

# A new look at measuring dependence

Wei Biao Wu and Jan Mielniczuk

**Abstract** This paper revisits the concept of dependence. We view statistical dependence as the state of variables being influenced by others. Our viewpoint accords well with the daily understanding of the notion of dependence, while classical dependence measures such as Pearson's correlation coefficient, Kendall's  $\tau$  and Spearman's  $\rho$  have different meanings. With this understanding of dependence, we introduce new dependence measures which are easy to work with and they are useful for developing an asymptotic theory for complicated stochastic systems. We also explore relations of the introduced dependence concept with nonlinear system theory, experimental design, information theory and risk management.

## 1 Introduction

According to the Merriam-Webster dictionary, *dependence* means that

"the quality or state of being dependent; especially: the quality or state of being influenced or determined by or subject to another."

This interpretation appears to accord well with our daily understanding of the notion of dependence. In probability theory and statistics, however, the meaning of dependence is understood somewhat differently. As pointed out in Drouet Mari and Kotz (2001, p. 31):

---

Wei Biao Wu  
University of Chicago,  
Department of Statistics, The University of Chicago,  
Chicago, IL 60637, USA, e-mail: wbwu@galton.uchicago.edu

Jan Mielniczuk  
Polish Academy of Sciences  
Warsaw University of Technology, Warsaw, Poland  
e-mail: miel@ipipan.waw.pl

"A term 'statistical dependence' rather than dependence should perhaps be used to emphasize that we are dealing with new concepts specific for probability theory which may not coincide with the daily meaning of this concept."

Various concepts and measures of statistical dependence have been introduced, including Pearson's correlation coefficient, Kendall's  $\tau$  and Spearman's  $\rho$  among others. These dependence measures are not in conformity with the above understanding of dependence. Others, which can be broadly described as regression dependence concepts, including Pearson's correlation ratio  $\eta^2$ , Goodman-Kruskal's  $\tau$  and the notion of positive and negative regression dependence (Lehmann, 1966), agree with it. Literary use of correlation which covers any notion of dependence contributes to creating frequent misconceptions about formalization of this concept.

A historical account of the dependence concepts mentioned above can be found in Drouet Mari and Kotz (2001). Joe (1997) discussed using copulas in this context and also other dependence concepts and measures as well as applications in reliability theory. See also Lai and Xie (2006), Szekli (1995) and the monograph edited by Block, Sampson and Savits (1990). For other important contributions to the concept of dependence see Tjøstheim (1996) and Skaug and Tjøstheim (1993).

The goal of the paper is to provide another look at regression approach to "statistical dependence" which is in line with our daily understanding of dependence. In particular, we shall view statistical dependence as the quality or state of variables being influenced by others, in accordance the interpretation given in the Meriam-Webster dictionary. With this notion of dependence, we will introduce its measures which quantify how the output variable varies with respect to changes in the input variables. The novelty here is that the presence of all other variables influencing output variable, summarily described as noise, is taken into account. In particular, the proposed measures depend on strength of interaction between the input variables and the noise in contrast e.g. to Pearson correlation ratio  $\eta^2$ . This approach has tight connections with various areas including causal inference, nonlinear system theory, prediction, experimental design, information theory and risk management. More importantly, the proposed dependence measures provide a useful tool for statistical analysis of complicated stochastic systems. They are a convenient vehicle for developing an asymptotic theory which is useful for the statistical inference, control, prediction and other objectives.

The rest of the paper is structured as follows. In Section 2 we discuss in detail our understanding of dependence and introduce various bivariate dependence measures, as well as their connections with information theory. In Section 3 we shall mention some connections with reliability theory regarding concepts of positive dependence. Multivariate dependence is treated in Section 4. Section 5 presents moment inequalities and limit theorems for sums of random variables with complicated dependence structures. We consider here a special type of causal processes being measurable functions of Bernoulli shifts pertaining to independent and identically distributed (iid) random variables. Some useful tools which are of independent interest are also provided.

## 2 Bivariate dependence

For a pair of jointly distributed random variables  $(X, Y)$ , let  $F_{XY}(x, y) = \mathbb{P}(X \leq x, Y \leq y)$ ,  $x, y \in \mathbb{R}$ , be the joint distribution function and  $F_{Y|X}(y|x) = \mathbb{P}(Y \leq y|X = x)$  the conditional distribution function of  $Y$  given  $X = x$ . For  $u \in (0, 1)$ , define the conditional quantile function

$$G(x, u) = \inf\{y \in \mathbb{R} : F_{Y|X}(y|x) \geq u\}. \tag{1}$$

By definition, for fixed  $x$ ,  $G(x, u)$  is a nondecreasing function in  $u$ . The conditional quantile function  $G$  plays an important role in our theory of dependence. Let  $U$  be a uniform(0,1) distributed random variable and assume that  $U$  and  $X$  are independent. Then the distributional equality

$$(X, Y) =_{\mathcal{D}} (X, G(X, U)) \tag{2}$$

holds. Based on (2), we can view  $Y$  as the outcome of the bivariate function  $G(\cdot, \cdot)$ :

$$Y = G(X, U) \tag{3}$$

In other words,  $Y$  is viewed as the output from a random physical system with  $X$  and  $U$  being the input and the noise or error, respectively. If we call  $X$  an independent variable and  $Y$  a dependent variable, then (3) becomes a regression problem if the goal is to find or estimate  $G$ . Observe that the right hand side of (1) can be written as a transformation of two *independent* uniform(0,1) distributed random variables  $U$  and  $V$ , namely  $(F_X^{-1}(V), G(F_X^{-1}(V), U))$ . This is different from copula construction when  $(X, Y)$  is transformed to a vector  $(V_1, U_1) = (F_X(X), F_Y(Y))$  and dependence of  $X$  and  $Y$  is reflected by dependence of  $U_1$  and  $V_1$ . These two frameworks have different ranges of applicability. Ours can lead to dependence measures that are useful for developing moment inequalities and limit theorems for stochastic processes (cf. Section 5).

*Remark 0.1.* If  $F_{Y|X}(y|x)$  is strictly increasing and continuous in  $y$ , then it is obvious that  $U := F_{Y|X}(Y|X)$  is uniform(0,1) distributed and it is independent of  $X$ . In this case we have a stronger version of (2): the identity  $(X, Y) = (X, G(X, U))$  holds almost surely, while (2) holds only in the distributional sense. For a general construction yielding almost sure equality see Rüschendorf (1981).  $\diamond$

With (3), we generically interpret dependence as how the output  $Y$  depends on the input  $X$  in *the presence of the noise*  $U$ . This interpretation seems to provide some clarifications of the troublesome confusion on the understanding of dependence mentioned in Mosteller and Tukey (1977, p. 262):

We must be clearer about the abused word "dependence" and its relatives. When we say "y depends on x", sometimes we intend *exclusive dependence*, meaning that, if  $x$  is given, then the value of  $y$  follows, usually because the existence of a law is implied. In mathematics if  $y$  is the area of a circle and  $r$  the radius, then  $y = \pi r^2$  illustrates this regular type of *exclusive dependence*.

At other times "y depends on x" means *failure of independence*, usually in the sense of "other things being equal" as in "the temperature of the hot water depends on how far the faucet is from the heater." Clearly, it may depend on other things, such as the setting of the heater and the building's temperature, to say nothing of whether the long pipe between the faucet and the heater is on a cold outside wall or a warm inside one. These are two quite different ideas of dependence, and the use of one word for both has often led to troublesome, if not dangerous, confusion.

Then there are the mathematical usages "dependent variable" and "independent variable". These have been extremely effective in producing confusion when dealing with data.

If  $G(x, u)$  does not depend on  $u$ , then we can write  $Y = g(X)$  and hence  $Y$  functionally depends on  $X$ . Namely if  $X = x$  is given, then the value of  $Y$  is explicitly known. This is the exclusive dependence discussed above. On the other hand, if  $G(x, u)$  does not depend on  $x$ , then  $Y$  is functionally as well as statistically independent of  $X$ . For situations between this two extremes, we interpret the meaning of "Y depends on X" as "the state of Y being influenced by X", as indicated by the explanation of dependence in the Merriam-Webster dictionary. For example, in the second example, if we let  $Y$  be the temperature of the hot water and  $X$  the distance from the heater to the faucet. Then  $Y$  is certainly influenced by other factors which are summarily described by  $U$  in the definition above. With this understanding, one can have a unified look at the two quite different ideas of dependence and thus help clarify the "troublesome, if not dangerous, confusion".

Furthermore, our approach leads to dependence measures that are useful in developing an asymptotic theory for complicated stochastic systems. Certainly there are many ways to implement and quantify the degree of dependence of  $Y$  on  $X$ . Here we consider two types: global dependence measures and local dependence measures, which are dealt with in Sections 2.1 and 2.2, respectively.

