Successioni parzialmente c.i.d. di variabili aleatorie

Sonia Petrone

con **Sandra Fortini** e **Polina Sporysheva**

Università Bocconi, Milano

Alcuni argomenti di Probabilità, Statistica e Teoria della Misura
Convegno in onore della Prof. ssa Patrizia Berti
Modena, 8-9 giugno 2015

SCAMBIABILITÀ , TEOREMI LIMITE, E SUCCESSIONI C.I.D.

Patrizia ha dato contributi importanti sui fondamenti della probabilità , nell'impostazione soggettiva e predittiva; sulla scambiabilità , e su teoremi limite per leggi scambiabili e per successioni con dipendenza stocastica.

Questo lavoro è basato sui risultati in
P. Berti, L. Pratelli, P. Rigo (2004). Limit theorems for a class of identically distributed random variables. *Annals of Probability*, **32**, 2029–2052.

e su diversi loro lavori successivi.

CONTENTS

Exchangeability is important in many areas of probability & related fields (KINGMAN (1978); ALDOUS (1985; 2010) and has a fundamental role in Bayesian Statistics.

(X_n) exchangeable $\Rightarrow$ (X_n) is stationary.
Stationarity is a key assumption.

CONTENTS

Exchangeability is important in many areas of probability & related fields (KINGMAN (1978); ALDOUS (1985; 2010) and has a fundamental role in Bayesian Statistics.

(X_n) exchangeable $\Rightarrow$ (X_n) is stationary.
Stationarity is a key assumption.

But, in many problems, stationarity is restrictive.

BERTI, PRATELLI, RIGO (2004) give a notion of **conditionally identically distributed sequences** (**c.i.d.**) that, roughly speaking, corresponds to exchangeability without the assumption of stationarity.

$\rightarrow$ We give a parallel notion of **partially c.i.d.** sequences, that corresponds to partial exchangeability under the assumption of stationarity.

OUTLINE

- motivations
- removing stationarity assumption:
  – c.i.d. and partially c.i.d. sequences.
  – weaker assumptions.
- properties and limit theorems
- examples.

Exchangeability

The sequence (X_n) is exchangeable if

$$(X_1, \ldots, X_n) \stackrel{d}{=} (X_{\pi(1)}, \ldots, X_{\pi(n)}),$$

for any $n \geq 1$ and any finite permutation π of $(1, \ldots, n)$.

Order in which data are recorded is irrelevant.

By the representation theorem for exchangeable sequences:

$$\mathbf{P}(X_1 \in A_1, \ldots, X_n \in A_n) = \int \prod_{i=1}^{n} P(A_i)\, d\mathbf{P}(P),$$

where P is the weak limit of the sequence of the empirical, and predictive, distributions.
Thus, $X_i \mid P \stackrel{i.i.d}{\sim} P$, and P has a distribution induced by $\mathbf{P}$. The random measure P is the *directing measure*

PARTIAL EXCHANGEABILITY (DE FINETTI)

Definition. *The sequence* (X_n, Y_n) *is partially exchangeable (in the sense of de Finetti) if*

$$(X_1, \ldots, X_n, Y_1, \ldots, Y_m) \stackrel{d}{=} (X_{\sigma(1)}, \ldots, X_{\sigma(n)}, Y_{\tau(1)}, \ldots, Y_{\tau(m)}),$$

for any $n, m \geq 1$ *and any permutations* σ *of* $(1, \ldots, n)$ *and* τ *of* $(1, \ldots, m)$.

Order inside groups is irrelevant, but observations cannot be permuted across groups.

