How to Explain the Definition of Stochastic Affiliation to Economics Students

Tonghui Wang[1], Olga Kosheleva[2], and Vladik Kreinovich[3]
[1]Department of Mathematical Sciences
New Mexico State University
Las Cruces, NM 80003, USA
twang@nmsu.edu
[2]College of Education
[3]Department of Computer Science
University of Texas at El Paso
500 W. University
El Paso, TX 79968, USA
olgak@utep.edu, vladik@utep.edu

Abstract

To formally describe the intuitive idea of "positive correlation" between two quantities, it is often helpful to use the notion of stochastic affiliation. While this notion is useful, its usual definition is not intuitively clear – which make it difficult to explain this notion to, e.g., economics students. To help students understand this notion, in this paper, we show how the notion of stochastic affiliation can be explained in clear probabilistic terms.

1 Formulation of the Problem: The Notion of Stochastic Affiliation is Difficult to Explain

Positive correlation as an important intuitive idea. In the statistical analysis of economic situations, it is often important to check which pairs of quantities x and y are "positively correlated" – in the sense that:

- the increase in x makes it more probable that y increases, and, vice versa,
- the increase in y makes it more probable that x increases.

The notion of stochastic affiliation. One of the most useful formalizations of the idea of positive correlation is the notion of stochastic affiliation; see, e.g., [1, 2]. Two random variables X and Y whose joint distribution is described by a probability density function $f(x, y)$ are called *stochastically affiliated* if for every $x_1 \le x_2$ and $y_1 \le y_2$, we have $f(x_1, y_2) \cdot f(x_2, y_1) \le f(x_1, y_1) \cdot f(x_2, y_2)$.

A pedagogical problem. The above definition leads to many interesting mathematical results (see, e.g., [1, 2]). However, this definition is difficult to understand: why do we compare the two products? This lack of intuitive understanding makes it especially difficult to teach this concept to non-mathematics students – e.g., to economics students.

What we do in this paper. In this paper, we explain how the above intuitive idea of positive correlation leads to the above formal definition. This explanations makes the definition of stochastic affiliation easier to understand for economics students.

2 Stochastic Affiliation: Our Explanation

Main idea behind the proposed explanation. Suppose that we have selected values $x_1 \le x_2$ and we consider only situations in which the variable X takes one of these two values. Suppose also that we have selected values $y_1 \le y_2$ and we consider only situations in which the variable Y takes one of these two values.

For each of the two values $y = y_1$ and $y = y_2$, we can consider the probabilities $p(X = x_1 \,|\, Y = y)$ and $p(X = x_2 \,|\, Y = y)$. Positive correlation means that when we increase the y-value from y_1 to y_2, then the probability of the larger x-value also increases, i.e., that

$$p(x_2 \,|\, Y = y_2) \ge p(x_2 \,|\, Y = y_1).$$

Let us show that this intuitively clear definition leads to the original (seemingly unclear) definition of stochastic affiliation.

Intuitive meaning of conditional probability: reminder. The above idea is based on the notion of conditional probability. Thus, before we show how the above idea can lead to a good understanding of the notion of stochastic affiliation, let us first recall the intuitive understanding of conditional probability. By definition, conditional probability $P(A \,|\, B)$ of an event A under the condition B is defined as the ratio $\dfrac{P(A \,\&\, B)}{P(B)}$.

To understand why this formal definition makes intuitive sense let us recall what is probability in the first place. The probability $P(A)$ of any event A can be described as the fraction of cases in which A is true, i.e., as $p(A) \approx \dfrac{N(A)}{N}$, where N is the total number of cases, and $N(A)$ is the number of cases when A is true. The larger the sample size N, the more accurate is this approximation.

Based on this approximate formula, we conclude that when we know the total number of cases N and we know the probability $p(A)$ that A is true, we can then estimate the number of cases $N(A)$ in which A is true as $N(A) \approx N \cdot p(A)$.

The definition of the conditional probability $p(A \mid B)$ is similar to the definition of the probability, except that instead of considering all possible cases, we only consider cases in which B is true. In order words, $p(A \mid B) \approx \dfrac{N(A \,\&\, B)}{N(B)}$. Here, as we have mentioned before, $N(A \,\&\, B) \approx N \cdot p(A \,\&\, B)$ and $N(B) \approx N \cdot p(B)$. Substituting these expression into the above formula for the conditional probability, we get $p(A \,\&\, B) \approx \dfrac{N \cdot p(A \,\&\, B)}{N \cdot p(B)}$. Canceling the common factor N in both numerator and denominator, we conclude that $p(A \mid B) \approx \dfrac{p(A \,\&\, B)}{p(B)}$.

