n)$ form an interval

$$[u_1 \cdot \ldots \cdot u_n, \min(u_1, \ldots, u_n)].$$

The above maximum entropy approach results in a uniform distribution on this interval. Thus, we get the following copula:

$$C(u_1, \ldots, u_n) = \frac{u_1 \cdot \ldots \cdot u_n + \min(u_1, \ldots, u_n)}{2}.$$

**What if we have non-positive dependence.** In the case of non-positive dependence, the condition $C(u, v) \leq u \cdot v$ implies that $\max(u + v - 1, 0) \leq C(u, v) \leq u \cdot v$. In this case, the set of possible values of $C(u, v)$ is the interval $[\max(u + v - 1, 0), u \cdot v]$. Here also, the maximum entropy approach results in a uniform distribution on this interval. Thus, we get the following copula:

$$C(u, v) = \frac{\max(u + v - 1, 0) + u \cdot v}{2}.$$

Here also, the resulting copula is different from the independent case – and reflects the fact that we have non-positive dependence.

*Comment on multi-D case.* In the multi-D case, non-positive dependence means $C(u_1, \ldots, u_n) \le u_1 \cdot \ldots \cdot u_n$. Thus, possible values of $C(u_1, \ldots, u_n)$ form an interval

$$[\max(u_1 + \ldots + u_n - (n-1), 0), u_1 \cdot \ldots \cdot u_n].$$

The above maximum entropy approach results in a uniform distribution on this interval. Thus, we get the following expected value

$$C(u_1, \ldots, u_n) = \frac{\max(u_1 + \ldots + u_n - (n-1), 0) + u_1 \cdot \ldots \cdot u_n}{2}.$$

However, in contrast to the case of $n = 2$, the lower bound

$$\max(u_1 + \ldots + u_n - (n-1), 0)$$

is not a copula, and thus, its use does not necessarily lead to a joint probability distribution.

*General comment.* It is worth mentioning that in the case when we have no information about the dependence, i.e., when we only know that $C(u, v)$ belongs to the interval $[\max(u + v - 1, 0), \min(u, v)]$, the modified maximum entropy approach leads to

$$C(u, v) = \frac{\max(u + v - 1, 0) + \min(u, v)}{2},$$

which is general, different from the copula $C(u, v) = u \cdot v$ which results from the usual maximum entropy approach.

# 3 How to Describe Strong Positive Dependence? Weak Positive Dependence? A Natural Idea

**How to describe strong and weak positive dependence? Main idea.** The above idea shows that for the case of positive dependence, the "average" dependence is described by the copula

$$C_{\rm av}(u, v) = \frac{u \cdot v + \min(u, v)}{2}.$$

The larger the value $C(u, v)$, the larger the dependence.

It is reasonable to say that strong dependence means that $C(u, v)$ is larger than this average value, while weak positive dependence means that $C(u, v)$ is smaller than this average.

Let us analyze which copulas correspond to these notions.

**Copula corresponding to strong positive dependence.** In the case of strong positive dependence, we have

$$\frac{u \cdot v + \min(u, v)}{2} \le C(u, v) \le \min(u, v).$$

Similarly, we conclude that we have a uniform distribution on this interval and thus, the resulting copula is equal to its midpoint

$$C(u,v)=\frac{1}{4}\cdot u\cdot v+\frac{3}{4}\cdot\min(u,v).$$

We can similarly consider very strong (= stronger than strong), for which we get

$$C(u,v)=\frac{1}{8}\cdot u\cdot v+\frac{7}{8}\cdot\min(u,v),$$

and somewhat strong (= between average and strong), for which we get

$$C(u,v)=\frac{3}{8}\cdot u\cdot v+\frac{5}{8}\cdot\min(u,v),$$

**Copula corresponding to weak positive dependence.** In the case of weak positive dependence, we have

$$u\cdot v\le C(u,v)\le\frac{u\cdot v+\min(u,v)}{2}.$$

Similarly, we conclude that we have a uniform distribution on this interval and thus, the resulting copula is equal to its midpoint

$$C(u,v)=\frac{3}{4}\cdot u\cdot v+\frac{1}{4}\cdot\min(u,v).$$

We can similarly consider very weak (= weaker than weak), for which we get

$$C(u,v)=\frac{7}{8}\cdot u\cdot v+\frac{1}{8}\cdot\min(u,v),$$

and somewhat weak (= between average and weak), for which we get

$$C(u,v)=\frac{5}{8}\cdot u\cdot v+\frac{3}{8}\cdot\min(u,v),$$

**Copulas corresponding to strong and weak negative dependence.** Similarly, for negative dependence, the further away the value $C(u,v)$ is from $u\cdot v$ – i.e., in this case, the smaller $C(u,v)$ – the larger the dependence.