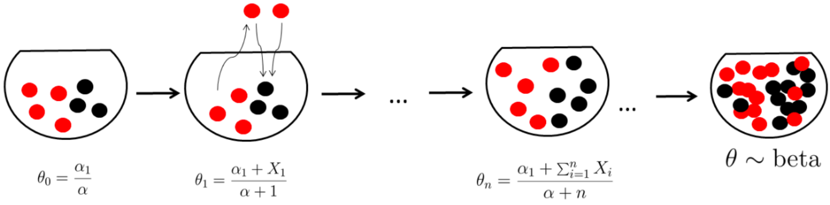
By the representation theorem for partially exchangeable sequences,

$$\mathbf{P}(X_1 \in A_1, \ldots, X_n \in A_n, Y_1 \in B_1, \ldots, Y_m \in B_m) = \int \prod_{i=1}^{n} P_x(A_i) \prod_{j=1}^{m} P_y(B_j) d\mathbf{P}(P_x, P_y),$$

where (P_x, P_y) is the weak limit of the sequence of the empirical distributions $(\hat{F}_{n,x}, \hat{F}_n, y)$.

MOTIVATING EXAMPLE: EVOLUTIONARY PHENOMENA

Two color Polya urn. (spread of contagion):



evolutionary phenomena, described by the generative rule:

$$\begin{aligned} X_1 &\sim Bernoulli(\theta_0), \quad \theta_0 = \frac{\alpha_1}{\alpha} \\ X_{n+1} \mid X_{1:n} &\sim Bernoulli(\theta_n), \quad \theta_n = \frac{\alpha_1 + \sum_{i=1}^n X_i}{\alpha + n}, \quad n \geq 1. \end{aligned}$$

(X_n) is exchangeable.

By the representation theorem, although (X_n) describes an ***evolutionary*** phenomena, it is probabilistically equivalent to **static** sampling:

- pick $p \sim Beta(\alpha_1, \alpha - \alpha_1)$;
- then sample i.i.d. from the urn with composition p.

TWO PROPERTIES

– The urn composition (θ_n), that is also the predictive probability of white ball, is a martingale

$$\mathbb{E}(\theta_{n+1} \mid X_1, \ldots, X_n) = \theta_n$$

– The predictive distribution is a symmetric function of $X_1, \ldots, X_n$.

- How does the contagion spread? asymptotic urn composition?
  The martingale property gives

  $$\theta_n \to \theta; \quad \theta \text{ random},$$

  and one can show that $\theta \sim Beta(\alpha_1, \alpha - \alpha_1)$.
- How about the relative frequency $\bar{X}_n$?

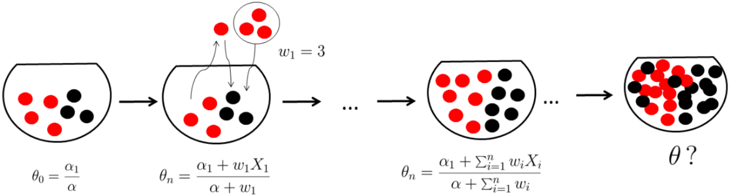
$$\bar{X}_n \to \theta, \text{ and CLT: } \quad \bar{X}_n \approx \int N(\theta, \frac{\theta(1-\theta)}{n})\, d\mathbf{P}(\theta).$$

Convergence of the predictive probabilities implies (Aldous, 1985) that (X_n) is asymptotically exchangeable.
Moreover, (X_n) is stationary. Thus, it is exchangeable.

RANDOMLY REINFORCED URNS (RRU)

What if the reinforcement is **random**?

$$\theta_n = P(X_{n+1} = 1 \mid X_{1:n}, W_{1:n}) = \frac{\alpha_1 + \sum_{i=1}^n W_i X_i}{\alpha + \sum_{i=1}^n W_i}$$



In general, the process (X_n) is no longer stationary $\to$ no longer exchangeable.
Can we still establish the asymptotic behavior of the urn composition (predictive probability)?

Athreya (1969) and Athreya, Ney (1972); Pemantle (1989); Berti, Pratelli, Rigo (2004; 2011)

EXAMPLE 2: BAYESIAN STATISTICS

Subjective probability. Exchangeability is subjective, order does not matter.
(P_n) predictive rule: learning from experience.

What if the order of past observations is relevant?

