Formal Epistemology Workshop
Carnegie-Mellon University
June 2, 2007

# REMARKS ON J. ZHANG'S PAPER, "Exchangeability and Invariance: A Causal Theory"

Carl Wagner
Department of Mathematics
The University of Tennessee
Knoxville, TN 37996-1300

## 1. Exchangeability

If $(\Omega, \mathbf{A}, P)$ is a probability space, a sequence of random variables $(X_1,\ldots X_n)$ defined on $\Omega$ is *exchangeable* iff, for all permutations $\pi$ of $\{1,\ldots,n\}$ and all real numbers $x_1,\ldots, x_n$,

$$(1.1)\quad P(X_1 \leq x_1, \ldots , X_n \leq x_n) = P(X_{\pi(1)} \leq x_1, \ldots , X_{\pi(n)} \leq x_n).$$

An equivalent, sometimes more salient, formulation of exchangeability permutes the subscripts on the values $x_i$ :

$$(1.2)\quad P(X_1 \leq x_1, \ldots , X_n \leq x_n) = P(X_1 \leq x_{\pi(1)}, \ldots , X_n \leq x_{\pi(n)}).$$

Remark 1.1. If the $X_i$ are discrete random variables, then the inequalities $\leq$ may be replaced by equalities in (1.1) and (1.2).

An infinite sequence $(X_i)$ of random variables is said to be exchangeable iff each finite initial segment $(X_1,\ldots X_n)$ of $(X_i)$ is exchangeable, or, equivalently, iff every finite subsequence of $(X_i)$ is exchangeable.

THEOREM 1.1. The random variables in any finite (hence, any infinite) exchangeable sequence are identically distributed.

THEOREM 1.2. If the sequence $(X_i)$ is exchangeable, then it is *stationary*, i.e., for all positive integers n and k,

$$(1.3) \qquad (X_1,\ldots X_n) \quad =_d \quad (X_{1+k},\ldots X_{n+k}),$$

where the symbol $=_d$ denotes identical joint distribution.

THEOREM 1.3. If the sequence $(X_i)$ is exchangeable, then it is *spreadable*, i.e., for all positive integers n and $i_1 < \ldots < i_n$ ,

$$(1.4) \qquad (X_1,\ldots X_n) \quad =_d \quad (X_{i1},\ldots X_{in}) \quad .$$

All of the preceding theorems may easily be verified using appropriate marginalizations.

Remark 1.2. It is a remarkable, if not immediately intuitive, fact that exchangeability is equivalent to spreadability (de Finetti, Ryll-Nardzewski). See: Olaf Kallenberg, Foundations of Modern Probability, 2d.ed., Springer, Theorem 11.10, on p. 212.

Remark 1.3. Since many interesting results about finite exchangeable sequences rely on these sequences being extendable to infinite, or at least much longer finite, exchangeable sequences, it may be worth noting that there are finite exchangeable sequences that can not be extended by even a single additional random variable. For example, if X is a Bernoulli random variable with $P(X=1) = P(X=0) = ½$, and $Y = 1 – X$, then there exists no random variable Z such that the sequence (X,Y,Z) is exchangeable[exercise].

## 2. Exchangeable 0,1 -Valued Random Variables

THEOREM 2.1. (De Finetti). If $(X_i)$ is an infinite sequence of exchangeable, 0,1-valued random variables, then there exists a random variable $\Theta : \Omega \rightarrow [0,1]$ such that

(2.1) $(X_i) \mid \Theta = \theta$ are independent, identically distributed Bernoulli random variables, with $P(X_i = 1) = \theta$,

whence, for all positive integers n, and all $x_1, \ldots, x_n$ in $\{0,1\}$,

(2.2) $P(X_1 = x_1, \ldots, X_n = x_n \mid \Theta = \theta) = \theta^{\Sigma x_i} (1-\theta)^{n - \Sigma x_i}$ ,

and, with $F_\Theta(\theta) := P(\Theta \le \theta)$ ,

(2.3) $P(X_1 = x_1, \ldots, X_n = x_n) = \int_0^1 \theta^{\Sigma x_i} (1-\theta)^{n - \Sigma x_i} \, dF_\Theta(\theta)$.

COROLLARY 2.1. With $X_i$ and $\Theta$ as above,

(2.4) $E[X_i] = E[\Theta]$ for all i,

(2.5) $E[X_i X_j] = E[\Theta^2]$ for all i, j , and so

(2.6) $\mathrm{Cov}[X_i, X_j] := E[X_i X_j] - E[X_i] E[X_j] =$

$$E[\Theta^2] - E[\Theta]^2 = \mathrm{Var}[\Theta].$$

Hence, if $\Theta$ is nondeterministic, then $\mathrm{Cov}[X_i, X_j] > 0$, i.e., the sequence $(X_i)$ is “positively correlated.”

Remark 2.1. The proofs of (2.4) - (2.6) follow straightforwardly from the “Double Expectation Theorem” :

(2.7) $E[X] = E[\,E[X|Y]\,] = \int_{-\infty}^{\infty} E[X|Y=y] \, dF_Y(y)$, where $F_Y(y) := P(Y \le y)$. See Sheldon Ross, A First Course in Probability, 7th edition,Prentice-Hall, 2006.

Corollary 2.1 already hints at the connection between earlier and later values of the variables $X_i$. The next corollary elaborates on that connection, and indicates why, for n sufficiently large, upon observing the values $X_1 = x_1, \ldots, X_n = x_n$, it is reasonable to estimate $P(X_{n+1} = 1)$ by the quantity $(x_1 + \ldots + x_n)/n$.

COROLLARY 2.2. With $X_i$ and $\Theta$ as above,

(2.8) $\lim_{n\to\infty} \frac{1}{n}\sum_1^n X_i = \Theta$ almost surely , i.e.,

(2.9) $P(\{ \lim_{n\to\infty} \frac{1}{n}\sum_1^n X_i = \Theta \}) = 1$, where

$$\{ \lim_{n\to\infty} \frac{1}{n}\sum_1^n X_i = \Theta \} = \{\omega\in\Omega: \lim_{n\to\infty} \frac{1}{n}\sum_1^n X_i(\omega) = \Theta(\omega) \}$$

*Proof.* Define a 0,1-valued random variable L by:

(2.10) $L(\omega) = 1$ iff $\lim_{n\to\infty} \frac{1}{n}\sum_1^n X_i(\omega) = \Theta(\omega)$ .

As with any 0,1-valued random variable, $E[L] = P(L=1)$. Hence, (2.9) holds iff $P(L=1) =1$ iff $E[L] =1$. By the double expectation theorem,

$$E[L] = E[\, E[\, L|\Theta \,] \,] = \int_0^1 E[\, L \mid \Theta=\theta \,]\, dF_\Theta(\theta)$$

$$= \int_0^1 P(\, L = 1 \mid \Theta=\theta)\, dF_\Theta(\theta)$$

$$= \int_0^1 P(\lim_{n\to\infty} \frac{1}{n} \sum_1^n Y_i = \theta)\, dF_\Theta(\theta),$$

where $Y_i := X_i \mid \Theta=\theta$ . By De Finetti's Theorem, the $Y_i$ are i.i.d. Bernoulli ($\theta$), and so, by the strong law of large numbers, the preceding integral is equal to

$$\int_0^1 1\, dF_\Theta(\theta) = 1. \quad \square$$