## 2.1 Global dependence measures

From now on assume that a copy of the bivariate random vector under consideration is given for which (3) holds. To define our global dependence measures, we need to apply the idea of coupling. This idea, which is frequently used in contemporary statistics, can be used in a natural way to measure the strength of dependence. Let  $X'$  be an iid copy of  $X$  and let  $X$  and  $X'$  be independent of  $U$ . Measures defined below are asymmetric and are meant to gauge how strongly  $Y$  depends on  $X$ .

**Definition 0.1.** Let  $Y = G(X, U) \in \mathcal{L}^p$ ,  $p > 0$ ; let  $Y' = G(X', U)$ . Define  $\delta_p(X, Y) = \|Y - Y'\|_p$ , and, for  $p \geq 1$ ,  $\tau_p(X, Y) = \|Y - \mathbb{E}(Y|U)\|_p$ .

In Definition 0.1,  $\delta_p = \delta_p(X, Y)$  is defined under the principle of *ceteris paribus* (with other things being the same), a desirable property in the theory of design of experiments. Specifically, the random noises  $U$  in  $Y = G(X, U)$  and  $Y' = G(X', U)$  are the same, and the only difference is  $X$  and  $X'$ . If we interpret (3) as a physical system with the causality relationship of  $X$  probabilistically causing  $Y$  (Mellor, 1998), then

we view  $X$  and  $Y$  as the input and the output, respectively, and thus  $\delta_p$  measures the overall dependence of  $Y$  on  $X$ . In this sense we call  $\delta_p$  *physical* or *functional dependence measure*. In the early practice of statistics, the absence of causal relation has been linked with probability through the concept of independence and statistical inference has been developed as a method for inquiry into features of causes (see Glymour (2006) and Stigler (1986)). Here  $\delta_p$  is meant to be a measure of causal relationship when a cause ( $X$ ) is understood as any quantity whose prior variation would result in the subsequent variation of another quantity ( $Y$ ).

In certain situations, especially in observational studies, it is not always possible to change  $X$  and the causality relationship may not be observable. In such cases  $\delta_p$  is only a mathematical object which may not be practically meaningful. Nonetheless, the measure  $\delta_p$  is useful in developing an asymptotic theory. Different concept of causality has been introduced e.g. by Granger (1969) who defines that  $X$  is causing  $Y$  when we are able to predict  $Y$  with a greater precision using the value of  $X$  than using all other information our disposal other than  $X$ . Other approaches of describing complicated dependence patterns include graphical Markov models (see e.g. Cowell et al., 1999) and Rubin's framework (Rubin, 1990).

Observe that both introduced measures depend in a complicated way on  $G$ , the conditional quantile function of  $Y$  given  $X$ . Under appropriate regularity conditions, based on a random sample  $(X_i, Y_i)_{i=1}^n$ ,  $G$  can be estimated non-parametrically and ensuing estimates  $\widehat{\delta}_p(X, Y)$  and  $\widehat{\tau}_p(X, Y)$  can then be computed or simulated. If  $Y$  is known to depend on  $X$  in a parametric way, then it suffices to estimate the unknown parameter and then  $\widehat{\delta}_p(X, Y)$  and  $\widehat{\tau}_p(X, Y)$  can be computed through simulation. In general  $\delta_p(X, Y)$  and  $\tau_p(X, Y)$  are not directly estimable. However, we will argue that they can be very useful in theoretical analysis of dependence.

The two dependence measures  $\delta_p$  and  $\tau_p$  are closely related; see Proposition 0.1 below. In the definition of  $\tau_p$ , the conditional expectation  $\mathbb{E}(Y|U)$  smoothes out the impact of  $X$  on  $Y$ . Thus  $\tau_p$  is a distance between  $Y$  and its analogue for which the impact of  $X$  has been averaged. Both measures quantify the degree of global dependence of  $Y = G(X, U)$ , as a random function of  $X$ , on  $X$ . With this interpretation, our dependence measures are naturally asymmetric and the degree of dependence of  $Y$  on  $X$  and that of  $X$  on  $Y$  are different. There certainly exist other ways to measure the discrepancy between  $Y$  and  $Y'$  (or  $\mathbb{E}(Y|U)$ ) by using e.g. different metrics.

**Proposition 0.1.** *We have for  $p \geq 1$  that  $\tau_p(X, Y) \leq \delta_p(X, Y) \leq 2\tau_p(X, Y)$ .*

*Proof.* Write  $\tau_p = \tau_p(X, Y)$  and  $\delta_p = \delta_p(X, Y)$ . Since  $\mathbb{E}(Y|U) = \mathbb{E}(Y'|X, U)$ ,  $Y - \mathbb{E}(Y|U) = \mathbb{E}(Y - Y'|X, U)$ . By Jensen's inequality,  $\tau_p \leq \delta_p$ . Since  $X, X', U$  are independent,  $\mathbb{E}(G(X, U)|U) = \mathbb{E}(G(X', U)|U)$ . Then

$$\delta_p \leq \|G(X, U) - \mathbb{E}(G(X, U)|U)\|_p + \|\mathbb{E}(G(X', U)|U) - G(X', U)\|_p = 2\tau_p.$$

◇

**Proposition 0.2.**  *$\delta_p(X, Y)$  and  $\tau_p(X, Y)$  are strong measures of dependence in the sense that  $\delta_p = 0$  ( $\tau_p = 0$ ) if and only if  $X$  and  $Y$  are independent.*

*Proof.* By Proposition 0.1, it is enough to check the property for  $\tau_p$ . Clearly,  $\tau_p = 0$  if and only if  $G(X, U) = \mathbb{E}(G(X, U)|U)$   $P_X \times P_U$  -a.e. which implies that for any  $0 < u < 1$   $F_{Y|X}^{-1}(u|x)$  does not depend on  $x$ . This is equivalent to independence of  $X$  and  $Y$ .  $\diamond$

Thus both measures satisfy one of Rényi postulates on measures of dependence. Observe, however, that they are both regression type measures and the symmetry postulate fails for them.

A particular important case is  $p = 2$ . In this case  $\delta_2 = \sqrt{2}\tau_2$  as  $G(X, U) - \mathbb{E}(G(X, U)|U)$  and  $G(X', U) - \mathbb{E}(G(X', U)|U)$  are conditionally independent given  $U$  with a conditional mean equal to 0.

Let us call  $\mathbb{E}(Y|X)$  a main effect of  $X$  on  $Y$ . We have the following decomposition due to orthogonality

$$\begin{aligned} \|Y - \mathbb{E}Y\|^2 &= \|Y - \mathbb{E}(Y|X) - \mathbb{E}(Y|U) + \mathbb{E}Y\|^2 \\ &\quad + \|\mathbb{E}(Y|X) - \mathbb{E}Y\|^2 + \|\mathbb{E}(Y|U) - \mathbb{E}Y\|^2. \end{aligned} \tag{4}$$

Thus the variability of  $Y$  is decomposed into three terms: variability of two main effects of  $U$  and  $X$  on  $Y$  and the variability of their interaction. Moreover, we have that the sum of two first terms in the above formula equals  $\|Y - \mathbb{E}(Y|U)\|^2 = \tau_2^2$ . Define Pearson correlation ratio (nonparametric R-squared) of arbitrary square integrable random variables  $Z$  and  $Y$  as  $\eta_{Z,Y}^2 = \text{var}(\mathbb{E}(Y|Z))/\text{var}(Y)$  (cf e.g. Doksum and Samarov (1995)). Note that the numerator of  $\eta_{Z,Y}^2$  equals  $\text{var}(Y) - \mathbb{E}\text{var}(Y|Z)$  thus it measures relative decrease of variability of  $Y$  given  $Z$ ; other measures of dependence such as Goodman-Kruskal's  $\tau$  for categorical data are constructed in the same vein. The following proposition reveals the relationship between  $\tau_2$ ,  $\text{var}(Y)$  and  $\eta_{U,Y}$ .

**Proposition 0.3.** *We have*

- (i)  $\text{var}(Y) = \tau_2^2 + \|\mathbb{E}(Y|U) - \mathbb{E}Y\|^2 = \tau_2^2 + \eta_{U,Y}^2 \times \text{var}(Y)$ ;
- (ii)  $\tau_2^2 \geq \eta_{X,Y}^2 \times \text{var}(Y)$ .

Inequality (ii) follows by noting that  $\mathbb{E}(Y - \mathbb{E}(Y|U))|X) = \mathbb{E}(Y|X) - \mathbb{E}Y$ .

Observe that it follows from (4) that when we consider  $\tau_2$  as the measure of dependence of  $Y$  on  $X$ , we subtract from the overall variability of  $Y$  only the variability of the main effect of  $U$ , leaving the interaction term. This corresponds to the idea of quantifying how the output  $Y$  depends on the input  $X$  in the presence of the noise  $U$ .