- Can we still say that $P_n(\cdot) \approx \hat{F}_n(\cdot)$, for large n?
- $\rightarrow$ *Bayesian consistency*: $d(P_n, \hat{F}_n)$
  (DIACONIS+FREEDMAN (1993), BERTI+RIGO (1997),..)
- $\rightarrow$ Computations for large n? Can we approximate P_n by an estimate, here $\hat{F}_n$, and how about the approximation error?
  - CLT for $\sqrt{n}(P_n(A) - \hat{F}_n(A))$;
  - uniform limits: limit law of $\sqrt{n}(P_n(\cdot) - \hat{F}_n(\cdot))$

EXAMPLE 2: BAYESIAN STATISTICS

Subjective probability. Exchangeability is subjective, order does not matter.
(P_n) predictive rule: learning from experience.

What if the order of past observations is relevant?

- Can we still say that $P_n(\cdot) \approx \hat{F}_n(\cdot)$, for large n?
- $\rightarrow$ *Bayesian consistency*: $d(P_n, \hat{F}_n)$
  (DIACONIS+FREEDMAN (1993), BERTI+RIGO (1997),..)
- $\rightarrow$ Computations for large n? Can we approximate P_n by an estimate, here $\hat{F}_n$, and how about the approximation error?
  - CLT for $\sqrt{n}(P_n(A) - \hat{F}_n(A))$;
  - uniform limits: limit law of $\sqrt{n}(P_n(\cdot) - \hat{F}_n(\cdot))$

The predictive distribution for dependent random probability measures is analytically complicated: can we find an approximation, for partially exchangeable sequences? would it hold, even if not stationary?

REMOVING THE STATIONARITY ASSUMPTION

Theorem *(Kallenberg, 1978). A stationary sequence that satisfies*

$$(X_1, \ldots, X_n, X_{n+1}) \stackrel{d}{=} (X_1, \ldots, X_n, X_{n+k})$$

for any n and $k \geq 1$, is exchangeable.

The above condition implies

$$X_{n+1} \mid X_1, \ldots .X_n \stackrel{d}{=} X_{n+k} \mid X_1, \ldots, X_n$$

future observations are conditionally identically distributed (**c.i.d.**).

Thus,

c.i.d. + stationarity = exchangeability.

Conditionally identically distributed (c.i.d.) sequences

Berti, Pratelli, Rigo (*Ann. Probab. 2004*) give a general notion of c.i.d. sequences with respect to a filtration $\mathcal{G} = (\mathcal{G}_n)$. Let $X_n \in \mathcal{X}$, a Polish space.

Definition. *The sequence* (X_n) *is* **$\mathcal{G}$-c.i.d.** *if*

$$\mathbf{P}(X_{n+k} \in \cdot \mid \mathcal{G}_n) = \mathbf{P}(X_{n+1} \in \cdot \mid \mathcal{G}_n),$$

for every $n \geq 0$ *and* $k \geq 1$.

If $\mathcal{G} = \sigma(X_1, \ldots, X_n)$ is the natural filtration, we say that (X_n) is c.i.d..
If (X_n) is $\mathcal{G}$-c.i.d., it is also c.i.d..

Equivalently, (X_n) is $\mathcal{G}$-c.i.d. if the predictive measure $P(X_{n+1} \in \cdot \mid \mathcal{G}_n)$ is a $\mathcal{G}$-martingale:

$$(Z_{n,f} \equiv \mathbb{E}(f(X_{n+1}) \mid \mathcal{G}_n), n \geq 0) \quad \text{is a } \mathcal{G}\text{-martingale, for every integr. meas. } f.$$

LIMIT THEOREMS FOR C.I.D. SEQUENCES: PREDICTIVE DIST.

Let (X_n) be a $\mathcal{G}$-c.i.d. sequence, and $Z_{n,f} = \mathbb{E}(f(X_{n+1} \mid \mathcal{G}_n), f$ measurable, with $\mathbb{E}|f(X_n)| < \infty$.
For $f(X) = X$, $Z_{n,f} = \mathbb{E}(X_{n+1} \mid \mathcal{G}_n)$, point prediction.
For $f(X) = I_A(X)$, $Z_{n,f} = \mathbf{P}(X_{n+1} \in A \mid \mathcal{G}_n)$, predictive probability of A.

