§4. Generalization and further properties of Correlation

Up to now we have been concerned only with finite probability distributions, for which we have defined information and correlation. We shall now generalize the definition of correlation so as to be applicable to joint probability distributions over arbitrary sets of unrestricted cardinality.

We first consider the effects of refinement of a finite distribution. For example, we may discover that the event \( X \) is actually the disjunction of several exclusive events \( X_1, \ldots, X_n \), so that \( X \) occurs if any one of the \( X_i \) occurs, i.e., the single event \( X \) results from our failing to distinguish between the \( X_i \). The probability distribution which distinguishes between the \( X_i \) will be called a refinement of the distribution which does not. In general, we shall say that a distribution \( P' = P'\left(X_1', \ldots, X_i'ight) \) is a refinement of \( P = P\left(X_1, \ldots, X_i\right) \) if

\[
P(X_1, \ldots, X_i) = \sum_{X_i} P'(X_1', \ldots, X_i') \quad \text{(all } i \text{, } \ldots \text{)}
\]

We now state an important theorem concerning the behavior of correlation under a refinement of a joint probability distribution:

**Theorem 2:** \( P' \) is a refinement of \( P \) \( \Rightarrow \{X_1, \ldots, X_i\} \approx \{X_1', \ldots, X_i'\} \)

So that correlations never decrease upon refinement of a joint probability distribution. (Proof in Appendix)

As an example, suppose that we have a continuous probability density \( p(x, y) \). By division of the axes into a finite number of intervals, \( X_k \) and \( Y_j \), we arrive at
a finite joint distribution $P_{ij}$ by integration of $P(x, y)$ over the rectangle whose sides are the intervals $x_i$ and $y_j$. And which represents the probability that $x \in x_i$ and $y \in y_j$.

If we now subdivide the intervals, the new distribution $P'$ will be a refinement of $P$, and by theorem 2 the correlation $\rho_{ij}^2$ computed by $P'$ will never be less than that computed by $P$. Theorem 2 is seen to be simply the mathematical verification of the intuitive notion that closer analysis of a situation in which quantities $X$ and $Y$ are dependent can never lessen the knowledge about $Y$ which can be obtained from $X$.

This theorem allows us to give a definition of correlation which will apply to joint distributions over completely arbitrary sets, i.e., for any probability measure on an arbitrary product space in the following manner:

Assume that we have two arbitrary sets $X = \{x_1, x_2, \ldots, x_n\}$ and $Y = \{y_1, y_2, \ldots, y_m\}$, and a probability measure on the product, $P(X \times Y)$. Let $P'''$ be any finite partition of $X$ into subsets $X''_i$ and $Y$ into subsets $Y''_j$, such that the subsets $(X''_i, Y''_j)$ are measurable in the probability measure $P$. Another partition $P''$ is a refinement of $P'''$, $P'' \subset P'''$, if $P''$ results from $P'''$ by further subdivision of the subsets $X''_i$ and $Y''_j$. Each partition results in a finite probability distribution, for which the correlation $\rho_{ij}'$ is always defined. Furthermore, a refinement of a partition leads to a refinement of the probability distribution, so that by theorem 2:

\[ P'' \subset P''' \Rightarrow \{X, Y\}^P = \{X, Y\}^P'' \]
Ref: \( \lim_{x \to a} f(x) \) exists and is equal to \( a \) if for every \( \varepsilon > 0 \) there exists an \( \delta > 0 \) such that \( |f(x) - a| < \varepsilon \) for every \( x \in \mathbb{R} \) for which \( 0 < |x - a| < \delta \).
Now the set of all partitions is partially ordered under the refinement relation. Moreover, because for any pair of partitions $P, P'$ there is always a partition $P''$ which is a refinement of both (common lower bound), the set of all partitions forms a directed set. For functions $f$, on a directed set $\mathcal{D}$, one defines a directed set limit, $\lim f$, as:

$$\lim f = \alpha \iff \forall \varepsilon > 0 \exists \delta > 0 \text{ such that } \forall \theta \in \mathcal{D}, f(\theta) - \alpha < \varepsilon \text{ holds.}$$

It is easily seen from the directed set property of common lower bounds that if this limit exists it is necessarily unique.

By (4.8), the correlation $\{X, Y\}$ for each partition $P$ is a monotonic function on the directed set of all partitions. If the directed set limit always exists (it may be infinite, but is in any case well defined) which we shall take as the basic definition of the correlation $\{X, Y\}$.

