Société de Calcul Mathématique SA
Outils d'aide à la décision

Increasing probabilistic relationships

Bernard Beauzamy

February 2015

It is quite simple to define what is an increasing real function $y = f(x)$; it satisfies $f(x_1) \leq f(x_2)$ if $x_1 \leq x_2$. But usually, the relationship between two quantities x and y is not deterministic, but probabilistic. So, in this case, what is an "increasing" relationship ?

In a probabilistic setting, at best what has been recorded is a joint law of two variables, for instance number of vehicles passing in a street (traffic) and concentration of some pollutant at the same place. We would like to investigate the question : is the concentration an increasing function of the traffic, whatever this means ?

We have at our disposal a joint law $\varphi(x, y)$ obtained from a certain number of measurements: at some places, at some times, were measured together the traffic and the concentration. Typically, the traffic x is decomposed into bins of width 100 vehicle/hour ; the concentration y is decomposed into bins of width 10 mg/l. These decompositions are arbitrary.

Let $n_{i,j}$ be the number of observations which fall into the bin $x = i, y = j$, and let $N = \sum_{i,j} n_{i,j}$ be the total number of observations. The joint law is the collection of all numbers $n_{i,j}$.

Definition

We will say that there is an increasing probabilistic relationship between the two variables T and C (traffic and concentration) if, for all $t_1 \leq t_2$, all c :

$$P(C \geq c \mid T = t_1) \leq P(C \geq c \mid T = t_2) \tag{1}$$

In other words, there are more chances to see a high pollution when T is large than when T is small.

We will prove :

Proposition. – *This definition implies that :*

For all $t_1 \leq t_2$, *all* c *:*

$$P\left(C \geq c \,|\, T \geq t_1\right) \leq P\left(C \geq c \,|\, T \geq t_2\right) \tag{2}$$

Proof

We have:

In order to show (2), it is enough to show that :

$$P\left(C \geq c \,|\, T = t_1 \; or \; t_1 + 1\right) \leq P\left(C \geq c \,|\, T = t_2 \; or \; t_2 + 1\right) \tag{3}$$

Indeed, just reiterating this procedure yields (2).

We set:

$$a_1^0 = \sum_{j \geq j_0} n_{i_1, j} \;,\; a_1' = \sum_{j < j_0} n_{i_1, j}$$

$$b_1^0 = \sum_{j \geq j_0} n_{i_1+1, j} \;,\; b_1' = \sum_{j < j_0} n_{i_1+1, j}$$

and similarly:

$$a_2^0 = \sum_{j \geq j_0} n_{i_2, j} \;,\; a_2' = \sum_{j < j_0} n_{i_2, j}$$

$$b_2^0 = \sum_{j \geq j_0} n_{i_2+1, j} \;,\; b_2' = \sum_{j < j_0} n_{i_2+1, j}$$

We have:

$$P\left(C \geq c \,|\, T = t_1 \; or \; t_1 + 1\right) = \frac{a_1^0 + b_1^0}{a_1^0 + b_1^0 + a_1' + b_1'}$$

and similarly:

$$P\left(C \geq c \,|\, T = t_2 \; or \; t_2 + 1\right) = \frac{a_2^0 + b_2^0}{a_2^0 + b_2^0 + a_2' + b_2'}$$

Condition (1), applied to $\left(t_1, t_2\right)$ and to $\left(t_1 + 1, t_2 + 1\right)$, implies:

$$\frac{a_1^0}{a_1^0 + a_1'} \le \frac{a_2^0}{a_2^0 + a_2'}$$

which is equivalent to:

$$a_1^0\left(a_2^0 + a_2'\right) \le a_2^0\left(a_1^0 + a_1'\right)$$

and:

$$\frac{a_1^0}{a_2^0} \le \frac{a_1'}{a_2'} \tag{4}$$

Similarly, we have:

$$\frac{b_1^0}{b_1^0 + b_1'} \le \frac{b_2^0}{b_2^0 + b_2'}$$

which implies:

$$\frac{b_1^0}{b_2^0} \le \frac{b_1'}{b_2'} \tag{5}$$

The condition

$$\frac{a_1^0 + b_1^0}{a_1^0 + b_1^0 + a_1' + b_1'} \le \frac{a_2^0 + b_2^0}{a_2^0 + b_2^0 + a_2' + b_2'}$$

is equivalent to:

$$\frac{a_1^0 + b_1^0}{a_1' + b_1'} \le \frac{a_2^0 + b_2^0}{a_2' + b_2'} \tag{6}$$

In order to prove (6), we may assume that $a_1^0 = a_2^0 = b_1^0 = b_2^0 = 1$, and therefore (4) and (5) reduce to $a_2' \le a_1'$ and $b_2' \le b_1'$. The condition (6) is now $a_2' + b_2' \le a_1' + b_1'$, which follows from (4) and (5). Our Proposition is proved.

The reverse implication from (2) to (1) is not true in general, as the following example shows.

3	1	0	10
2	1	1	1
1	10	1	0
C/T	1	2	3

We have

P(C>=2|T>=1)= 0,56
P(C>=2|T>=2)= 0,92

P(C>=3|T>=1)= 0,44
P(C>=3|T>=2)= 0,77

so (2) is satisfied, but:

P(C>=3|T=1)= 0,08
P(C>=3|T=2)= 0