- By definition of c.i.d., $(Z_{n,f})$ is a $\mathcal{G}$-martingale. Therefore,

$$Z_{n,f} \to Z_f$$

  and $Z_{n,f} = E(Z_f \mid \mathcal{G}_n)$.

- BERTI, PRATELLI, RIGO (2004) show that $Z_{n,f} \to Z_f$ **stably** (Renyi (1963), Aldous (1985)).
  Thus, there exists a measure P such that $Z_f = E_P(f(X))$; in other words, a.s.-$\mathbf{P}$, **the sequence of predictive distributions P_n converges weakly to a random measure P**.

- Convergence of the predictive distributions implies (Aldous, 1985) that the sequence (X_n) is **asymptotically exchangeable**:

$$(X_{n+1}, X_{n+2}, \ldots) \overset{d}{\approx} (S_1, S_2, \ldots) \quad \text{for large } n$$

  where $(S_1, S_2, \ldots)$ is an exchangeable sequence with directing measure P.

LIMIT THEOREMS FOR C.I.D. SEQUENCES: PREDICTIVE DIST.

Let (X_n) be a $\mathcal{G}$-c.i.d. sequence, and $Z_{n,f} = \mathbb{E}(f(X_{n+1} \mid \mathcal{G}_n), f$ measurable, with $\mathbb{E}|f(X_n)| < \infty$.
For $f(X) = X$, $Z_{n,f} = \mathbb{E}(X_{n+1} \mid \mathcal{G}_n)$, point prediction.
For $f(X) = I_A(X)$, $Z_{n,f} = \mathbf{P}(X_{n+1} \in A \mid \mathcal{G}_n)$, predictive probability of A.

- By definition of c.i.d., $(Z_{n,f})$ is a $\mathcal{G}$-martingale. Therefore,

$$Z_{n,f} \to Z_f$$

  and $Z_{n,f} = E(Z_f \mid \mathcal{G}_n)$.

- BERTI, PRATELLI, RIGO (2004) show that $Z_{n,f} \to Z_f$ **stably** (Renyi (1963), Aldous (1985)).
  Thus, there exists a measure P such that $Z_f = E_P(f(X))$; in other words, a.s.-$\mathbf{P}$, **the sequence of predictive distributions P_n converges weakly to a random measure P**.

- Convergence of the predictive distributions implies (Aldous, 1985) that the sequence (X_n) is **asymptotically exchangeable**:

$$(X_{n+1}, X_{n+2}, \dots) \overset{d}{\approx} (S_1, S_2, \dots) \quad \text{for large } n$$

  where $(S_1, S_2, \dots)$ is an exchangeable sequence with directing measure P.

It follows easily that, if (X_n) is stationary and asymptotically exchangeable, then it is exchangeable.

CENTRAL LIMIT THEOREMS

Theorem. If (X_n) is $\mathcal{G}$-c.i.d., the sequence of empirical distributions converges to the same limit P

$$\frac{\sum_{i=1}^n \delta_{X_i}}{n} \Rightarrow P, \quad a.s. - \mathbf{P}.$$

Berti, Pratelli, Rigo (2004) also provide central limit theorems for

$$\sqrt{n}\left(\frac{\sum_{i=1}^n f(X_i)}{n} - L_n\right),$$

for different choices of the random centering L_n. In particular: $L_n = \mathbb{E}_P(f(X))$ and $L_n = \mathbb{E}(f(X) \mid \mathcal{G}_n)$. For $f(X) = I_A(X)$, $L_n = P(A)$ and $L_n = \mathbf{P}(X_{n+1} \in A \mid \mathcal{G}_n)$.

> **Theorem**. *Let (X_n) be a $\mathcal{G}$-c.i.d. sequence. Then, under conditions,* $\sqrt{n}\,[\frac{\sum_{i=1}^n \delta_{X_i}(A)}{n} - \mathbf{P}(X_{n+1} \in A \mid \mathcal{G}_n)]$ *converges stably to* $N(0, P(A)(1-P(A)))$.