We have succeeded in our endeavor to give a completely general definition of correlation, applicable to all distributions. It is a further consequence of (4.8) that this directed set limit is the supremum of $\{X, Y\}$, so that:

$$\lim \{X, Y\} = \sup \{X, Y\}$$

which we could equally well have taken as the definition.
There is a joint probability distribution over $X$ and $Y$ due to one over $U$. $U$ is jointly distributed.

$L = v$, $V = g(X)$

where $v = f(U)$.

$V = y(x)$

$X \sim Y$.

Add more variables.

$Y \sim X$.

$X \sim Y$.

$X \sim Y$.

$Y \sim X$.

$X \sim Y$.

$X \sim Y$.

$X \sim Y$.

$X \sim Y$.

$X \sim Y$.

$X \sim Y$.

$X \sim Y$.

$X \sim Y$.

$X \sim Y$.

$X \sim Y$.

$X \sim Y$.

$X \sim Y$.

$X \sim Y$.

$X \sim Y$.

$X \sim Y$.

$X \sim Y$.

$X \sim Y$.

$X \sim Y$.

$X \sim Y$.

$X \sim Y$.

$X \sim Y$.

$X \sim Y$.

$X \sim Y$.

$X \sim Y$.

$X \sim Y$.

$X \sim Y$.

$X \sim Y$.

$X \sim Y$.

$X \sim Y$.

$X \sim Y$.

$X \sim Y$.

$X \sim Y$.

$X \sim Y$.

$X \sim Y$.

$X \sim Y$.

$X \sim Y$.

$X \sim Y$.

$X \sim Y$.

$X \sim Y$.

$X \sim Y$.

$X \sim Y$.

$X \sim Y$.

$X \sim Y$.

$X \sim Y$.

$X \sim Y$.

$X \sim Y$.

$X \sim Y$.

$X \sim Y$.

$X \sim Y$.

$X \sim Y$.

$X \sim Y$.

$X \sim Y$.

$X \sim Y$.

$X \sim Y$.

$X \sim Y$.

$X \sim Y$.

$X \sim Y$.

$X \sim Y$.

$X \sim Y$.

$X \sim Y$.

$X \sim Y$.

$X \sim Y$.

$X \sim Y$.

$X \sim Y$.

$X \sim Y$.

$X \sim Y$.

$X \sim Y$.

$X \sim Y$.

$X \sim Y$.

$X \sim Y$.

$X \sim Y$.

$X \sim Y$.

$X \sim Y$.

$X \sim Y$.

$X \sim Y$.

$X \sim Y$.

$X \sim Y$.

$X \sim Y$.

$X \sim Y$.

$X \sim Y$.

$X \sim Y$.

$X \sim Y$.

$X \sim Y$.

$X \sim Y$.

$X \sim Y$.

$X \sim Y$.

$X \sim Y$.

$X \sim Y$.

$X \sim Y$.

$X \sim Y$.

$X \sim Y$.

$X \sim Y$.

$X \sim Y$.

$X \sim Y$.

$X \sim Y$.

$X \sim Y$.

$X \sim Y$.

$X \sim Y$.

$X \sim Y$.

$X \sim Y$.

$X \sim Y$.

$X \sim Y$.

$X \sim Y$.

$X \sim Y$.

$X \sim Y$.

$X \sim Y$.

$X \sim Y$.

$X \sim Y$.

$X \sim Y$.

$X \sim Y$.

$X \sim Y$.

$X \sim Y$.

$X \sim Y$.

$X \sim Y$.

$X \sim Y$.

$X \sim Y$.

$X \sim Y$.

$X \sim Y$.

$X \sim Y$.

$X \sim Y$.

$X \sim Y$.

$X \sim Y$.

$X \sim Y$.

$X \sim Y$.

$X \sim Y$.

$X \sim Y$.

$X \sim Y$.

$X \sim Y$.

$X \sim Y$.

$X \sim Y$.

$X \sim Y$.

$X \sim Y$.

$X \sim Y$.

$X \sim Y$.

$X \sim Y$.

$X \sim Y$.

$X \sim Y$.

$X \sim Y$.

$X \sim Y$.

$X \sim Y$.

$X \sim Y$.

$X \sim Y$.

$X \sim Y$.

$X \sim Y$.

$X \sim Y$.

$X \sim Y$.

$X \sim Y$.

$X \sim Y$.

$X \sim Y$.

$X \sim Y$.

$X \sim Y$.

$X \sim Y$.

$X \sim Y$.

$X \sim Y$.

$X \sim Y$.

$X \sim Y$.

$X \sim Y$.

$X \sim Y$.

