TSI TSI TSI TSI TSI TSI TSI TSI TSI TSI TSI TSI TSI TSI TSI TSI TSI TSI TSI TSI TSI TSI TSI TSI TSI TSI TSI TSI

A Relationship Strictly Between Uncorrelated and Independent

Dr. Ralph deLaubenfels

TSI TSI TSI TSI TSI TSI TSI TSI TSI TSI TSI TSI TSI TSI TSI TSI TSI TSI TSI TSI TSI TSI TSI TSI TSI TSI TSI TSI

Teacher-Scholar Institute

Columbus, Ohio

2011

A RELATIONSHIP STRICTLY BETWEEN INDEPENDENT AND UNCORRELATED

RALPH DELAUBENFELS

ABSTRACT. For X, Y two random variables, we say that Y is *unpredictable* by X if $E(Y|X = x) = E(Y)$, for all x. We show that X and Y independent implies Y is unpredictable by X, and Y unpredictable by X implies X and Y are uncorrelated. We show by example that these implications are strict; that is, we produce examples, as simple as possible, of dependent X and Y such that Y is unpredictable by X, and of uncorrelated X and Y such that Y is predictable by X (meaning the double negative "Y is not unpredictable by X").

I. INTRODUCTION. An interesting paper in The American Statistician (David 2009) described surprisingly recent confusion between being independent and being uncorrelated. Thus it seems of interest that there is a natural relationship between two random variables that is strictly between being independent and being uncorrelated.

Very roughly, the motivation for considering such relationships is the following. One may have two random variables X and Y such that X is easy to measure but of no intrinsic value, while Y is something we care about but is very hard to measure directly. It is then desirable to use measurements of X to gain information about Y.

The goal of this short note is to make clear, by simple examples and equivalences, the relationship between being uncorrelated, independent, and what we will call unpredictable.

Definition. Let X and Y be two random variables. We will say that Y *is unpredictable by* X if $E(Y|X = x) = E(Y)$, for all x in the support of the distribution of X. Otherwise we will say that Y *is predictable by* X.

Being unpredictable is strictly between being uncorrelated and being independent; that is, being unpredictable implies being uncorrelated, but not the converse, while being independent implies being unpredictable, but not the converse (see Section III). Thus unpredictability helps clarify the relationship between uncorrelated and independent in the same sense that the discovery of the New World helped clarify the relationship between Europe and Asia.

Similar separations of uncorrelated and independent may be found in the definition of subindependent, in (Hamedani and Volkmer 2009), and an interesting example in (De Paula 2008), which, in the language of this paper, consists of random variables X, Y such that Y^n is unpredictable by X and X^n is unpredictable by Y, for all natural numbers n, while X and Y are dependent.

Unpredictability is shown in this paper to be a particularly simple and natural separation of uncorrelated and independent, warranting an explicit designation, in two ways.

First, a geometric separation is presented in Section II, and may be equivalently described as follows, for random variables X and Y. Y is unpredictable by X if and only if Y and $g(X)$ are uncorrelated, for all (Borel) functions g, while independence is equivalent to $h(Y)$ and $g(X)$ being uncorrelated, for all functions h, g. See the informal comments after each equivalence in Section II for further clarification of the relationship between uncorrelated, unpredictable, and independent.

Second, Example 3.2 in Section III separates uncorrelated, unpredictable, and independent with random variables X, Y such that X is supported on two points and Y is supported on three. This is a bare minimum of probabilistic variability, in which these relationships manifest themselves distinctively; Proposition 3.9 shows that, if both X and Y have at most two points in their support, then uncorrelated, unpredictable, and independent are equivalent.

We shall see (Propositions 3.5 and 3.7) that unpredictability is not a symmetric relationship. Thus we need to also define X, Y to be *mutually predictable (unpredictable)* if Y is predictable (unpredictable) by X and X is predictable (unpredictable) by Y.

If X and Y are dependent but Y is unpredictable by X, although measurements of X theoretically contain information about Y, they might have no value from the point of view of prediction.

Throughout the paper, assume X and Y have finite variances, $V(X)$ and $V(Y)$. "Borel" will be short for "Borel measurable on $\mathbf{R}$."