This implies

$$\frac{\sum_{i=1}^n \delta_{X_i}(A)}{n} - \mathbf{P}(X_{n+1} \in A \mid \mathcal{G}_n) \approx \int N(0, \frac{P(A)(1-P(A))}{n}) d\mathbf{P}(P), \quad \text{for } n \text{ large.}$$

Roughly speaking, the required conditions control the convergence rates of the predictive distributions (see Berti *et al.*, 2011).

Remark. In some cases, weaker assumptions are sufficient for the asymptotic results and CLTs.
In particular, if $X_i \in \{0, 1\}$, central limit theorems can be obtained under the weaker assumption that $\theta_n = P(X_{n+1} = 1 \mid \mathcal{G}_n)$ is a uniformly integrable quasi-martingale (which holds for c.i.d. sequences). See Aletti, May, Secchi (2009); Berti, Crimaldi, Pratelli, Rigo (2011).

This is of interest for applications in randomly reinforced urn processes and clinical trials.

The previous CLTs refer to a fixed function f. Stronger results regard

$$\sqrt{n}\Big(\frac{\sum_{i=1}^{n} f(X_i)}{n} - L_n\Big)$$

as a **process**, and are provided by Berti, Pratelli, Rigo (2004).
The c.i.d. condition is a basic assumption for **uniform limits theorems** for predictive inference.

PARTIAL EXCHANGEABILITY AND PARTIAL C.I.D.

Theorem *(Kallenberg, 1978). (X_n) is c.i.d. and stationary iff (X_n) exchangeable.*

PARTIAL EXCHANGEABILITY AND PARTIAL C.I.D.

Theorem *(Kallenberg, 1978). (X_n) is c.i.d. and stationary iff (X_n) exchangeable.*

For multiple experiments (m evolutionary processes; or, sampling from m populations,...):
interest for a notion of partially c.i.d. sequences $(\,(X_n, Y_n)\,)$ such that

Theorem. *(X_n, Y_n) is partially c.i.d. and stationary iff (X_n, Y_n) is partially exchangeable.*

PARTIALLY C.I.D. SEQUENCES

Definition. *A sequence* $((X_n, Y_n), n \geq 1)$ *is partially* $\mathcal{G}$*-c.i.d. if, for every* $n \geq 0$*,* $k \geq 1$*:*

$$\begin{aligned} \mathbf{P}(X_{n+k} \in \cdot \mid \mathcal{G}_n, Y_{n+1}) &= \mathbf{P}(X_{n+1} \in \cdot \mid \mathcal{G}_n, Y_{n+1}) \\ \mathbf{P}(Y_{n+k} \in \cdot \mid \mathcal{G}_n, X_{n+1}) &= \mathbf{P}(Y_{n+1} \in \cdot \mid \mathcal{G}_n, X_{n+1}) \end{aligned}$$

That is, (X_n) is c.i.d. with respect to the filtration $(\mathcal{G}_n \vee Y_{n+1})$, and (Y_n) is c.i.d. with respect to the filtration $(\mathcal{G}_n \vee X_{n+1})$.

If $\mathcal{G}_n = \sigma(X_1, \ldots, X_n, Y_1, \ldots, Y_n)$, we simply say that $(\,(X_n, Y_n)\,)$ is partially c.i.d..

PARTIALLY C.I.D. SEQUENCES

Definition. *A sequence* $((X_n, Y_n), n \geq 1)$ *is partially* $\mathcal{G}$*-c.i.d. if, for every* $n \geq 0$, $k \geq 1$*:*

$$\begin{aligned} \mathbf{P}(X_{n+k} \in \cdot \mid \mathcal{G}_n, Y_{n+1}) &= \mathbf{P}(X_{n+1} \in \cdot \mid \mathcal{G}_n, Y_{n+1}) \\ \mathbf{P}(Y_{n+k} \in \cdot \mid \mathcal{G}_n, X_{n+1}) &= \mathbf{P}(Y_{n+1} \in \cdot \mid \mathcal{G}_n, X_{n+1}) \end{aligned}$$

That is, (X_n) is c.i.d. with respect to the filtration $(\mathcal{G}_n \vee Y_{n+1})$, and (Y_n) is c.i.d. with respect to the filtration $(\mathcal{G}_n \vee X_{n+1})$.

