ARTICLE IN PRESS

Available online at www.sciencedirect.com

ScienceDirect

Stochastic Processes and their Applications ▮ (▮▮▮▮) ▮▮▮–▮▮▮

stochastic processes and their applications

www.elsevier.com/locate/spa

# On a notion of partially conditionally identically distributed sequences

Sandra Fortini*, Sonia Petrone, Polina Sporysheva

*Bocconi University, Milan, Italy*

Received 30 July 2016; received in revised form 6 March 2017; accepted 11 June 2017
Available online xxxx

**Abstract**

A notion of conditionally identically distributed (c.i.d.) sequences has been studied as a form of stochastic dependence weaker than exchangeability, but equivalent to it in the presence of stationarity. We extend such notion to families of sequences. Paralleling the extension from exchangeability to partial exchangeability in the sense of de Finetti, we propose a notion of *partially c.i.d.* dependence, which is shown to be equivalent to partial exchangeability for stationary processes. Partially c.i.d. families of sequences preserve attractive limit properties of partial exchangeability, and are asymptotically partially exchangeable. Moreover, we provide strong laws of large numbers and two central limit theorems. Our focus is on the asymptotic agreement of predictions and empirical means, which lies at the foundations of Bayesian statistics. Natural examples of partially c.i.d. constructions are interacting randomly reinforced processes satisfying certain conditions on the reinforcement.


*Keywords:* Exchangeability; Partial exchangeability; Reinforced processes; Spreadability; Limit theorems; Prediction; Bayesian nonparametrics

## 1. Introduction

Exchangeability is a central notion in many areas of probability and related fields; we refer to Kingman [34], Aldous [2], Austin [8], Kallenberg [33] and Aldous [3] for classical, wide references. In Bayesian statistics, exchangeability is the fundamental probabilistic structure at the basis of learning, expressing the subjective probabilistic description of repeated experiments

* Correspondence to: Department of Decision Sciences, Bocconi University, Via Roentgen 1, 20136 Milano, Italy
*E-mail address:* sandra.fortini@unibocconi.it (S. Fortini).

http://dx.doi.org/10.1016/j.spa.2017.06.008
0304-4149/

under similar conditions. Exchangeable sequences are conditionally independent and identically distributed (i.i.d.).

However, forms of competition, selection, and other sources of non stationarity, may break exchangeability, although the system may converge, asymptotically, to an exchangeable steady state. Thus, weaker notions of stochastic dependence, that do not assume stationarity, yet preserve some main asymptotic properties of exchangeable processes, are of theoretical and applied interest. Based on results by Kallenberg [32], Berti et al. [14] introduce a notion of *conditionally identically distributed* (c.i.d.) sequences, as a form of stochastic dependence weaker than exchangeability but equivalent to it for stationary sequences. Roughly speaking, a sequence of random variables $(X_n)_{n\geq 1}$ is c.i.d. if, for every $n \geq 0$, $X_{n+1}, X_{n+2}, \ldots$ are conditionally identically distributed, given the past $X_1, \ldots, X_n$ (for $n = 0$ the property reads $X_1, X_2, \ldots$ are identically distributed). A precise definition is given in Section 2. These processes are the starting point of our study.

Notions of partial exchangeability are needed for more complex phenomena, which can still be described by some form of probabilistic invariance, under specific subclasses of permutations. See [20], [33] and [3]. A basic notion is partial exchangeability in the sense of de Finetti [23] (called internal exchangeability for a family of sequences by Aldous [2]). In Bayesian statistics, partial exchangeability in the sense of de Finetti is the fundamental probabilistic dependence behind inference for multiple experiments. Roughly speaking, observations are exchangeable within each experiment, but not across experiments; the probabilistic dependence among sequences allows borrowing strength across experiments. Throughout this paper, by partial exchangeability we mean partial exchangeability in the sense of de Finetti. Again, different forms of non-stationarity may break the symmetry of partial exchangeability. It seems natural to ask how the notion of c.i.d. sequences can be extended to a notion of *partially* c.i.d. processes, the way that partial exchangeability extends exchangeability. Such extension is the main objective of the present work.

We introduce a notion of *partially c.i.d.* families of sequences that is shown to be equivalent to partial exchangeability under stationarity. Then, we prove that partially c.i.d. sequences preserve some main limit properties of partially exchangeable sequences. In particular, the joint predictive distributions and the joint empirical distributions converge (weakly) to the same random limit, almost surely. Moreover, partially c.i.d. sequences are asymptotically partially exchangeable. The asymptotic agreement of frequencies and predictions is of fundamental interest in Bayesian statistics, where probability has a subjective interpretation, showing the frequentist basis of the subject's probabilistic learning. Such agreement is ensured (in the subject's opinion) for exchangeable and partially exchangeable sequences. Our result shows that it is still ensured when relaxing the assumption of stationarity from partial exchangeability. Marginally, these results are not surprising, as partially c.i.d. sequences are marginally c.i.d. and the limit behavior of c.i.d. sequences has been studied [14,15]. Yet, for multiple sequences, the *joint* limit behavior is not obvious, as the sequences are stochastically dependent. Notice that they remain asymptotically dependent, if the random marginal limit measures are dependent.

These limit results are refined in Section 5, where we provide a strong law of large numbers for partially c.i.d. sequences, and in Section 6, where we give two central limit theorems, for the scaled cumulative forecast errors and the scaled difference between empirical means and predictions, respectively. Beyond fundamental issues, the possibility of approximating predictions by empirical means, with an approximation error given by a central limit theorem, can be of interest for hypothesis testing and model checking as well as for facilitating computations in Bayesian prediction with large sample size.

Please cite this article in press as: S. Fortini, et al., On a notion of partially conditionally identically distributed sequences, Stochastic Processes and their Applications (2017), http://dx.doi.org/10.1016/j.spa.2017.06.008.