II. EQUIVALENCES. The fact that being unpredictable is between being uncorrelated and being independent (Corollary 3.1) will be clear from straightforward equivalences, arranged in this section for easy comparison. These probably well-known equivalences will also demonstrate how naturally unpredictability arises in the context of independent versus uncorrelated random variables.

Let's adopt the geometrically natural and useful outlook (see (deLaubenfels 2006; Dudley 2002)) of defining two random variables X and Y to be *orthogonal* if their covariance $\mathrm{Cov}(X,Y) = 0$.

The *orthogonal projection* $P(Y)$ of a random variable Y onto a vector space of random variables $\mathcal{A}$ is a random variable in $\mathcal{A}$ such that $(Y - P(Y))$ is orthogonal to $\mathcal{A}$, that is, to every random variable in $\mathcal{A}$.

It is a fact, immediately believable by a picture, that $P(Y)$ is the best approximation of Y from $\mathcal{A}$, in the sense that it minimizes the variance of the difference:

$$V(Y - P(Y)) \leq V(Y - W), \quad \forall W \in \mathcal{A}.$$

In particular, for X, Y random variables, the conditional expectation $E(Y|X)$ is the orthogonal projection of Y onto $\{g(X) \,|\, g \text{ is Borel}, \; V(g(X)) < \infty\}$, hence the best approximation of Y by functions of X. See (Dudley, Theorem 10.2.9).

Equivalence 2.1. The following are equivalent.

(a) X and Y are uncorrelated.

(b) The orthogonal projection of Y onto $\{f(X) \,|\, f \text{ is linear}\}$ is a constant.

(c) Y is orthogonal to $\{f(X) \,|\, f \text{ is linear}\}$.

(d) $\mathrm{Cov}(X,Y) = 0$.

Proof: The equivalence of (d) and (b) follows from the explicit representation of the orthogonal projection $P(Y)$ onto the vector space in (b) and (c):

$$P(Y) = E(Y) + \frac{\mathrm{Cov}(X,Y)}{V(X)} \left(X - E(X)\right).$$

The equivalence of (b) and (c) follows from the definition of orthogonal projection, while (a) and (d) are equivalent by definition. $\square$

Informally, the equivalences of Equivalence 2.1 are saying there is no (worthwhile) linear relationship between X and Y; the best approximation of Y by linear functions of X is a constant, thus is unaffected by a measurement of X.

Equivalence 2.2. The following are equivalent.

(a) Y is unpredictable by X.

(b) The orthogonal projection of Y onto $\{g(X) \,|\, g \text{ is Borel}, \; V(g(X)) < \infty\}$ is a constant.

(c) Y is orthogonal to $\{g(X) \,|\, g \text{ is Borel}, \; V(g(X)) < \infty\}$.

(d) $\mathrm{Cov}(g(X), Y) = 0$, for all Borel g such that $V(g(X)) < \infty$.

Proof: This follows from the definition of Y being unpredictable by X and the fact that the orthogonal projection in (b) is $E(Y|X)$. $\square$

Informally, Equivalence 2.2 is saying that, when X and Y are mutually unpredictable, there is no (worthwhile) functional relationship between X and Y. When Y is unpredictable by X, there is no informative function of X for Y; the best approximation of Y by functions of X is a constant, thus, as in Equivalence 2.1, is unaffected by a measurement of X.

From a prediction point of view, this is saying that we cannot use measurements of X to predict a measurement of Y; at least not with our favorite predictor, $E(Y|X)$.

Equivalence 2.3. The following are equivalent.

(a) X and Y are independent.

(b) For any Borel h such that $V(h(Y)) < \infty$, the orthogonal projection of $h(Y)$ onto $\{g(X) \mid g \text{ is Borel}, V(g(X)) < \infty\}$ is a constant.

(c) $\{h(Y)) \mid h \text{ is Borel}, V(h(Y)) < \infty\}$ is orthogonal to $\{g(X) \mid g \text{ is Borel}, V(g(X)) < \infty\}$.

(d) $\mathrm{Cov}(g(X), h(Y)) = 0$, for all Borel g, h such that $V(g(X)), V(h(Y)) < \infty$.

Proof: The equivalence of (b), (c), and (d) is clear by our definition of orthogonal. The equivalence of (a) and (d) is relatively well known; we will outline the argument.