If $\mathcal{G}_n = \sigma(X_1, \ldots, X_n, Y_1, \ldots, Y_n)$, we simply say that $(\,(X_n, Y_n)\,)$ is partially c.i.d..

Clearly, a simple sufficient condition for two $\mathcal{G}$-c.i.d. sequences (X_n) and (Y_n) to be partially $\mathcal{G}$-c.i.d. is that

$$\begin{aligned} (X_{n+1}, X_{n+2}) &\perp\!\!\!\perp Y_{n+1} \mid \mathcal{G}_n \\ (Y_{n+1}, Y_{n+2}) &\perp\!\!\!\perp X_{n+1} \mid \mathcal{G}_n \end{aligned}$$

LIMIT THEOREMS: PREDICTIVE DISTRIBUTIONS

Marginal limit theorems follow from the marginal c.i.d. property.

- Both $\mathbf{P}(X_{n+1} \in \cdot \mid \mathcal{G}_n)$ and $\mathbf{P}(X_{n+1} \in \cdot \mid X_1, \dots, X_n)$ converge weakly to a random measure P_x, a.s., and (X_n) is asymptotically exchangeable.
- Both $\mathbf{P}(Y_{n+1} \in \cdot \mid \mathcal{G}_n)$ and $\mathbf{P}(Y_{n+1} \in \cdot \mid Y_1, \dots, Y_n)$ converge weakly to a random measure P_y, a.s., and (Y_n) is asymptotically exchangeable.

LIMIT THEOREMS: PREDICTIVE DISTRIBUTIONS

Marginal limit theorems follow from the marginal c.i.d. property.

- Both $\mathbf{P}(X_{n+1} \in \cdot \mid \mathcal{G}_n)$ and $\mathbf{P}(X_{n+1} \in \cdot \mid X_1, \ldots, X_n)$ converge weakly to a random measure P_x, a.s., and (X_n) is asymptotically exchangeable.
- Both $\mathbf{P}(Y_{n+1} \in \cdot \mid \mathcal{G}_n)$ and $\mathbf{P}(Y_{n+1} \in \cdot \mid Y_1, \ldots, Y_n)$ converge weakly to a random measure P_y, a.s., and (Y_n) is asymptotically exchangeable.

BUT we need the joint distribution, and need to prove asymptotic conditional independence of (X_n) and (Y_n).

Theorem. *If $((X_n, Y_n))$ is $\mathcal{G}$-c.i.d.,*

$$\mathbb{E}(f(X_{n+1})g(Y_{m+1}) \mid \mathcal{G}_n) \overset{stably}{\rightarrow} \mathbb{E}_{P_x}(f(X))\, \mathbb{E}_{P_y}(g(Y)),$$

for all f, g continuous and bounded.
This implies that there exist a representing measure P such that

$$\mathbb{E}(f(X_{n+1})g(Y_{m+1}) \mid \mathcal{G}_n) \rightarrow \mathbb{E}_P(f(X)g(Y))$$

and such P is equal to $P_x\ P_y$.

- Thus, the sequence of joint predictive distribution : $P_n = \mathbf{P}((X_{n+1}, Y_{n+1}) \in \cdot \mid \mathcal{G}_n)$ converges weakly to $P = (P_x \times P_y)$, $\mathbf{P}$-a.s.
- It follows that $(X_n, Y_n))$ is asymptotically exchangeable, with directing measure $P = P_x \times P_y$. Therefore, $(\,(X_n, Y_m)\,)$ is asymptotically partially exchangeable, with directing measures P_x and P_y.