Let  $\rho(X, Y) = \text{cov}(X, Y)/\sqrt{\text{var}(X)\text{var}(Y)}$  be correlation coefficient of  $X$  and  $Y$  and observe that comparing distances of  $Y - \mathbb{E}Y$  from its projections on the subspaces of measurable and linear functions of  $U$  we have that  $\rho^2(U, Y) \leq \eta_{U,Y}^2$  (Cramér, 1946). Moreover, using independence of  $X$  and  $U$  we have

$$\text{cov}^2(X, Y) = \mathbb{E}^2(X - \mathbb{E}X)(Y - \mathbb{E}(Y|U)) \leq \text{var}(X)\tau_2^2.$$

Thus we obtain

**Proposition 0.4.** *The following inequality holds:*

$$\rho^2(X, Y) \leq \tau_2^2 / \text{var}(Y) \leq 1 - \rho^2(U, Y).$$

Let us compare now  $\tau_2^2$  with  $\eta_{X,Y}^2$ .

*Example 2.1.* Consider the heteroscedastic regression model

$$Y = f(X) + \sigma(X)\varepsilon,$$

where  $\varepsilon$  and  $X$  are independent,  $\mathbb{E}\varepsilon = 0$  and  $\sigma(X) > 0$   $P_X$ -a.e. Then we have

$$F_{Y|X}^{-1}(u|x) = f(X) + \sigma(X)F_\varepsilon^{-1}(u),$$

where  $F_\varepsilon$  is the distribution function of  $\varepsilon$  and  $F_\varepsilon^{-1}$  is the quantile function. Thus

$$\begin{aligned} \tau_2^2 &= \|Y - \mathbb{E}(Y|U)\|^2 = \|f(X) - \mathbb{E}f(X) + (\sigma(X) - \mathbb{E}\sigma(X))F_\varepsilon^{-1}(U)\|^2 \\ &= \text{var}(f(X)) + \text{var}(\sigma(X))\mathbb{E}\varepsilon^2. \end{aligned}$$

Note that  $\text{var}(Y) = \text{var}(f(X)) + \mathbb{E}(\sigma^2(X))\mathbb{E}\varepsilon^2$ . Thus

$$\tau_2^2 = \text{var}(Y) - (\mathbb{E}\sigma(X))^2\mathbb{E}\varepsilon^2.$$

Consider the following special cases

(i) (additive model) Let  $\sigma(X) \equiv 1$ . Then  $\tau_2 = \sigma_{f(X)}$  and  $\tau_2^2 = \text{var}(Y) - \text{var}(\varepsilon)$ . Note that this is also true if  $\varepsilon \equiv 0$ . In particular, for  $(X, Y) \sim N(m_X, m_Y, \sigma_X, \sigma_Y, \rho)$  we have that  $Y = m_X + \sigma_Y(\rho/\sigma_X(X - m_X) + (1 - \rho^2)^{1/2}\varepsilon)$  where  $\varepsilon \sim N(0, 1)$  is independent of  $X$  and thus  $\tau_2 = |\rho|\sigma_Y$ .

(ii) (multiplicative model) Let  $f(X) \equiv 0$ . Then  $\tau_2 = \sigma_{\sigma(X)}(\mathbb{E}\varepsilon^2)^{1/2}$ .

Observe that it follows from the above examples that for the additive model  $\eta_{X,Y}^2 = \tau_2^2 / \text{var}(Y)$  thus  $\eta_{X,Y}^2$  coincides with  $\tau_2^2$  up to a positive constant whereas for the multiplicative model  $Y = X\varepsilon$  with  $\mathbb{E}\varepsilon = 0$  we have that  $\eta_{X,Y}^2 = 0$  whereas  $\tau_2 > 0$ . Thus in the last case the interaction effect is present which is detected by  $\tau_2$  but not by  $\eta_{X,Y}^2$ .

*Example 2.2.* (One way ANOVA). Assume that  $Y = G(X, U)$  in (3) has the special form  $Y = \mu_X + \varepsilon$ , where  $X$  is uniformly distributed over  $\{1, 2, \dots, k\}$ , and  $\varepsilon = H(U)$  has finite second moment. Let  $\bar{\mu} = \sum_{i=1}^n \mu_i / k$ . Then  $\mathbb{E}(Y|U) = \varepsilon + \bar{\mu}$ , and  $\tau_2^2 = k^{-1} \sum_{i=1}^n (\mu_i - \bar{\mu})^2$  corresponds to the treatment effect.

*Example 2.3.* Let  $\tilde{G}$  be a conditional quantile function corresponding to a distribution function (copula) of a transformed vector  $(F_X(X), F_Y(Y))$ . Then it is easy to check that  $F_{F(Y)|F(X)}^{-1}(u|F(x)) = F_Y(F_{Y|X}^{-1}(u|x))$  and thus  $\tilde{G}(u, x) = F_Y(G(u, F_X^{-1}(x)))$ . Moreover, the definition of  $\delta_p$  for the pair  $(F_X(X), F_Y(Y))$  is  $\|\tilde{G}(W, U) - \tilde{G}(W', U)\|_p$ , where  $W, W', U$  are independent  $U[0, 1]$  random variables. Equivalently,  $\delta_p = \|F_Y(Y) - F_Y(Y')\|_p$ , where  $Y = G(X, U)$ ,  $Y' = G(X', U)$  and  $X'$  is an independent copy of  $X$ .

*Example 2.4.* Let  $X$  and  $Y$  be 0/1-valued and  $\mathbb{P}(X = i, Y = j) = p_{ij}$ ,  $i, j = 0, 1$ . Let  $p_0 = p_{00} + p_{01}$  and  $p_1 = p_{10} + p_{11}$ . Then  $p_0 + p_1 = 1$  and  $\text{cov}(X, Y) = p_{11} - (p_{10} +$

$p_{11})(p_{01} + p_{11}) = p_{11}p_{00} - p_{01}p_{10}$ . Let  $G(X, U) = X\mathbf{1}_{U \leq p_{11}/p_1} + (1 - X)\mathbf{1}_{U \leq p_{01}/p_0}$ . Then (2) holds. Since  $X$  and  $U$  are independent and  $\mathbb{E}X = p_1$ , we have

$$\begin{aligned} \tau_2^2(X, Y) &= \|Y - \mathbb{E}(Y|U)\|_2^2 = \text{var}(X)\mathbb{E}|\mathbf{1}_{U \leq p_{11}/p_1} - \mathbf{1}_{U \leq p_{01}/p_0}| \\ &= \text{var}(X)|p_{11}/p_1 - p_{01}/p_0| = |p_{11}p_{00} - p_{01}p_{10}| = |\text{cov}(X, Y)|. \end{aligned}$$

Observe that in this case we have symmetry:  $\tau_2(X, Y) = \tau_2(Y, X)$ .

## 2.2 Local dependence measures

In Definition 0.1 we introduced global dependence measures which quantify the difference between  $Y = G(X, U)$  and its coupled version  $Y' = G(X', U)$ . In certain applications it is desirable to know the local dependence of  $Y$  on  $X$  varies with changing  $x$ . A natural way is to consider a properly normalized distance between  $G(x, U)$  and  $G(x + \delta, U)$ , where  $\delta \in \mathbb{R}$  is small.

**Definition 0.2.** Assume that the limit  $G'(x, U) = \lim_{\delta \rightarrow 0} \delta^{-1}[G(x + \delta, U) - G(x, U)]$  exists almost surely. For  $p > 0$  define  $\delta_p(x) = \|G'(x, U)\|_p$  if the latter exists.

The local dependence measure  $\delta_p(x)$  quantifies the dependence of  $Y = G(X, U)$  on  $X$  at  $X = x$  in the presence of the noise  $U$ . If  $G(x, u)$  does not depend on  $u$  i.e.  $Y$  is a function of  $X$ , we write  $G(x, u) = G(x)$  and then  $G'(x, U)$  is just the derivative  $G'(x)$ . The following proposition shows that the global dependence measure  $\delta_p$  is bounded by a  $\delta_p(x)$ -weighted distance between  $X$  and  $X'$ .

**Proposition 0.5.** Let  $p \geq 1$ . Then  $\delta_p \leq \|\int_X^{X'} \delta_p(x) dx\|_p$ .

*Proof.* The inequality trivially holds if  $p = 1$ . For  $p > 1$  let  $q = p/(p - 1)$  and  $\lambda(z) = \delta_p^{-1/q}(z)$ . Then  $q^{-1} + p^{-1} = 1$  and by Hölder's inequality,

$$\begin{aligned} |G(x, U) - G(x', U)| &= \left| \int_x^{x'} G'(z, U) dz \right| \\ &\leq \left| \int_x^{x'} |G'(z, U)|^p \lambda^p(z) dz \right|^{1/p} \times \left| \int_x^{x'} \lambda(z)^{-q} dz \right|^{1/q}. \end{aligned}$$

Then  $\mathbb{E}[|G(x, U) - G(x', U)|^p] \leq |\int_x^{x'} \delta_p(z) dz|^p$  and the proposition follows.  $\diamond$

In Definition 0.2,  $\delta_p(x)$  involves the Euclidean distance between  $G(x, U)$  and  $G(x + \delta, U)$ . An alternative way is to use the information divergence. Let  $p(\cdot|x)$  be the density function of  $G(x, U)$ . There exist many ways to measure the distance between  $p(\cdot|x)$  and  $p(\cdot|x + \delta)$ .