The fact that (a) implies (d) (since $g(X)$ and $h(Y)$ are independent) is well known and elementary; for the converse, it is sufficient to take, for arbitrary real s, t,

$$g(x) \equiv e^{-isx}, \quad h(y) \equiv e^{-ity},$$

so that (d) implies that $E(e^{-isX}e^{-itY}) = E(e^{-isX})E(e^{-itY})$ for all real s, t, and uniqueness of Fourier-Stieltjes transforms implies that the joint distribution of X, Y is the product of the marginal distributions of X and Y; that is,

$$dP_{(X,Y)} = dP_X \, dP_Y,$$

which is equivalent to (a). □

Informally, independence means there is no relationship between X and Y, so that measuring X gives no information about Y.

III. RELATIONSHIPS BETWEEN UNCORRELATED, UNPREDICTABLE, AND INDEPENDENT.

Corollary 3.1 follows immediately from the equivalences of the previous section.

Corollary 3.1. Suppose X and Y are random variables, and consider the following relationships.

(a) X and Y are independent.

(b) Y is unpredictable by X.

(c) X and Y are uncorrelated.

Then (a) $\rightarrow$ (b) $\rightarrow$ (c).

We must next demonstrate that these implications are strict; that is, that there exist dependent X, Y such that Y is unpredictable by X, and there exist uncorrelated X, Y such that X is predictable by Y. Both of these will be furnished by the following example.

Example 3.2. For $0 < q < \frac{1}{4}$, define the discrete joint distribution $p(x, y) \equiv P(X = x, Y = y)$ by

$$p(0, \pm 1) = (\frac{1}{4} - q), \quad p(1, \pm 1) = q, \quad p(0, 0) = 2q, \quad p(1, 0) = \frac{1}{2} - 2q.$$

The marginal distributions are

$$\begin{array}{ccc} x & 0 & 1 \\ p(x) & \frac{1}{2} & \frac{1}{2} \end{array} \quad \text{and} \quad \begin{array}{cccc} y & -1 & 0 & 1 \\ p(y) & \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \end{array}$$

This example distinguishes uncorrelated, unpredictable, and independent. Propositions 3.4 and 3.7 demonstrate that X and Y are uncorrelated with X predictable by Y (for $q \neq \frac{1}{8}$); Propositions 3.3 and 3.5 that Y is unpredictable by X, with X and Y dependent (for $q \neq \frac{1}{8}$).

Proposition 3.3. X and Y are independent if and only if $q = \frac{1}{8}$.

Proof:

$$P(X = 1, Y = 1) = q, \quad \text{while} \quad P(X = 1)P(Y = 1) = \frac{1}{8};$$

thus, for independence, it is necessary that $q = \frac{1}{8}$; the sufficiency follows from a few more identical calculations. □

Proposition 3.4. X and Y are uncorrelated.
Proof:

$$E(XY) = P(X=1, Y=1) - P(X=1, Y=-1) = q - q = 0 = (-\frac{1}{4} + \frac{1}{4}) = E(Y).$$

□

Proposition 3.5. Y is unpredictable by X.
Proof:

$$E(Y|X=0) = \frac{(\frac{1}{4}-q)-(\frac{1}{4}-q)}{\frac{1}{2}} = 0 = \frac{q-q}{\frac{1}{2}} = E(Y|X=1).$$

□

Remark 3.6. Proposition 3.5 follows automatically from Proposition 3.4 and Equivalences 2.1 and 2.2, since the support of X contains only two points, thus

$$\{f(X)\,|\,f \text{ is linear }\} = \{g(X)\,|\,g \text{ is Borel, } V(g(X)) < \infty\}.$$

Proposition 3.7. X is unpredictable by Y if and only if $q = \frac{1}{8}$.
Proof:

$$E(X|Y=-1) = \frac{q}{\frac{1}{4}} = 4q = E(X|Y=1), \quad \text{and} \quad E(X|Y=0) = \frac{\frac{1}{2}-2q}{\frac{1}{2}} = 1-4q,$$

thus X is unpredictable by Y if and only if $1-4q = 4q$, or $q = \frac{1}{8}$. □

Remark 3.8. A small calculation shows that $E(X|Y) = (1-4q) + (8q-1)Y^2$; this also implies Proposition 3.7, and, since $E(Y) = 0$, Proposition 3.4.

The following proposition shows that Example 3.2 is the simplest possible.

Proposition 3.9. If X and Y each have at most two points in their support, then (a), (b), and (c) of Corollary 3.1 are equivalent.