LIMIT THEOREMS: EMPIRICAL DISTRIBUTIONS

Again, convergence of the marginal empirical distributions follow from the c.i.d. property. But we also have convergence for the joint empirical distribution:

Theorem. *If $((X_n, Y_n))$ is partially $\mathcal{G}$-c.i.d.*

$$\frac{1}{n}\sum_{i=1}^{n}\delta_{(X_i,Y_i)} \Rightarrow P_x \times P_y, \quad a.s. - \mathbf{P}.$$

A central limit theorem. Let

$$\begin{aligned} C_{n,x} &= \sqrt{n}\ \Big[\frac{\sum_{i=1}^{n} f(X_i)}{n} - \mathbb{E}(f(X_{n+1} \mid \mathcal{G}_n)\ \Big] \\ C_{n,y} &= \sqrt{n}\ \Big[\frac{\sum_{i=1}^{n} g(Y_i)}{n} - \mathbb{E}(g(X_{n+1} \mid \mathcal{G}_n)\ \Big] \end{aligned}$$

If $((X_n, Y_n))$ is partially $\mathcal{G}$-c.i.d., under assumptions,

$$(C_{n,x}, C_{n,y}) \to \int N(0,\Sigma)d\mu(\Sigma),$$

where μ is the probability law of the random covariance matrix $\Sigma = \begin{bmatrix} U & T \\ T & V \end{bmatrix}$.

A CENTRAL LIMIT THEOREM

For real measurable functions f and g such that, for all n, $\mathbb{E}|f(X_n)| + \mathbb{E}|g(Y_n)| < \infty$, let

$$M_n = f(X_n) - n\mathbb{E}\left[f(X_{n+1}) \mid \mathcal{G}_n\right] + (n-1)\mathbb{E}\left[f(X_n) \mid \mathcal{G}_{n-1}\right],$$
$$L_n = g(Y_n) - n\mathbb{E}\left[g(Y_{n+1}) \mid \mathcal{G}_n\right] + (n-1)\mathbb{E}\left[g(Y_n) \mid \mathcal{G}_{n-1}\right].$$

Theorem. *Suppose $(X_n, Y_n)_n$ are partially $\mathcal{G}$-c.i.d.. If*

$$\mathbb{E}f(X_1)^2 + \mathbb{E}g(Y_1)^2 + \sup_n \mathbb{E}\, C_{n,x}^2 + \sup_n \mathbb{E}\, C_{n,y}^2 < \infty \text{ and}$$

$$\frac{1}{n}\sum_{i=1}^n M_i^2 \to U,\ \frac{1}{n}\sum_{i=1}^n L_i^2 \to V,\ \frac{1}{n}\sum_{i=1}^n M_i L_i \to T \ \ \mathbf{P}\text{-a.s.},$$

then $(C_{n,x}, C_{n,y}) \to \mathcal{N}(0, \Sigma)$ stably, where Σ is the random covariance matrix $\Sigma = \begin{bmatrix} U & T \\ T & V \end{bmatrix}$.

Remark. This result is for fixed f, g. We are working on uniform limits theorems.

Example 1: randomly reinforced urns (RRU)

Two color urn; sampling with random reinforcement. Initial composition α_1/α white balls. For $n \geq 1$

$$\theta_n = P(X_{n+1} = 1 \mid X_{1:n}, W_{1:n}) = \frac{\alpha_1 + \sum_{i=1}^n W_i X_i}{\alpha + \sum_{i=1}^n W_i}$$

for random weights $W_1, W_2, \ldots$.

- weight is associated to individuals, (independently on the color taken by X_n), or to time:

  $$W_n \perp\!\!\!\perp X_n \mid X_1, \ldots, X_{n-1}, W_1, \ldots, W_{n-1}.$$

  Then the process (X_n) is $\mathcal{G}$-c.i.d., with $\mathcal{G}_n = \sigma(X_{1:n}, W_{1:n})$ (Pemantle (1989); Athreya (1969) and Athreya, Ney (1972); Berti, Pratelli, Rigo (2004; 2011)).

- (random) weights are associated to colors, and possibly depend on an observable variable: $W_n = w(X_n, Y_n)$. Then the process is *quasi*- $\mathcal{G}$-c.i.d.: the sequence (θ_n) is a quasi-martingale. Yet, for binary X_n, this is sufficient to prove asymptotic properties.

RRU FOR ADAPTIVE CLINICAL TRIALS

Two treatments, A and B (or K treatments, e.g. doses). Y response to treatment (success–insuccess; or discrete, or continuous). Let $P(Y_n = 1 \mid X_n = 1) = \alpha_A$ and $P(Y_n = 1 \mid X_n = 0) = \alpha_B$.