**Definition 0.3.** (i)  $I_1$  local dependence measure. Assume that as  $t \rightarrow 0$ , one has the approximation



$$\int [p(y|x+t) - p(y|x)] \log \frac{p(y|x+t)}{p(y|x)} dy = t^2 I_1(x) + o(t^2), \tag{5}$$

where the left hand side of (5) is symmetrized Kullback-Leibler divergence between  $p(\cdot|x)$  and  $p(\cdot|x+t)$  and

$$I_1(x) = \int \frac{[\partial p(y|x)/\partial x]^2}{p(y|x)} dy \tag{6}$$

Note that  $I_1(x)$  may be viewed as Fisher’s information of  $p(\cdot|x)$  when  $x$  is treated as its parameter.

(ii)  $I_2$  local dependence measure. Assume that as  $t \rightarrow 0$ , one has the approximation

$$\int [p(y|x+t) - p(y|x)]^2 dy = t^2 I_2(x) + o(t^2), \tag{7}$$

where

$$I_2(x) = \int [\partial p(y|x)/\partial x]^2 dy. \tag{8}$$

Observe that equation (5) is satisfied under regularity conditions in view of equality  $\log(1 + \Delta) = \Delta + \mathcal{O}(\Delta^2)$  with  $\Delta = (p(y|x+t) - p(y|x))/p(y|x)$ .

In the context of nonlinear prediction theory, Fan and Yao (2003) introduced local measures  $I_1$  and  $I_2$  to quantify sensitivity to initial values. Wu (2008) applied  $I_2(x)$  to study weak convergence of empirical processes of dependent random variables. Blyth (1994) noticed that  $I_1(x)$  has a nice property:  $I_1(x) \geq [g'(x)]^2 / \text{var}(Y|X = x)$ , where  $g(x) = \mathbb{E}(Y|X = x) = \mathbb{E}G(x, U)$  and  $\text{var}(Y|X = x) = \mathbb{E}(Y^2|X = x) - g^2(x)$ . The last inequality is just the conditional version of the Cramér-Rao inequality.

Hellinger distance is another widely used metric to measure distances between distributions. Proposition 0.6 shows that the local Hellinger dependence measure is closely related to  $I_1(x)$  and they are equivalent in the local sense. It easily follows from the Lebesgue dominated convergence theorem.

**Proposition 0.6.** *Assume that  $\frac{\partial}{\partial x} \sqrt{p(y|x)}$  exists for  $y \in \mathbb{R}$  and for some  $g_x(\cdot) \in \mathcal{L}^2(\mathbb{R})$  it holds that  $|\sqrt{p(y|x+t)} - \sqrt{p(y|x)}| \leq g_x(y)|t|$  for all  $y \in \mathbb{R}$  and sufficiently small  $t$ . Then we have*

$$\lim_{t \rightarrow 0} \frac{1}{t^2} \int [\sqrt{p(y|x+t)} - \sqrt{p(y|x)}]^2 dy = \frac{I_1^2(x)}{4}. \tag{9}$$

A special class of local divergencies  $\mathcal{D}$  such that  $\mathcal{D}(p(y|x+t), p(y|x)) \sim t^2 I_1(x)$  when  $t \rightarrow 0$  is considered in Blyth (1994). It is proved that for such measures, called local Rao divergencies, we have

$$\lim_{t \rightarrow 0} \frac{\mathcal{D}(p(y|x+t), p(y|x))}{t^2} \geq \frac{g'^2(x)}{\text{var}(Y|X = x)}.$$

*Example 2.5.* Consider the special case  $G(x, u) = g(x) + H(u)$ , where  $H(U)$  has density function  $h$  and  $g$  is differentiable. Then  $p(y|x) = h(y - g(x))$  and  $\partial p(y|x)/\partial x = h'(y - g(x))g'(x)$ . Let  $I_h = \int [h'(y)]^2/h(y)dy$  be the Fisher information pertaining to  $h$ . Then by (6), the Kullback-Leibler local dependence measure  $I_1(x) = [g'(x)]^2 I_h$ .

*Example 2.6.* Consider again the heteroscedastic regression model. As  $G(x, u) = f(x) + \sigma(x)F_\epsilon^{-1}(u)$  we have

$$\begin{aligned} \delta_2^2(x) &= \int_0^1 (f'(x) + \sigma'(x)F_\epsilon^{-1}(u))^2 du \\ &= f'^2(x) + \sigma'^2(x) \int (F_\epsilon^{-1}(u))^2 du + 2f'(x)\sigma'(x) \int F_\epsilon^{-1}(u) du \\ &= f'^2(x) + \sigma'^2(x)\text{var}(\epsilon). \end{aligned}$$

Let  $g(x) = \mathbb{E}(Y|X = x)$  and observe that the derivative  $g'(x)$  indicates how the conditional expected value of  $Y$  given  $X = x$  changes with changing  $x$ . Bjerve and Doksum (1993) introduced correlation curve  $\rho(x)$ , still another measure of local dependence, which is based on  $g'(x)$  and takes the form

$$\rho(x) = \frac{\sigma_1 g'(x)}{(g'(x)^2 \sigma_1^2 + \sigma^2(x))^{1/2}},$$

where  $\sigma^2(x) = \text{var}(Y|X = x)$  and  $\sigma_1^2 = \text{var}(X)$ . This is a local analogue of equality holding for a correlation coefficient of bivariate normal vector  $\rho = \beta \sigma_1 / (\beta^2 \sigma_1^2 + \sigma^2)^{1/2}$  with  $\beta$  denoting slope of  $Y$  regressed on  $X$  and  $\sigma^2 = \text{var}(Y|X)$ . Correlation curve  $\rho(x)$  is easily estimated by means of local linear smoothers which yield natural estimator of  $g'(x)$ . Observe, however, that for multiplicative model  $\rho(x) \equiv 0$ . This is due to the fact that correlation curve measures strength of locally *linear* relationship. In order to remedy this drawback Doksum and Froda (2000) consider correlation ratio given  $X$  belongs to a small neighborhood of  $x$  calibrated to coincide with  $\rho^2$  in linear models. The following proposition relates  $\delta_p(x)$  with numerator of the ratio defining correlation curve.

**Proposition 0.7.** *Let  $G_\eta(x, u) = \sup_{|x-y| \leq \eta} |\partial G(y, u)/\partial y|$ . Assume that for some  $\eta > 0$ , we have  $G_\eta(x, \cdot) \in \mathcal{L}^p$  for some  $p \geq 1$ . Then  $|g'(x)| \leq \delta_p(x)$ .*

*Proof.* Define  $D_\delta(u) = [G(x + \delta, u) - G(x, u)]/\delta$ . Then

$$\begin{aligned} \left| \lim_{\delta \rightarrow 0} \frac{g(x + \delta) - g(x)}{\delta} \right| &= \left| \lim_{\delta \rightarrow 0} \int_0^1 D_\delta(u) du \right| \leq \\ \lim_{\delta \rightarrow 0} \left( \int_0^1 |D_\delta(u)|^p du \right)^{1/p} &= \left( \int_0^1 \lim_{\delta \rightarrow 0} |D_\delta(u)|^p du \right)^{1/p}, \end{aligned}$$

where the last equality follows from the Lebesgue dominated convergence theorem. Thus the proposition follows.  $\diamond$

Observe that for Example 2.6 the proposition is trivially satisfied as  $g(x) = f(x)$ .

### 3 Connections with reliability theory

In reliability theory the notion of dependence has been widely studied and various concepts of positive dependence have been proposed; see Lehmann (1966) and the recent exposition by Lai and Xie (2006). Here we shall discuss how our concept of dependence is related to those in reliability theory.

Note that from the definition of  $G$  in (1),  $G(x, \cdot)$  is nondecreasing for any fixed  $x$ . Imposing certain conditions on  $G(\cdot, u)$  amounts to assumptions on the character of dependence. In particular, it follows immediately from the definition of  $G$  that a property that  $G(\cdot, u)$  is nondecreasing is equivalent to positive regression dependence (PRD) introduced in Lehmann (1966), for which family of conditional distributions  $F_{Y|X}(\cdot|x)$  are ordered w.r.t.  $x$ . This property is also called  $Y$  stochastically increasing on  $X$  (SI) in Barlow and Proschan (1975).