The larger the same size N, the more accurate is this formula, i.e., the smaller the bound on the difference $p(A \mid B) - \dfrac{p(A \,\&\, B)}{p(B)}$ between the left-hand and right-hand sides of the above approximate equality. Since this difference does not depend on N at all, this means that this difference is smaller than an arbitrarily small positive number – i.e., equal to 0.

From the above idea to the formal definition of stochastic affiliation. By definition of the probability density function, for each pair (x, y), the probability that X is between x and $x+dx$ and that Y is between y and $y+dy$ is equal to $f(x, y) \cdot dx \cdot dy$. Thus, the conditional probability that X is between x and $x + dx$ under the condition that Y is between y and $y + dy$ is equal to the ratio $\dfrac{f(x, y)\, dx \cdot dy}{p(y \le Y \le y + dy)}$. By definition of the marginal probability density $f_Y(y)$, the probability in the denominator is equal to $f_Y(y) \cdot dy$. Thus, the desired conditional probability $p_y(X = x) \stackrel{\text{def}}{=} p(x \mid y)$ is equal to the ratio $\dfrac{f(x, y) \cdot dx}{f_Y(y)}$. In particular:

- the probability $p_y(X = x_1)$ to get x_1 is equal to $\dfrac{f(x_1, y) \cdot dx}{f_Y(y)}$, and
- the probability $p_y(X = x_2)$ to get x_2 is equal to $\dfrac{f(x_2, y) \cdot dx}{f_Y(y)}$.

If we also limit ourselves to two possible values x_1 and x_2 of the variable X, then the resulting probability $p_y(x_2) \stackrel{\text{def}}{=} p(X = x_2 \mid Y = y \,\&(X = x_1 \vee X = x_2))$ of x_2 is equal to

$$p_y(X = x_2 \mid (X = x_1 \vee X = x_2)) = \frac{p_y(X = x_2)}{p_y(X = x_1 \vee X = x_2)} =$$

$$\frac{p_y(X = x_1)}{p_y(X = x_1) + p_y(X = x_2)}.$$

Substituting the above expressions for $p_y(X = x_1)$ and $p_y(X = x_2)$ into this formula, we get

$$p(X = x_2 \,|\, Y = y) = \frac{\dfrac{f(x_2, y) \cdot dx}{f_Y(y)}}{\dfrac{f(x_1, y) \cdot dx}{f_Y(y)} + \dfrac{f(x_2, y) \cdot dx}{f_Y(y)}}.$$

Multiplying both numerator and denominator by $f_Y(y)$ and dividing by dx, we get

$$p(X = x_2 \,|\, Y = y) = \frac{f(x_2, y)}{f(x_1, y) + f(x_2, y)}.$$

The requirement that this probability increases with y means that

$$p(X = x_2 \,|\, Y = y_2) \geq p(X = x_2 \,|\, Y = y_1),$$

i.e., that

$$\frac{f(x_2, y_2)}{f(x_1, y_2) + f(x_2, y_2)} \geq \frac{f(x_2, y_1)}{f(x_1, y_1) + f(x_2, y_1)}.$$

This inequality leads to the following inequality between the inverses:

$$\frac{f(x_1, y_2) + f(x_2, y_2)}{f(x_2, y_2)} \leq \frac{f(x_1, y_1) + f(x_2, y_1)}{f(x_2, y_1)},$$

i.e., to

$$\frac{f(x_1, y_2)}{f(x_2, y_2)} + 1 \leq \frac{f(x_1, y_1)}{f(x_2, y_1)} + 1.$$

Subtracting 1 from both sides of this inequality, we get

$$\frac{f(x_1, y_2)}{f(x_2, y_2)} \leq \frac{f(x_1, y_1)}{f(x_2, y_1)}.$$

Multiplying both sides by both (non-negative) denominators, we get the desired inequality

$$f(x_1, y_2) \cdot f(x_2, y_1) \leq f(x_1, y_1) \cdot f(x_2, y_2).$$

The formula that forms the traditional definition of stochastic affiliation has thus been derived.

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, and by a travel grant from Chiang Mai University, Thailand.

The authors are thankful to all the participants of the 7th International Conference of Thailand Econometric Society TES'2014 (Chiang Mai, Thailand, January 8–10, 2014) for valuable discussions.

References

[1] P. R. Milgrom and R. J. Weber, "A Theory of Auctions and Competitive Bidding," *Econometrica*, 1982, Vol. 50, pp. 1089–1122.

[2] Z. Wei, T. Wang, and W. Panichkitkosolkul, "Dependence and association concepts through copulas", In: V.-N. Huynh, V. Kreinovich, and S. Sriboonchitta, *Modeling Dependence in Econometrics*, Springer Verlag, Berlin, Heideelberg, 2014, pp. 113–126.