Thus, strong negative dependence means that

$$\max(u+v-1,0)\le C(u,v)\le\frac{\max(u+v-1,0)+u\cdot v}{2}$$

and thus, the average is equal to the midpoint of the corresponding interval:

$$C(u,v)=\frac{3}{4}\cdot\max(u+v-1,0)+\frac{1}{4}\cdot u\cdot v.$$

Weak negative dependence means that

$$\frac{\max(u+v-1,0)+u\cdot v}{2} \le C(u,v) \le u\cdot v$$

and thus, the average is equal to the midpoint of the corresponding interval:

$$C(u,v) = \frac{1}{4}\cdot \max(u+v-1,0) + \frac{3}{4}\cdot u\cdot v.$$

Similarly, we can have very strong negative correlation

$$C(u,v) = \frac{7}{8}\cdot \max(u+v-1,0) + \frac{1}{8}\cdot u\cdot v,$$

somewhat strong negative correlation

$$C(u,v) = \frac{5}{8}\cdot \max(u+v-1,0) + \frac{3}{8}\cdot u\cdot v,$$

very weak negative correlation

$$C(u,v) = \frac{1}{8}\cdot \max(u+v-1,0) + \frac{7}{8}\cdot u\cdot v,$$

and somewhat weak negative correlation

$$C(u,v) = \frac{3}{8}\cdot \max(u+v-1,0) + \frac{5}{8}\cdot u\cdot v.$$

*Comment.* Similar formulas can be obtained in the multi-D case. For example, strong positive dependence corresponds to

$$C(u_1,\ldots,u_n) = \frac{1}{4}\cdot u_1\cdot\ldots\cdot u_n + \frac{3}{4}\cdot \min(u_1,\ldots,u_n);$$

very strong positive dependence corresponds to

$$C(u_1,\ldots,u_n) = \frac{1}{8}\cdot u_1\cdot\ldots\cdot u_n + \frac{7}{8}\cdot \min(u_1,\ldots,u_n);$$

and somewhat strong positive dependence corresponds to

$$C(u_1,\ldots,u_n) = \frac{3}{8}\cdot u_1\cdot\ldots\cdot u_n + \frac{5}{8}\cdot \min(u_1,\ldots,u_n).$$

**Relation to Hurwicz optimism-pessimism criterion.** The resulting formulas are similar to another way of selecting an alternative under uncertainty: the optimism-pessimism criterion proposed by a Nobelist L. Hurwicz [1, 5, 6]: when we only know bounds $\underline{u}$ and $\overline{u}$ on a value $u$ (for which the larger the value, the better for us), then we should select the value $\alpha\cdot\overline{u}+(1-\alpha)\cdot\underline{u}$, where $\alpha\in[0,1]$ describes the decision maker's degree of optimism:

- the value $\alpha = 1$ characterizes an optimist who only takes into account the best-case scenario, with value $\overline{u}$, and ignores the possibility of the worst-case situations;
- the value $\alpha = 0$ characterizes a pessimist who only takes into account the worst-case scenario, with value $\underline{u}$, and ignores the possibility of the best-case situations;
- finally, values $\alpha$ between 0 and 1 characterize realists who take into account that both good and bad situations are possible.

For the case of positive dependence, if we apply Hurwicz criterion instead of maximum entropy, we get

$$C(u, v) = \alpha \cdot \min(u, v) + (1 - \alpha) \cdot u \cdot v.$$

The above maximum entropy description of positive dependence corresponds to $\alpha = 1/2$, strong positive to $\alpha = 3/4$, etc.

*Comment.* Similar operations were described in [4].

**Acknowledgments.** This work was supported in part by the National Science Foundation grants HRD-0734825 and HRD-1242122 (Cyber-ShARE Center of Excellence) and DUE-0926721.

# References

[1] L. Hurwicz, *Optimality Criteria for Decision Making Under Ignorance*, Cowles Commission Discussion Paper, Statistics, No. 370, 1951.

[2] P. Jaworski, F. Durante, W. K. Härdle, and T. Ruchlik (eds.), *Copula Theory and Its Applications*, Springer Verlag, Berlin, Heidelberg, New York, 2010.

[3] E. T. Jaynes, *Probability Theory: The Logic of Science*, Cambridge University Press, Cambridge, Massachisetts, 2003.

[4] V. Kreinovich, "Dynamic Fuzzy Logic Leads to More Adequate 'And' and 'Or' Operations", *Proceedings of the 2011 Sixth International Conference on Soft Computing, Computing with Words and Perceptions in System Analysis, Decision and Control ICSCCW'2011*, Antalya, Turkey, September 1–2, 2011, pp. 21–29.

[5] V. Kreinovich, "Decision Making under Interval Uncertainty (and beyond)", In: P. Guo and W. Pedrycz (eds.), *Human-Centric Decision-Making Models for Social Sciences*, Springer Verlag, 2014, pp. 163–193.

[6] R. D. Luce and R. Raiffa, *Games and Decisions: Introduction and Critical Survey*, Dover, New York, 1989.

[7] R. B. Nelsen, *An Introduction to Copulas*, Springer Verlag, Berlin, Heidelberg, New York, 1999.

[8] D. J. Sheskin, *Handbook of Parametric and Nonparametric Statistical Procedures*, Chapman and Hall/CRC Press, Boca Raton, Florida, 2011.