**Proposition 0.8.**  $G(\cdot, u)$  is nondecreasing for any  $u$  if and only if  $\mathbb{P}(Y > y|X = \cdot)$  is nondecreasing for any  $y$ .

Define  $G^{-1}(y|x) = \inf\{u : G(x, u) \geq y\}$ . Obviously, in the case when  $F_{Y|X}(\cdot|x)$  is continuous and strictly increasing for any  $x$  then  $G^{-1}(y|x) = F_{Y|X}(y|x)$ . Define  $G_{x',x}(u) = G^{-1}(G(x, u)|x')$ . Then positive regression dependence is equivalent to  $G_{x',x}(u) \leq u$  for any  $x' > x$  (Capéraà and Genest, 1990).

Note that the above obvious equivalence of Proposition 0.8 yields a straightforward proof that positive regression dependence implies association i.e. that

$$\text{Cov}(f_1(X, Y), f_2(X, Y)) \geq 0 \tag{10}$$

holds for any  $f_1, f_2$  which are coordinate-wise nondecreasing and such that the covariance exists. This has been proved first by Esary et al. (1967). Namely, observe that the functions  $h_i(x, u) = f_i(x, G(x, u))$ ,  $i = 1, 2$ , are nondecreasing in both arguments. Thus

$$\mathbb{E}(h_1(X, U') - h_1(X', U'))(h_2(X, U) - h_2(X', U)) \geq 0$$

and

$$\mathbb{E}(h_1(X, U) - h_1(X, U'))(h_2(X, U) - h_2(X, U')) \geq 0,$$

where  $(X', U')$  is an independent copy of  $(X, U)$ . Summing both inequalities we obtain  $2\text{cov}(h_1(X, U), h_2(X, U)) \geq 0$ . The property has natural generalization to multivariate case which we will discuss later. Lehmann (1966) has shown that PRD implies that  $\mathbb{P}(Y > y|X \leq \cdot)$  is nondecreasing, whereas the fact that  $\mathbb{P}(Y > y|X > \cdot)$  is nondecreasing implies association of  $X$  and  $Y$  (Esary and Proschan, 1972).

In particular, association implies  $\text{cov}(f(X), g(Y)) \geq 0$  for any nondecreasing  $f$  and  $g$ . This is easily seen to imply positive quadrant dependence (PQD) condition  $\mathbb{P}(X > x, Y > y) \geq \mathbb{P}(X > x)\mathbb{P}(Y > y)$  for  $x, y \in \mathbb{R}$ . The following generalization is possible. If  $(X_i, Y_i), i = 1, 2, \dots, n$  are independent copies of  $(X, G(X, U))$  with

$G(\cdot, u)$  nondecreasing then it follows from Lehmann (1966) that if  $r, s : \mathbb{R}^n \rightarrow \mathbb{R}$  are concordant functions such that direction of monotonicity for each univariate argument is the same then  $\text{cov}(r(\mathbf{X}), s(\mathbf{Y})) \geq 0$ , where  $\mathbf{X} = (X_1, X_2, \dots, X_n)$  and  $\mathbf{Y} = (Y_1, Y_2, \dots, Y_n)$ .

Observe also that for nondecreasing  $G(\cdot, u)$ ,  $Y = G(X, U)$  is a monotone mixture with  $X$  (c.f. Jogdeo (1978)) meaning that the function  $h(x) := \mathbb{E}(f(G(X, U)) | X = x)$  is nondecreasing for any nondecreasing function  $f$ .

A pair  $(X, Y)$  is likelihood ratio dependent (LRD) if its bivariate density  $f(x, y)$  is totally positive of order 2 i.e. for any  $x < x'$  and  $y < y'$ ,  $f(x, y)f(x', y') \geq f(x, y')f(x', y)$ . Then provided that  $F_{Y|X}(\cdot | x)$  is continuous and strictly increasing, LRD is equivalent to convexity of  $G_{x'x}$  for any  $x < x'$ , where  $G_{x'x}$  is the function defined above in terms of the function  $G$ .

We now discuss how some multivariate extensions of these concepts of positive dependence can be interpreted in the presented framework. Note that in the construction of function  $G$  one can assume that  $\mathbf{X} = (X_1, X_2, \dots, X_m)$  is a multivariate vector. Thus we have the representation  $(\mathbf{X}, Y) =_{\mathcal{D}} (\mathbf{X}, G(\mathbf{X}, U))$ . Another sequential construction is discussed in the next section. We now prove the extension of (10) to the case when  $G(\mathbf{x}, u)$  is nondecreasing in any coordinate of  $\mathbf{x}$ .

**Proposition 0.9.** *Assume that  $\mathbf{X}$  is associated and  $G(\mathbf{x}, u)$  is nondecreasing in any coordinate of  $\mathbf{x}$ . Then  $(\mathbf{X}, G(\mathbf{X}, u))$  is associated.*

*Proof.* Define  $h_i(\mathbf{x}, u) = f_i(\mathbf{x}, G(\mathbf{x}, u))$  for  $i = 1, 2$ . Then

$$\mathbb{E}\{(h_1(\mathbf{X}, U) - h_1(\mathbf{X}, U'))(h_2(\mathbf{X}', U) - h_2(\mathbf{X}', U'))\} \geq 0.$$

Moreover, for any fixed  $u$  as  $h_i(\mathbf{x}, u)$  are nondecreasing in each coordinate of  $\mathbf{x}$  we have

$$0 \leq 2\text{cov}(h_1(\mathbf{X}, u), h_2(\mathbf{X}, u)) = \mathbb{E}\{(h_1(\mathbf{X}, u) - h_1(\mathbf{X}', u))(h_2(\mathbf{X}, u) - h_2(\mathbf{X}', u))\}$$

since  $\mathbf{X}$  is associated. Integrating the second inequality w.r.t. distribution of  $U$  and adding the result to the first inequality we get  $2\text{cov}(h_1(\mathbf{X}, U), h_2(\mathbf{X}, U)) \geq 0$  which proves the assertion.  $\diamond$

Proposition 0.9 is a generalization of (10) by noting that a univariate variable  $X$  is always associated and thus  $(X, G(X, U))$  is associated for nondecreasing  $G(\cdot, u)$ .

Consider the following generalization of positive regression dependence. The random vector  $\mathbf{X}$  is conditionally increasing in sequence (CIS) if  $\mathbb{P}(X_i > x | \mathbf{X}_{i-1} = \mathbf{x}_{i-1})$  is increasing in  $x_1, \dots, x_{i-1}$  for any  $i$ , where  $\mathbf{X}_{i-1} = (X_1, \dots, X_{i-1})$  and  $\mathbf{x}_{i-1} = (x_1, \dots, x_{i-1})$ . Define  $G_i$  as the function  $G$  for  $\mathbf{X} = \mathbf{X}_{i-1}$  and  $Y = X_i$  for  $i = 2, \dots, n$ . Then CIS is equivalent to  $G_i(\mathbf{x}, u)$  being nondecreasing in each coordinate of  $\mathbf{x}$ ,  $i = 2, \dots, n$ . Then it follows easily by induction from Proposition 0.9 that CIS implies positive association (for a different proof see Barlow and Proschan (1975)). Also the stronger property of  $\mathbf{X}$  being conditionally increasing (CI) (see e.g. Drouet Mari and Kotz (2005), p. 39) can be expressed by monotonicity of appropriately defined functions  $G$ .

### 4 Multivariate dependence

When analyzing complicated stochastic systems, one has to deal with multivariate dependence. To this end, we can introduce a multivariate version of (3). Earlier related discussions are given in Rosenblatt (1952). Consider the random vector  $(X_1, \dots, X_n)$ . Let  $\mathbf{X}_m = (X_1, \dots, X_m)$ . As in (2), there exists a measurable function  $G_n$  and a standard uniform random variable  $U_n$ , independent of  $\mathbf{X}_{n-1}$ , such that

$$\mathbf{X}_n = \mathcal{G}(\mathbf{X}_{n-1}, G_n(\mathbf{X}_{n-1}, U_n)). \tag{11}$$

Here  $G_n$  is the conditional quantile of  $X_n$  given  $\mathbf{X}_{n-1} = (X_1, \dots, X_{n-1})$ . In the theory of risk management,  $G_n(\mathbf{X}_{n-1}, u)$  is the value-at-risk (VaR) at level  $u$  [cf. J. P. Morgan (1996)]. Namely, given  $\mathbf{X}_{n-1}$  which represents the information available up to time  $n - 1$ ,  $G_n(\mathbf{X}_{n-1}, u)$  is the threshold or risk level at which the probability of  $X_n$  exceeds is  $u$ . Let  $\mathbf{U}_m = (U_1, \dots, U_m)$ . Dependence structure of  $(X_1, X_2, \dots, X_n)$  is essential in determining VaRs. For a readable introduction see Embrechts et al. (2002).