$X \sim Y$.

$X \sim Y$.

$X \sim Y$.

$X \sim Y$.

$X \sim Y$.

$X \sim Y$.

$X \sim Y$.

$X \sim Y$.

$X \sim Y$.

$X \sim Y$.

$X \sim Y$.

$X \sim Y$.

$X \sim Y$.

$X \sim Y$.

$X \sim Y$.

$X \sim Y$.

$X \sim Y$.

$X \sim Y$.

$X \sim Y$.

$X \sim Y$.

$X \sim Y$.

$X \sim Y$.

$X \sim Y$.

$X \sim Y$.

$X \sim Y$.

$X \sim Y$.

$X \sim Y$.

$X \sim Y$.

$X \sim Y$.

$X \sim Y$.

$X \sim Y$.

$X \sim Y$.

$X \sim Y$.

$X \sim Y$.

$X \sim Y$.

$X \sim Y$.

$X \sim Y$.

$X \sim Y$.
We can now prove a very important theorem about correlation, which concerns its invariant nature. Let $X$ and $F$ be arbitrary sets with probability measure $M_p(X \times F)$ on their product. Let $f$ be any one to one mapping of $X$ onto a set $V_1$, $g$ a one to one mapping of $Y$ onto $V_2$. Then the probability measure induced on the product $V_1 \times V_2$ is simply the measure which assigns to each subset $V_1 \times V_2$ the measure of the image set, $M_p(V_1 \times V_2) = M_p (f^{-1}(V_1) \times g^{-1}(V_2)).$

Consider any partition $P$ of $X, Y$ into the subsets $B_1, B_2$ with probability distribution $P_{1,2}$. Then there is a corresponding partition $P'$ of $V_1 \times V_2$ into the sets $B_1 \times B_2$ where $V_1 = f(B_1)$ and $V_2 = g(B_2)$, the partition of the image sets of $P$, which has the probability distribution $P'_{1,2} = M_p(X_1 \times Y_2) = M_p(X_1 \times V_2) = P_{1,2}$, which is identical to the distribution for $P$ in $X \times Y$, so that:

$$\{X, Y\}^P = \{V_1, V_2\}^{P'}.$$
Due to the correspondence between the $P\subseteq$ and $Q\subseteq$ we have that

$\sup_{P} \{ x, y \} = \sup_{Q} \{ u, v \}$

and by virtue of (4.9) we have proved the theorem:

**Theorem 3:** $\{ X, Y \} = \{ U, V \}$ where $U$ is any one to one image of $X$, $V$ any one to one image of $Y$.

In other notation: $\{ X, Y \} = \{ f(x), g(Y) \}$ where $f$ and $g$ are arbitrary one to one functions.
This means changing variables to functionally related variables preserves the correlation. Again this is plausible on intuitive grounds, since a knowledge of \( f(x) \) is just as good as knowledge of \( x \), provided that \( f \) is one-to-one.

A special consequence of Theorem 3 is that for any continuous probability density \( P(x, y) \) over real numbers the correlation between \( f(x) \) and \( g(y) \) is the same as between \( x \) and \( y \), where \( f \) and \( g \) are any real valued one-to-one functions. As an example consider a probability distribution for the position of two particles, so that the random variables are the position coordinates. Theorem 3 then asserts that the position correlation is independent of the coordinate system, even if different coordinate systems are used for each particle! Also for a joint distribution, for a pair of events in space-time the correlation is invariant to arbitrary space-time coordinate transformations, again even allowing different transformations for the coordinates of each particle.

The general invariance of correlation expressed in Theorem 3 indicates the fundamental nature of this quantity for probability distributions. It is to be understood that all of the preceding results hold equally well for group correlations. Due to the fact that the correlation is defined as a limit for discrete distributions, all of the relations 3.7 to 3.15 which contain only correlation brackets remain true for arbitrary distributions. Only 3.11 and 3.12, which contain information terms, cannot be extended.
Assume that we have a collection of arbitrary sets \( X, Y, \ldots, Z \) and a probability measure \( M_p(X \times Y \times \cdots \times Z) \) on their cartesian product. Let \( P \) be any finite partition of \( X \) into subsets \( X_1, X_2, \ldots \), \( Y \) into subsets \( Y_1, Y_2, \ldots \), and \( Z \) into subsets \( Z_1, Z_2, \ldots \), such that the sets \( X_1 \times Y_1 \times \cdots \times Z_1 \) of the cartesian product are measurable in the probability measure \( M_p \).

