A COPULA BASED DIFFERENTIAL MEASURE OF LOCAL CORRELATION

Daniel Bruynooghe and Henry P. Wynn
Department of Statistics, London School of Economics, U.K.
E-mail: d.hawellek@lse.ac.uk

Abstract

A copula based measure of local correlation is developed for two random variables X and Y. The measure is originally motivated through the limiting process of a sequence of correlations in shrinking local neighbourhoods around (x, y). It is shown that this method is better applied in 'copula space' to the transformed variables $F_X(x), F_Y(y)$ in a sense of capturing the independence case properly. Upon transforming back via the inverse marginal CDFs, we arrive at a novel measure of local correlation. We illustrate its geometry for the bivariate Gaussian case. Finally, a non-parametric estimator is presented and its asymptotic distribution identified.

Motivation

Let $f(x, y)$ be the joint probability density function of X and Y.

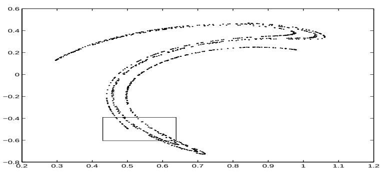

FIG. 1: Random sample from the attractor of the IKEDA map. The data clearly displays a negative correlation over the conditioning window A_ϵ.

Let $A_\epsilon := \left[x_0 \pm \frac{\epsilon_1}{2}\right] \times \left[y_0 \pm \frac{\epsilon_2}{2}\right]$ be the rectangular neighbourhood of (x_0, y_0). It can be shown that the first-order approximated correlation over A_ϵ equals:

$$-12^{-1}\epsilon_1\epsilon_2\Delta_{x_0}\Delta_{y_0}f^{-2}(x_0, y_0) + O(||\epsilon||^2),$$

where $\begin{pmatrix}\Delta_{x_0}\\ \Delta_{y_0}\end{pmatrix} := \nabla f(x, y)\Big|_{\binom{x}{y}=\binom{x_0}{y_0}}$.

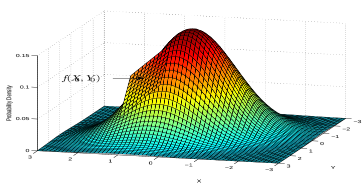

FIG. 2: First order approximation of a bivariate Gaussian density through a plane in the neighbourhood of $f(x_0, y_0)$.

This motivates the definition of local correlation as

$$Z_0 = -\frac{\Delta_{x_0}\Delta_{y_0}}{f^2(x, y)} = -\frac{\partial \ln f(x, y)}{\partial x}\frac{\partial \ln f(x, y)}{\partial y}. \quad (1)$$

The difficulty with this version is that independence of X and Y not necessarily implies $Z_0 = 0$.

The copula-based version

Let $U = F_X(X)$ and $V = F_Y(Y)$. The copula density of (U, V) is

$$f_{U,V}(u, v) = (f_X(x)f_Y(y))^{-1}f(x, y).$$

Computing the local correlation as defined in (1) for $f_{U,V}(u, v)$ and transforming back yields the quantity central to this poster:

$$Z_1 := -\frac{\partial}{\partial x}(\ln f(y|x))\frac{\partial}{\partial y}(\ln f(x|y))$$

It is immediately clear that independence of X and Y forces $Z_1 = 0$ for all (x, y).

Gaussian geometry

Assume (X, Y) has the bivariate Gaussian density with zero mean, variances σ_1^2, σ_2^2 and correlation coefficient ρ. Then the local correlation is given by

$$Z_1(x, y) = \frac{(\rho x\sigma_2 - y\sigma_1)(x\sigma_2 - \rho y\sigma_1)\rho^2}{\sigma_1^3\sigma_2^3(1-\rho^2)^2}.$$

The geometry of Z_1 is intuitively satisfactory as explained through Figure 3.

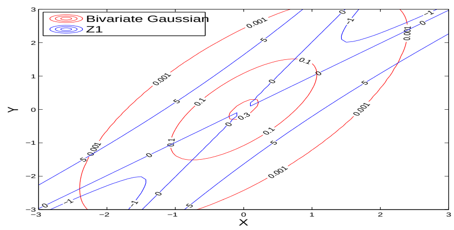

FIG. 3: $Z_1 = 0$ on the two straight lines $\rho\sigma_2 x' - \sigma_1 y' = 0$, $\sigma_2 x' - \rho\sigma_1 y' = 0$, which cross each other at the origin. These lines cross the elliptical contours of the density where the contours are vertical or horizontal. Between these lines Z_1 is positive or negative according to whether the contours of the density $f(x, y)$ have positive or negative slope.

A non-paramatric estimator

By definition, Z_1 is composed of the conditional densities $f(y|x), f(x|y)$ and their derivatives with respect to the conditioning variable. Fan et al. [1996] show ingeniously how to estimate a conditional density $f(y|x)$ and its derivatives through a weighted regression of $K(Y - y)$ on polynomials in $X - x$, where K is a kernel function. Consider the Taylor expansion:
$E\{K_{h_Y}(Y - y)|X = x\} =$

$$f_{Y|X}(y|x) + \frac{1}{2}\mu_2\frac{\partial^2}{\partial y^2}f_{Y|X}(y|x) + o(h_Y^2).$$

where $\mu_2 := \int z^2 K(z)dz$.

From this it follows that $E\{K_{h_Y}(Y - y)|X = x\} \to f(y|x)$ as $h_Y \to 0$, which makes it suitable as a regression target. A Taylor-expansion of $f_{Y|X}(y|X)$ about x yields:

$$f_{Y|X}(y|X) \simeq \sum_{j=0}^{p}\frac{f_{Y|X}^{(j)}(y|x)}{j!}(X - x)^j \quad (2)$$

which shows that the target is linear in the polynomials of $(X - x)$ which permits estimation through weighted least squares (WLS). With $\hat{\beta}$ as the WLS-estimator, β_0 estimates $f(y|x)$ and β_j the j-th derivative.
Since Z_1 is composed of conditional densities, we can adopt this approach and simultaneously estimate $f(y|x)$ and $f(x|y)$ and their derivatives. Let $\hat{Z}_1$ be the so obtained estimator:

$$\hat{Z}_1 := \frac{\hat{\beta}_1}{\hat{\beta}_0}\frac{\hat{\gamma}_1}{\hat{\gamma}_0}, \quad (3)$$

where γ is the WLS-estimator for the reverse regression of X on Y.

Asymptotic distribution

Asymptotic normality for $\hat{Z}_1$ can be shown under the same regularity assumptions imposed by Fan et al. [1996] for the independent case. The proof entails showing joint asymptotic normality for $(\hat{\beta}, \hat{\gamma})$ and an application of the 'delta-rule' to establish weak convergence of a continuous transformation of a weakly convergent random variable. Important intermediate observations include that the estimators of the derivatives converge slower then estimators of the densities themselves and hence dominate the asymptotic behaviour and that the derivative estimators can be shown to be asymptotically uncorrelated if the kernel functions are symmetric.

Conclusion and further work

The local correlation measure Z_1 has been derived as the limiting correlation in a shrinking neighbourhood for the copula density of two random variables. Its appealing geometry was shown for the bivariate Gaussian case. The properties of Z_1 for Non-Gaussian densities are currently explored. We conjecture that they are similar within the whole class of ellyptical densities.
The non-parametric estimator $\hat{Z}_1$ has been shown to be asymptotically normally distributed. Rate of convergence and distribution are yet to be verified numerically.

References

J. Fan et al. Estimation of conditional densities and sensitivity measures in nonlinear dynamical systems. *Biometrika*, 83(1):189, 1996.