Iterating (11), we have

$$\begin{pmatrix} X_1 \\ X_2 \\ \dots \\ X_n \end{pmatrix} = \mathcal{G} \left( \begin{pmatrix} X_1 \\ G_2(\mathbf{X}_1, U_2) \\ \dots \\ G_n(\mathbf{X}_{n-1}, U_n) \end{pmatrix} \right) = \mathcal{G} \left( \begin{pmatrix} H_1(\mathbf{U}_1) \\ H_2(\mathbf{U}_2) \\ \dots \\ H_n(\mathbf{U}_n) \end{pmatrix} \right), \tag{12}$$

where  $H_1, \dots, H_n$  are measurable functions. The above distributional equality serves as a standard method of simulating multivariate distributions (see e.g. Deák (1990), chapter 5) and the pertaining method is called the standard construction. The above representation is used in Rüschendorf and de Valk (1993) to characterize some subclasses of Markov sequences. Note that for (12) no regularity of distribution of  $\mathbf{X}$  is needed in contrast to the corresponding construction yielding  $F(\mathbf{X}) = \mathbf{U}_n$  almost surely for which absolute continuity of distribution of  $\mathbf{X}$  is usually assumed. For other uses of the standard construction see e.g. Arjas and Lehtonen (1978) and Li et al. (1996). If  $(X_i)$  is a Markov chain, then in view of definition of  $G_i$  as a conditional quantile  $G_i(\mathbf{X}_{i-1}, U_i)$  in (12) can be written as a function of  $X_{i-1}$ . Denote by  $H_j(\mathbf{Y}_{j-1}, u)$ ,  $j = 2, \dots, n$  the analogue of the function  $G_j$  defined in (12) for the vector  $\mathbf{Y}_n = (Y_1, Y_2, \dots, Y_n)$ . It is proved in Arjas and Lahtonen (1978) that if  $X_1$  is stochastically smaller than  $Y_1$  and  $G_i(\mathbf{x}_{i-1}, u) \leq H_i(\mathbf{y}_{i-1}, u)$  for any  $\mathbf{x}_{i-1} \leq \mathbf{y}_{i-1}$  then there exist copies of  $\tilde{\mathbf{X}}_n$  and  $\tilde{\mathbf{Y}}_n$  of  $\mathbf{X}_n$  and  $\mathbf{Y}_n$  respectively, such that  $\tilde{\mathbf{X}}_n \leq \tilde{\mathbf{Y}}_n$  almost surely. This is equivalent to  $\mathbb{E}f(\mathbf{X}) \leq \mathbb{E}f(\mathbf{Y})$  for any nondecreasing  $f$  (cf e.g. Liggett (1985)).

In the sequel we shall discuss the problem whether elements of an infinite sequence  $(X_i)_{i \in \mathbb{Z}}$  can be expressed as functions of iid random variables, namely whether an analogue of (12) with infinitely many rows holds. With such representations, one can establish limit theorems and moment inequalities under simple and easily workable conditions; see Section 5. The latter problem belongs to the stochas-

tic realization theory (Borkar, 1993). Wiener (1958) first considered this representation problem for stationary and ergodic processes. As a special case, if  $(X_i)$  is a Markov chain with the form  $X_i = G_i(X_{i-1}, U_i)$ , Theorem 4.1 asserts that, under some regularity conditions there exists a copy  $\tilde{X}_i$  of  $X_i$  such that  $(\tilde{X}_i)_{i \in \mathbb{Z}} =_{\mathcal{D}} (X_i)_{i \in \mathbb{Z}}$  and  $\tilde{X}_i$  is expressed as  $H_i(\dots, U_{i-1}, U_i)$ , a function of iid random variables.

**Theorem 4.1.** *Assume that  $(X_i)$  satisfies the recursion*

$$X_i = G_i(X_{i-1}, U_i) =: F_i(X_{i-1}), \quad i \in \mathbb{Z}, \tag{13}$$

where  $U_i$  are iid standard uniform random variables. Here  $F_i$  are independent random maps  $F_i(x) = G_i(x, U_i)$ . Assume that for some  $\alpha > 0$  we have

$$\sup_{i \in \mathbb{Z}} L_i < 1, \text{ where } L_i = \sup_{x \neq y} \frac{\|G_i(x, U) - G_i(y, U)\|_{\alpha}}{|x - y|} \tag{14}$$

and for some  $x_0$ ,

$$\sup_{i \in \mathbb{Z}} \|G_i(x_0, U)\|_{\alpha} < \infty. \tag{15}$$

Then the backward iteration  $F_i \circ F_{i-1} \circ F_{i-2} \dots$  converges almost surely and the limit forms a non-stationary Markov chain which is a solution to (13).

*Proof.* The theorem can be proved along the same line as Theorem 2 in Wu and Shao (2004), where the stationary case with  $G_i \equiv G$  is dealt with. Their argument improves the classical result by Diaconis and Freedman (1999) and it can be easily generalized to the current non-stationary setting. Here we provide an outline of the argument and omit the details. Let  $\mathcal{X}$  be the state space of  $X_i$ . For  $i \geq j$  define the random map  $F_{i,j}$  by

$$F_{i,j}(x) = F_i \circ F_{i-1} \circ \dots \circ F_j(x), \quad x \in \mathcal{X},$$

where  $\circ$  denotes function composition. Let  $q = \min(1, \alpha)$ . By (14) and (15), there exists a  $\rho \in (0, 1)$  such that  $\mathbb{E}\{|F_{i,j}(x_0) - F_{i,j+1}(x_0)|^q\} = O(\rho^{i-j})$ .

Hence  $\sum_{j=-\infty}^i |F_{i,j}(x_0) - F_{i,j+1}(x_0)|^q < \infty$  almost surely and the sequence  $F_{i,j}(x_0)$ ,  $j = i, i-1, i-2, \dots$ , has an almost sure limit  $Z_i$  (say). One can show that  $Z_i \in \mathcal{L}^{\alpha}$ . Also, following the argument of Theorem 2 in Wu and Shao (2004), we have for any  $x$ , the sequence  $F_{i,j}(x)$ ,  $j = i, i-1, i-2, \dots$ , has the same limit  $Z_i$  which does not depend on  $x$ . Let  $X_i = Z_i$ , then (13) holds.  $\diamond$

Next we consider the problem of whether stationary and ergodic processes can be expressed as functionals of iid random variables. Let  $(X_i)_{i \in \mathbb{Z}}$  be a stationary and ergodic process. Following (12), one may think whether it can be generalized to the case of infinite sequences, namely whether one can find iid standard uniform random variables  $U_j$  and a measurable function  $H$  (recalling stationarity) such that

$$(X_i)_{i \in \mathbb{Z}} =_{\mathcal{D}} (H(\dots, U_{i-1}, U_i))_{i \in \mathbb{Z}}. \tag{16}$$

Intuitively, using  $(X_i)_{i \in \mathbb{Z}} = (G(\mathbf{X}_{i-1}, U_i))_{i \in \mathbb{Z}}$ , one can recursively write

$$X_i =_{\mathcal{D}} G(\mathbf{X}_{i-1}, U_i) =_{\mathcal{D}} G_2(\mathbf{X}_{i-2}; U_{i-1}, U_i) =_{\mathcal{D}} G_3(\mathbf{X}_{i-3}; U_{i-2}, U_{i-1}, U_i) = \dots \quad (17)$$

A major difficulty here is that the sequence  $G_j(\cdot; U_{i-j+1}, \dots, U_{i-1}, U_i)$  may not have a limit. In Theorem 4.1, due to the special Markovian structure and conditions (14) and (15), it is proved that the latter sequence does converge. In general the convergence is not guaranteed. Wiener (1958) argued that (16) should hold in a stronger sense that the distributional equality is an equality. Rosenblatt (1959) showed that Wiener’s result is not true. See also Rosenblatt (1971), Kallianpur (1981) and Tong (1990) for more details. D.Volny (personal communications) pointed out that (16) is generally not valid. Processes of the form  $(H(\dots, U_{i-1}, U_i))_{i \in \mathbb{Z}}$  are necessarily Bernoulli (Ornstein, 1973). It is known that some stationary and ergodic processes are not Bernoulli (see Ornstein (1973) and Kalikow (1982)). One such example is given in Rosenblatt (2009).

## 5 Moment inequalities and limit theorems

Let  $(X_i)$  be a stochastic process. Knowledge of  $S_n = X_1 + \dots + X_n$  is of critical importance in the study of random processes and various dependence measures for stochastic processes are relevant tools for this purpose. We mention construction of various mixing coefficients (see e.g. Bradley (2005)), weak dependence coefficients discussed in Dedecker et al. (2007), projective measures of dependence (Gordin (1969)) and measures based on mutual information of adjacent blocks of observations such as an excess entropy (Crutchfeld and Feldman (2003)). Moreover, special types of dependence, such as association, are useful for studying asymptotic properties of the partial sums (see e.g. Newman (1984)). Doukhan and Louhichi (1999) proposed an important dependence measure which concerns the decay rates of covariances between functionals of the past and the future of the underlying processes.