Proof: This follows from Equivalences 2.1, 2.2, and 2.3, after the observation that, because both X and Y have at most two points in their support,

$$\{h(Y))\,|\,h \text{ is Borel, } V(h(Y)) < \infty\} = \{f(Y)\,|\,f \text{ is linear}\}$$

and

$$\{g(X)\,|\,g \text{ is Borel, } V(g(X)) < \infty\} = \{f(X)\,|\,f \text{ is linear}\}.$$

□

Remark 3.10. Notice that one consequence of Example 3.2 is that unpredictability is not symmetric: in that example, Y is unpredictable by X but X is predictable by Y (for $q \neq \frac{1}{8}$). See (Schechtman and Yitzhaki 1987) for another very interesting unsymmetric relationship between random variables, Gini covariance and correlation.

This lack of symmetry means we need two other separations of relationships between random variables X and Y: we need dependent X, Y such that X, Y are mutually unpredictable (Example 3.11) and we need uncorrelated X and Y such that X, Y are mutually predictable (Example 3.12).

Example 3.11. This slight modification of Example 3.2 produces X, Y that are mutually unpredictable but dependent.

For $0 < q < \frac{1}{8}$, define the discrete joint distribution $p(x,y) \equiv P(X=x, Y=y)$ by

$$p(\pm 1, \pm 1) = q, \quad p(\pm 1, 0) = p(0, \pm 1) = (\frac{1}{4} - 2q), \quad p(0,0) = 4q.$$

X and Y have the same marginal distribution as Y in Example 3.2.
The dependence of X and Y, when $q \neq \frac{1}{16}$, is clear from

$$P(X=0)P(Y=0) = \frac{1}{4}, \quad P(X=0, Y=0) = 4q.$$

For mutual unpredictability, a short calculation as in Example 3.2 shows that

$$E(X|Y=-1) = P(X|Y=0) = P(X|Y=1) = 0,$$

and similarly for $E(Y|X)$.

Finally, the following example presents uncorrelated X, Y that are mutually predictable.

Example 3.12. For $0 < q < \frac{1}{10}$, define

$$p(-1,-1) = p(1,1) = p(0,1) = 2q, \ p(-1,1) = q, \ p(1,-1) = 3q, \ p(1,0) = 1 - 10q.$$

Then, since $E(XY) = 0 = E(Y)$, X and Y are uncorrelated, while, since

$$E(Y|X=0) = E(X|Y=0) = 1, \ E(Y|X=-1) = -\frac{1}{3}, \ E(X|Y=1) = \frac{1}{5},$$

X and Y are mutually predictable.

Example and Acknowledgment 3.13. The following example, which began this investigation, is due to Ronald Christensen. Let X be any positive random variable, and, for any positive x, $Y|X = x$ be normally distributed with mean zero and variance x. The random variables X and Y are then dependent but Y is unpredictable by X, since $E(Y|X = x) = 0$, for any positive x.

Although measuring X gives information about Y (its variance), it does not change our prediction, $E(Y|X)$, of Y.

Remark 3.14. The null hypothesis of Y being unpredictable by X, that is,

$$H_0 : E(Y|X) = E(Y),$$

including appropriate test statistics, is discussed in (Hart 1997); see, in particular, Section 9.2. Statistics for the analogous null hypotheses of X and Y being uncorrelated and X and Y being independent may be found in most elementary textbooks.

Bibliography

H. A. David (2009), "A historical note on zero correlation and independence," *The American Statistician* 63, 185–186.

A. De Paula (2008), "Conditional Moments and Independence," *The American Statistician* 62, 219–221.

R. deLaubenfels (2006), "The victory of least squares and orthogonality in statistics," *The American Statistician* 60, 315–321.

R. M. Dudley (2002), *Real Analysis and Probability,* Cambridge University Press.

G. G. Hamedani and H. W. Volkmer (2009), "Conditional Moments, Subindependence and Independence," Technical Report 478, Marquette University, Dept. of MSCS.

J. D. Hart (1997), *Nonparametric Smoothing and Lack-of-Fit Tests,* Springer Series in Statistics.

E. Schechtman and S. Yitzhaki (1987), "A measure of association based on Gini's mean difference," *Communications in Statistics, Theory and Methods, Series A* 16, 207–231.