Another partition \( P'' \) is a refinement of \( P \) if \( P'' \) results from \( P \) by further subdivision of the subsets \( X_1, Y_1, \ldots, Z_1 \). Each partition \( P'' \) results in a finite probability distribution for which the correlation \( \{X, Y, \ldots, Z\}^{P''} \) is always defined through (3.3). Furthermore, a refinement of a partition leads to a refinement of the probability distribution, so that by theorem 7:

\[
(4.8) \quad P \leq P'' \Rightarrow \{X, Y, \ldots, Z\}^{P'} \subseteq \{X, Y, \ldots, Z\}^{P''}
\]

For a function \( f \) on a directed set \( S \), one defines a directed set limit, \( \lim_{\rightarrow} f \), by:

Def.: \( \lim_{\rightarrow} f \) exists and is equal to \( \alpha \) if and only if for every \( \varepsilon > 0 \) there exists an \( x \in S \) such that \( |f(\beta) - \alpha| < \varepsilon \) for every \( \beta \in S \) for which \( \beta \leq x \).
By (4.8) the convolution \( \{X, Y, \ldots, Z\}^p \) is a monotone function on the directed set of all partitions. Consequently the directed set limit, which we shall take as the basic definition of the convolution \( \{X, Y, \ldots, Z\} \), always exists. (It may be infinite, but is in every case well-defined.) Thus:

\[
\text{Def: } \{X, Y, \ldots, Z\} = \lim \{X, Y, \ldots, Z\}^p
\]
Due to the fact that the correlation is defined as a limit for discrete distributions, Theorem 4 and all of the relations (3.1) to (3.15), which contain only correlation brackets, remain true for arbitrary distributions. Only 3.11 and 3.12, which contain information terms, cannot be extended yet.
We can now prove an important theorem about correlation which concerns its invariant nature. Let \( X, Y, Z \) be arbitrary sets with probability measure \( M_p \) on their cartesian product. Let \( f \) be any one-to-one mapping of \( X \) onto a set \( U \), \( g \) a one-one map of \( Y \) onto \( V \), \( \ldots \), and \( h \) a map of \( Z \) onto \( W \). Then a joint probability distribution over \( X \times Y \times \ldots \times Z \) leads also to one over \( U \times V \times \ldots \times W \) where the probability measure \( M_p' \) induced on the product \( U \times V \times \ldots \times W \) is simply the measure which assigns to each subset of \( U \times V \times \ldots \times W \) the measure which is the image set in \( X \times Y \times \ldots \times Z \) for the original measure \( M_p \).

(We have simply transformed to a new set of random variables: \( U = f(X), V = g(Y), \ldots, W = h(Z) \).)

Consider any partition \( P \) of \( X \times Y \times \ldots \times Z \) into the subsets \( \{X_1\}, \{Y_1\}, \ldots, \{Z_1\} \) with probability distribution

\[ P_{1,\ldots,\kappa} = M_p(X_1 \times Y_1 \times \ldots \times Z_1) \].

Then there is a corresponding partition \( P' \) of \( U \times V \times \ldots \times W \) into the image sets of the sets of \( P \), \( \{U_1\}, \{V_1\}, \ldots, \{W_1\} \) where \( U_1 = f(X_1), V_1 = g(Y_1), \ldots, W_1 = h(Z_1) \). But the probability distribution for \( P' \) is the same as that for \( P \) since

\[ P_{1,\ldots,\kappa} = M_p'(U_1 \times V_1 \times \ldots \times W_1) = M_p(X_1 \times Y_1 \times \ldots \times Z_1) = P_{1,\ldots,\kappa} \]

so that

\[ \{X, Y, \ldots, Z\}' = \{U, V, \ldots, W\} \].

(4.10)
Theorem 3: \[ \{x, y, \ldots, z\} = \{y, v, \ldots, w\} \]
where \( W \) is any one-one image of \( X \), \( U \) a one-one image of \( Y \), \( W \) a one-one image of \( Z \). In other notation:
\[ \{x, y, \ldots, z\} = \{f(x), f(y), \ldots, f(z)\} \] for all one-one functions \( f, g, \ldots, h \).

\( x', y', w' \) are the one images of \( x, y, z \) respectively.
These examples illustrate clearly the intrinsic nature of the correlation of various groups for joint probability distributions, which is implied by its invariance against arbitrary (one-one) transformations of the random variables. These correlation quantities are thus fundamental properties of probability distributions.

A correlation is an absolute rather than relative quantity, in the sense that the correlation between (numerical valued) random variables is completely independent of the scale of measurement chosen for the variables.

End of Section 4