Here we shall study  $S_n$  for causal processes by generalizing dependence measures given in Definition 0.1 to stochastic processes. Let  $\varepsilon_i, i \in \mathbb{Z}$ , be iid random variables and let  $\xi_i = (\dots, \varepsilon_{i-1}, \varepsilon_i)$ . Consider the infinite stochastic system

$$X_i = H_i(\xi_i), \quad i \in \mathbb{Z}, \quad (18)$$

where  $H_i$  are measurable functions. We view  $\xi_i$  as the input and  $X_i$  as the output of the system. If  $H_i$  does not depend on  $i$ , then the process  $(X_i)$  generated by (18) is stationary. Here we are interested in the behavior of  $S_n = X_1 + \dots + X_n$ , which is helpful in understanding properties of the underlying stochastic system.

It is usually difficult to know or derive the close form of the distribution of  $S_n$ . For large values of  $n$  asymptotic theory provides a useful approximation. To this end we shall introduce functional and predictive dependence coefficients. They are defined for a certain class of causal processes defined in (18), where  $(\varepsilon_i)$  is an iid sequence. Let  $(\varepsilon'_i)_{i \in \mathbb{Z}}$  be an iid copy of  $(\varepsilon_i)_{i \in \mathbb{Z}}$  and  $\xi_{i,j} = (\dots, \varepsilon_{i-1}^*, \varepsilon_i^*)$ , where  $\varepsilon_l^* = \varepsilon_l$  if  $l \neq j$

and  $\varepsilon_j^* = \varepsilon_j'$ . Namely  $\xi_{i,j}$  is a coupled version of  $\xi_i$  with  $\varepsilon_j$  in the latter replaced by  $\varepsilon_j'$ . For  $k \geq 0$  and  $p \geq 1$  define

$$\delta_{k,p} = \sup_{i \in \mathbb{Z}} \|H_i(\xi_i) - H_i(\xi_{i,i-k})\|_p \tag{19}$$

and, let the projection operator  $\mathcal{P}_j(\cdot) = \mathbb{E}(\cdot | \xi_j) - \mathbb{E}(\cdot | \xi_{j-1})$ ,

$$\theta_{k,p} = \sup_{i \in \mathbb{Z}} \|\mathcal{P}_{i-k} H_i(\xi_i)\|_p = \sup_{i \in \mathbb{Z}} \|\mathbb{E}[H_i(\xi_i) | \xi_{i-k}] - \mathbb{E}[H_i(\xi_{i,i-k}) | \xi_{i-k}]\|_p. \tag{20}$$

The second equality in (20) is due to the fact that:

$\mathbb{E}[H_i(\xi_i) | \xi_{i-k-1}] = \mathbb{E}[H_i(\xi_{i,i-k}) | \xi_{i-k-1}] = \mathbb{E}[H_i(\xi_{i,i-k}) | \xi_{i-k}]$ . We observe that  $\delta_{k,p}$  equals  $\sup_{i \in \mathbb{Z}} \delta_p(\varepsilon_{i-k}, H_i(\xi_i))$  defined in Definition 0.1 i.e.  $H_i(\xi_i)$  plays a role of  $Y$  and  $\varepsilon_{i-k}$  a role of  $X$ . Thus the strength of dependence of the random variable  $X_i = H_i(\xi_i)$  observed at moment  $i$  on the innovation  $\varepsilon_{i-k}$  is quantified. Note that the counterpart of  $\tau_p$  would be  $\tau_{k,p} = \sup_{i \in \mathbb{Z}} \|H_i(\xi_i) - \mathbb{E}(H_i(\xi_i) | \xi_{i-k})\|_p$ . Wu (2005) introduced functional and predictive dependence coefficients for stationary processes. Note that in (19) we assume that  $\varepsilon_i, \varepsilon_j', i, j \in \mathbb{Z}$ , are iid. It is meaningless if the independence assumption is violated. Under the construction (12), they are automatically iid uniform(0,1) random variables.

The following theorem concerns moments of  $S_n$  and can be proved along the lines of Wu (2007). We refer to Gordin and Lifsic (1978) and Hannan (1979) who used a representation of a random variable as a sum of such projections to derive asymptotic results. A similar approach was later used by several authors, see Woodrooffe (1992), Volny (1993), Dedecker and Merlevède (2002), Wu and Woodrooffe (2004), Bryk and Mielniczuk (2005) and Merlevède and Peligrad (2006) among others.

**Theorem 5.1.** *Let  $p > 1$ ,  $p' = \min(2, p)$  and  $B_p = 18p^{3/2}/\sqrt{p-1}$ . Assume that*

$$\Theta_{0,p} := \sum_{i=0}^{\infty} \theta_{i,p} < \infty. \tag{21}$$

*For  $k \geq 0$  let  $\Theta_{k,p} = \sum_{i=k}^{\infty} \theta_{i,p}$ . Then we have (i)  $D_j = \sum_{i=j}^{\infty} \mathcal{P}_j X_i$ ,  $j \in \mathbb{Z}$ , form martingale differences with respect to  $\xi_j$  and  $M_n = \sum_{i=1}^n D_i$  satisfies*

$$\|S_n - M_n\|_p^{p'} \leq 3B_p^{p'} \sum_{j=1}^n \Theta_{k,p}^{p'}, \tag{22}$$

*and (ii)  $S_n^* = \max_{j \leq n} |S_j|$  satisfies  $\|S_n^*\|_p \leq C_p n^{1/p'} \Theta_{0,p}$ , where  $C_p = pB_p/(p-1)$ .*

Theorem 5.1(ii) is closely related to the Efron-Stein inequality. Consider the special case  $p = 2$  and let  $X = G(\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n)$  be a function of independent  $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n$  and  $X_i = G(\varepsilon_1, \varepsilon_2, \dots, \varepsilon_i', \dots, \varepsilon_n)$ . Then Efron-Stein inequality states that  $\text{var}(X) \leq 2^{-1} \sum_{i=1}^n \mathbb{E}(X - X_i)^2$  (Efron and Stein, 1981).

**Theorem 5.2.** *Assume that  $X_i \in \mathcal{L}^p$ ,  $2 < p \leq 4$ , and  $\mathbb{E}(X_i) = 0$ . Further assume that*



$$\sum_{k=1}^{\infty} (\delta_{k,p} + k\theta_{k,p}) < \infty. \tag{23}$$

Let  $\sigma_i = \|D_i\|$ , where  $D_j = \sum_{i=j}^{\infty} \mathcal{P}_j X_i$ . Then on a richer probability space, there exists a standard Brownian motion  $\mathcal{B}$  such that

$$\max_{i \leq n} |S_i - \mathcal{B}(\sigma_1^2 + \dots + \sigma_i^2)| = o_{a.s.}[n^{1/p}(\log n)^{1/2+1/p}(\log \log n)^{2/p}]. \tag{24}$$

The approximation (24) suggests that  $S_i$  can be "regularized" by a Gaussian process  $Z_i = \mathcal{B}(\sigma_1^2 + \dots + \sigma_i^2)$  which has independent but non-stationary increments  $Z_i - Z_{i-1}$ . If  $X_i$  are iid with  $X_i \in \mathcal{L}^p$ , then the classical Hungarian strong embedding (strong invariance principle) asserts that  $\max_{i \leq n} |S_i - \mathcal{B}(i\sigma^2)| = o_{a.s.}(n^{1/p})$ , where  $\sigma = \|X_i\|$ . For stationary processes Wu (2007) obtained (24). A careful check of the proof in the last paper shows that (24) also hold under condition (23). The details are omitted.

**Acknowledgements** We would like to thank a referee for her/his many useful comments.

## References

- [1] Arjas, E. and Lehtonen, T. (1978) Approximating many server queues by means of single server queues. *Math. Operation Research* **3**, 205-223.
- [2] Barlow, R. and Proschan, F. (1975) *Statistical theory of reliability and life testing*, Rinehart and Winston, New York.
- [3] Bjerve S. and Doksum, K. (1993) Correlation curves: measures of association as functions of covariates values. *Annals of Statistics* **21**, 890-902.
- [4] Blyth, S. (1994) Local divergence and association. *Biometrika* **81**, 579-584.
- [5] Block, H. W., Sampson A. R. and Savits, T. H. (1990) (eds) *Topics in statistical dependence*, IMS Lecture Notes-Monograph Series, **16** IMS, Hayward, CA.
- [6] Borkar, V. S. (1993) White-noise representations in stochastic realization theory. *SIAM J. Control Optim.* **31** 1093-1102.
- [7] Bradley, R. (2005) Basic properties of strong mixing conditions. A Survey and some open questions. *Probability Surveys* **2**, 107-144
- [8] Bryk, A. and Mielniczuk, J. (2005) Asymptotic properties of kernel density estimates for linear processes: application of projection method. *Nonparametric Statistics* **14**, 121-133.
- [9] Capéreaà, P. and Genest, C. (1990). Concepts de dépendance et ordres stochastiques pour des lois bidimensionnelles. *Canadian Journal of Statistics* **18**, 315-326.
- [10] Cowell, R. G., Dawid, A. P. and Spiegelhalter, D. J. (1999) *Probabilistic Networks and Systems*, Springer, New York.

