On a strange property of the Order 1 Variogramm

Alexey Ya. Chervonenkis

July 2001

1 The property

Let us consider two independent random variables x and y with corresponding, cumulative distribution functions $F_1(c) = Pr\{x < c\}$ and $F_2(c) = Pr\{y < c\}$.

We shall compare two values $E_x E_y (x-y)^2$ and $E_x E_y |x-y|$.

a) We denote $Ex = M_x$, $Ey = M_y$, $x* = x - M_x$, $y* = y - M_y$. Then $Ex* = 0$, $Ey* = 0$,

$$Ex^{*2} = \text{Var } x, Ey^{*2} = \text{Var } y.$$

Thus

$$\begin{aligned} E_x E_y (x-y)^2 = E_x E_y ((x^* - y^*) + (M_x - M_y))^2 = \\ = E_x E_y (x^{*2} - 2x^* y^* + y^{*2}) + (M_x - M_y)^2 = \\ = \text{Var } x + \text{Var } y + (M_x - M_y)^2. \end{aligned}$$

So we see that the result is the sum of the two variances and square of expectation difference. For equal variances (or in cases where the variance is expected to be equal) the result depends only on the difference between the two expectations.

b) Now we change square of difference to module of difference and have

$$E_x E_y |x-y| = E_x E_y \int (\Theta(x-c) - \Theta(y-c))^2 dc, \text{ where}$$

$\Theta(z) = 1$, if $z < 0$ and 0 otherwise, because

$$|x-y| = \int (\Theta(x-c) - \Theta(y-c))^2 dc.$$

During further calculations, we assume that the variables are limited, i.e. that their distribution is concentrated on some interval $[a,b]$. In this case, all of the integrals used below make sense. But the final result is true also for unlimited variables, because the integrals there converge if and only if expectations exist.

Then

$$
\begin{aligned}
&E_xE_y|x-y| = E_xE_y\int(\Theta^2(x-c)+\Theta^2(y-c)-2\Theta(x-c)\Theta(y-c))dc = \\
&= E_xE_y\int(\Theta(x-c)+\Theta(y-c)-2\Theta(x-c)\Theta(y-c))dc = \\
&= E_x\int\Theta(x-c)dc + E_y\int\Theta(y-c)dc - 2E_xE_y\int(\Theta(x-c)\Theta(y-c))dc = \\
&= \int E_x\Theta(x-c)dc + \int E_y\Theta(y-c)dc - 2\int(E_x\Theta(x-c)E_y\Theta(y-c))dc,
\end{aligned}
$$

as far as $\Theta^2(z) = \Theta(z)$ for all z, and x and y are independent.

But $E_x\Theta(x-c) = F_1(c) and E_y\Theta(y-c) = F_2(c)$.

Thus

$$
\begin{aligned}
&E_xE_y|x-y| = \int F_1(c)dc + \int F_2(c)dc - \int 2F_1(c)F_2(c)dc = \\
&= \int(F_1(c)-F_1^2(c))dc + \int(F_2(c)-F_2^2(c))dc + \int(F_1^2(c)-2F_1(c)F_2(c)+F_2^2(c))dc = \\
&= \int F_1(c)(1-F_1(c))dc + \int F_2(c)(1-F_2(c))dc + \int(F_1(c)-F_2(c))^2dc.
\end{aligned}
$$

All three integrals in the last expression converge over R if and only if expectations E_x and E_y exist.

So the final result is

$$E_xE_y|x-y| = \int F_1(c)(1-F_1(c))dc + \int F_2(c)(1-F_2(c))dc + \int(F_1(c)-F_2(c))^2dc.$$

The first two integrals are called the dispersion of the x and y distributions. They correspond to mean square deviations (although they exist even when variance is infinite). The third integral corresponds to the square of the distance between the distribution functions in functional space. So in this case the criteria "feels" not only difference in expectations, but also the difference of distributions. *And it is strange!*

1. Variogramms

We call a random field a real valued function $f(x,\omega)$ of two variables: $x \in R^n$ and $\omega \in \Omega$, where Ω is a random element of a random space with given proability measure P. Usually we shall write just $F(x)$, taking in mind that this function is random.

A common variogram $R(x,y)$ is defined as $R(x,y) = E(f(x)-f(y))^2$. A first order variogramm (a variogramm of order 1) is defined as $R^1(x,y) = E|f(x)-f(y)|$. A random feild is stationary in the close sense if the probability measure does not depend on a shift in R^n. In the stationary case variogramms depend only on the shift vector $v = y - x$:

$$
\begin{aligned}
&R(x,y) = R(y-x) = R(v) \text{ and} \\
&R^1(x,y) = R^1(y-x) = R^1(v).
\end{aligned}
$$

It is clear that variogramms are symetrical: $R(x,y) = R(y,x)$ and $R^1(x,y) = R^1(y,x)$.

We consider further the following construction of a random field. We assume first that there is given not a common real value random field, but a random field, in which values are distribution functions $F(x,c)$ $(0 \leq F \leq 1; F(c_1) \geq F(c_2)$ if $c_1 \geq c_2)$. The closeness between these values can be defined as the generalised variogramm

$$R^*(x,y) = E^* \int (F(c,x) - F(c,y)^2 dc,$$

where E^* denotes expectation in the sense of a generalised random field.

As before we consider the generalised random field to be stationary if its probanilistic measure is invariant to vector x shift.

A corresponding common random field is constructed on the basis of the generalised random field in the folowing way: given a realisation $F(x,c)$ of the generalised field at every point x a real value $f(x) = c$ is chosen, conditionally independent according to the distribution $F(x,c)$. If all distribution functions $F(x,c)$ have the density of the form of $d(c - f(x))$, where $f(x)$ is a realisation of the random field, we have just another definition of common random field. But our construction has advantiges in many applications.

Now we shall compare the generalised variogramm $R^*(x,y)$ with the variogramms $R(x,y)$ and $R(x,y)^1$.

According to the previous section

$$R(x,y) = E^*E(f(x) - f(y))^2 = E^*\text{Var }(x) + E^*\text{Var }(y) + E^*(M(x) - M(y))^2,$$

where E is expectation due to distribution $F, M(x)$, $\text{Var}(x), M(y)$, $\text{Var}(y)$ are its expectation and variance in the points x and y. So the usual variogramm "feels" only the changing of expectation (in the stationary case $E^*\text{Var }(x)$ is constant).

On the contrary for varogramm of order 1 we have

$$\begin{aligned}&R^1(x,y) = E^*E|f(x) - f(y)| = \\ &E^* \textstyle\int F(x,c)(1 - F(x,c))dc + E^* \int F(y,c)(1 - F(y,c))^2 dc + \\ &+E^* \textstyle\int (F(x,c) - F(y,c))^2 dc = E^*D(x) + E^*D(y) + E^* \int (F(x,c) - F(y,c))^2 dc.\end{aligned}$$

Average dispersions $E^*D(x)$ and $E^*D(y)$ in the stationary case are constant and really the first order variogramm depends on the average difference between distribution functions $F(x,c)$ and $F(y,c)$, but not only on the difference between their expectations. In particular, if expectation $M(x)$ is constant, though distribution changes, the common variogram will be constant (exept the case $x = y$) while first the order variogramm shows distribution changing. *That is strange.*