- [11] Crutchfield, J.P. and Feldman, D.P. (2003) Regularities unseen, randomness observed: The entropy convergence hierarchy. *Chaos* **15**, 25-54.
- [12] Deák, I. (1990) *Random Numbers Generators and Simulation*, Akadémiai Kiadó, Budapest.
- [13] Dedecker, J., Merlevède, F. (2002) Necessary and sufficient conditions for the conditional central limit theorem. *Ann. Probab.* **30**, 1044–1081.
- [14] Dedecker, J., P. Doukhan, G. Lang, J.R. Leon R., S. Louhichi and C. Prieur (2007) *Weak Dependence: With Examples and Applications*, Springer, New York.
- [15] Diaconis, P. and Freedman, D. (1999) Iterated random functions. *SIAM Review* **41** 41-76.
- [16] Doksum, K. and Froda, S. M. (2000) Neighborhood correlation. *J. Statist. Plan. Inf.* **91**, 267–294.
- [17] Doksum, K. and Samarov, A. (1995) Nonparametric estimation of global functionals and measure of the explanatory powers in regression. *Annals of Statistics* **23**, 1443-1473.
- [18] Doukhan, P. and Louhichi, S. (1999) A new weak dependence condition and applications to moment inequalities. *Stochastic Process. Appl.* **84** 313–342.
- [19] Drouot Mari, D. and Kotz, S. (2001) *Correlation and dependence*, Imperial College Press, London.
- [20] Efron, B. and Stein, C. (1981) The jackknife estimate of variance. *Annals of Statistics* **9**, 586–596.
- [21] Embrechts, P., McNeil, A. and Straumann, D. (2002) Correlation and dependence in risk management: properties and pitfalls. In: *Risk Management: Value at Risk and Beyond*. (M.A.H. Dempster, ed.), Cambridge University Press, 176-223.
- [22] Esary, J. D., Proschan, F. and Walkup, D.W. (1967) Association of random variables with applications, *Annals of Mathematical Statistics* **38**, 1466–1474.
- [23] Esary, J. D. and Proschan, F. (1972) Relationships between some concepts of bivariate dependence. *Annals of Mathematical Statistics* **43**, 651-655.
- [24] Glymour, C. (2006) Causation-II. *Encyclopedia of Statistical Sciences*, Wiley. 782–794
- [25] Gordin, M.I. (1969) The central limit theorem for stationary processes. *Dokl. Akad. Nauk SSSR* **188**, 739–741
- [26] Gordin, M.I. and Lifsic, B.A.(1973) The central limit theorem for stationary Markov processes. *Dokl. Akad. Nauk SSSR* **239**, 392–393
- [27] Granger, C.W.J. (1969) Investigating causal relations by econometric models and cross-spectral methods. *Econometrica* **37**, 424–438.
- [28] Hannan, E.J. (1979) The central limit theorem for time series regression. *Stochastic Processes Appl.*, **9**, 281–289
- [29] Joe, H. (1997) *Multivariate models and dependence concepts*, Chapman and Hall, London.
- [30] Jogdeo, K. (1978) On a probability bound of Marshall and Olkin. *Ann. Statist.* **6**, 232–234.

- [31] Jogdeo, K. (2006) Concepts of dependence. *Encyclopedia of Statistical Sciences*, Wiley, 163–1647.
- [32] Kalikow, S. A. (1982)  $T, T^{-1}$  transformation is not loosely Bernoulli. *Ann. Math.* **115**, 393–409.
- [33] Kallianpur, G. (1981) Some ramifications of Wiener's ideas on nonlinear prediction. In: *Norbert Wiener, Collected Works with Commentaries*, MIT Press, Mass., 402–424.
- [34] Lai, C. D. and Xie, M. (2006) *Stochastic Ageing and Dependence for Reliability*. Springer, New York.
- [35] Lehmann, E. L. (1966) Some concepts of dependence. *Ann. Math. Statist.* **37**, 1137–1153.
- [36] Li, H. , Scarsini, M. and Shaked, M. (1996) Linkages: A tool for construction of multivariate distributions with nonoverlapping multivariate marginals. *J. Multivariate. Anal.* **56**, 20–41.
- [37] Liggett, T.M. (1985) *Interacting Particle Systems*, Springer, New York
- [38] Mellor, D. H. (1998) *The Facts of Causation*, Routledge, New York.
- [39] Merlevède and F., Peligrad, M. (2006) On the weak invariance principle for stationary sequences under projective criteria. *J. Theor. Probab.* **19**, 647–689
- [40] J. P. Morgan (1996) *RiskMetrics*. Technical Document. New York.
- [41] Mosteller, F. and Tukey, J.W. (1977) *Data Analysis and Regression*, Addison–Wesley Reading, Mass.
- [42] Newman, C.M. (1984) Asymptotic independence and limit theorems for positively and negatively dependent random variables. In: *Inequalities in Statistics and Probability* (Y.L. Tong, ed.), IMS Lecture Notes - Monograph Series 7, 127–140.
- [43] Ornstein, D. S. (1973) An example of a Kolmogorov automorphism that is not a Bernoulli shift. *Advances in Math.* **10** 49–62
- [44] Rényi, A. (1959) On measures of dependence. *Acta Math. Acad. Sci. Hungar.* **10** 441–451
- [45] Rosenblatt, M. (1952). Remarks on a multivariate transformation. *Ann. Math. Statist.* **23** 470–472.
- [46] Rosenblatt, M. (1959). Stationary processes as shifts of functions of independent random variables. *J. Math. Mech.* **8**, 665–681.
- [47] Rosenblatt, M. (1971) *Markov Processes. Structure and Asymptotic Behavior*, Springer, New York.
- [48] Rosenblatt, M. (2009) A comment on a conjecture of N. Wiener. *Statist. Probab. Letters*, **79**, 347–348
- [49] Rubin, D. (1990) Formal modes of statistical inference for causal effects. *Journal of Statistical Planning and Inference* **25**, 279–292.
- [50] Rüschendorf, L. (1981) Stochastically ordered distributions and monotonicity of the OC-function of sequential probability ratio tests. *Mathematische Operationsforschung und Statistik. Series Statisticis* **12**, 327–338.
- [51] Rüschendorf, L. and de Valk, V.(1981) On regression representation of stochastic processes. *Stochastic Processes and Their Applications* **46**, 183–198.

- [52] Scarsini, M. (1984) On measures of concordance. *Stochastica* **8**, 201–218.
- [53] Skaug, H. J. and Tjøstheim, D. (1993) A nonparametric test of serial independence based on the empirical distribution function. *Biometrika* **80** 591–602.
- [54] Stigler, S. M. (1986) *The history of statistics*, Harvard University Press, Cambridge, MA.
- [55] Szekli, R. (1995) *Stochastic ordering and dependence in applied probability*, Springer, New York
- [56] Tjøstheim, D. (1996) Measures of dependence and tests of independence. *Statistics* **28** 249–284.
- [57] Tong, H. (1990) *Nonlinear time series analysis: A dynamic approach.*, Oxford University Press, Oxford.
- [58] Volný, D. (1993) Approximating martingales and the central limit theorem for strictly stationary processes. *Stochastic Processes and their Applications* **44**, 41–74
- [59] Wiener, N. (1958) *Nonlinear Problems in Random Theory.*, MIT Press, Cambridge, MA.
- [60] Woodroffe, M. (1992) A central limit theorem for functions of a Markov chain with applications to shifts. *Stochastic Processes and Their Applications* **41**, 33–44.
- [61] Wu, W. B. and Woodroffe, M. (2004) Martingale approximations for sums of stationary processes. *Ann. Probab.* **32**, 1674–1690.
- [62] Wu, W. B. (2005) Nonlinear system theory: Another look at dependence. *Proceedings of the National Academy of Sciences, USA*, **102**, 14150–14154.
- [63] Wu, W. B. (2007) Strong invariance principles for dependent random variables. *Ann. Probab.*, **35**, 2294–2320.
- [64] Wu, W. B. (2008) Empirical processes of stationary sequences. *Statistica Sinica*, **18** 313–333.
- [65] Wu, W. B. and Shao, X. (2004) Limit theorems for iterated random functions. *Journal of Applied Probability*, **41** 425–436.