At step n, pick a ball. If $X_n = 1$ (white), assign patient n to treatment A; if $X_n = 0$, treatment B. Observe the response Y_n: if $Y_n = 1$, reinforce with one ball of the same color; otherwise, no reinforcement.

Thus, $W_n = w(X_n, Y_n)$, and $W_n \sim \text{Bernoulli}(\alpha_A)$ if $X_n = 1$, while $W_n \sim \text{Bernoulli}(\alpha_B)$ if $X_n = 0$.

- If $\alpha_A = \alpha_B$, W_n is independent on X_n and the process is $\mathcal{G}$-c.i.d. This case is of interest as a null hypothesis.
- If $\alpha_A \neq \alpha_B$, the process (X_n) is no longer c.i.d. The sequence $(\theta_n = P(X_{n+1} = 1 \mid \mathcal{G}_n))$ is a quasi-martingale.

  In particular, if $\alpha_A > \alpha_B$ (say), $\theta_n = P(X_{n+1} = 1 \mid X_{1:n}, W_{1:n}) \to 1$: the probability that the next patient is given the best treatment converges to one.

  $\to$ The best treatment (color) tends to dominate.

EXAMPLE 1: INTERACTING URNS

Consider two (but can be a countable system of) randomly reinforced urns.

$$P(X_{n+1} = 1 \mid X_{1:n}, W_{1:n}) = \frac{\alpha_1 + \sum_{i=1}^n W_i\, X_i}{\alpha + \sum_{i=1}^n W_i}$$
$$P(Y_{n+1} = 1 \mid Y_{1:n}, W'_{1:n}) = \frac{\alpha'_1 + \sum_{i=1}^n W'_i\, Y_i}{\alpha + \sum_{i=1}^n W'_i}$$

Let $\mathcal{G}_n = \sigma(X_{1:n}, Y_{1:n}, W_{1:n}, W'_{1:n})$. If

$$X_{n+1} \perp\!\!\!\perp W_{n+1}, Y_{1:n+1}, W'_{1:n+1} \mid X_{1:n}, W_{1:n}$$
$$Y_{n+1} \perp\!\!\!\perp W'_{n+1}, X_{1:n+1}, W_{1:n+1} \mid Y_{1:n}, W'_{1:n}$$

then the process $((X_n, Y_n))$ is partially $\mathcal{G}$-c.i.d. It is also partially $\sigma(X_{1:n}, Y_{1:n})$-c.i.d..

Examples:

- $W_n = W'_n$
- $W_n = w_n(Y_{1:n})$ and $W'_n = w'_n(X_{1:n})$, with $X_{n+1} \perp\!\!\!\perp Y_{n+1} \mid X_{1:n}, Y_{1:n}$;
- $W_n = w_n(\mathbf{Z_{1:n}})$ and $W'_n = w'_n(\mathbf{Z_{1:n}})$, with $Z_{n+1} \perp\!\!\!\perp X_{n+1}, Y_{n+1} \mid X_{1:n}, Y_{1:n}$;
- $W_n = w_n(Y_{1,n}, Z_{1:n})$ and $W'_n = w'_n(X_{1:n}, Z_{1:n})$, with $Z_{1:n+1} \perp\!\!\!\perp X_{n+1}, Y_{n+1} \mid X_{1:n}, Y_{1:n}$

  Some results are given by PAGANONI, SECCHI (2004).

A CENTRAL LIMIT THEOREM FOR INTERACTING URNS

Let (X_n, Y_n) be partially $\mathcal{G}$-c.i.d. RRUs, and suppose the weights W_i are i.i.d., with $\mathbb{E}\left(W_1^2\right) < \infty$.
Tedious calculations give that, **P**-a.s.,

$$\begin{aligned} \frac{1}{n}\sum_{i=1}^{n} M_i^2 &\rightarrow \delta\theta_x(1-\theta_x); \\ \frac{1}{n}\sum_{i=1}^{n} L_i^2 &\rightarrow \delta\theta_x(1-\theta_x); \\ \frac{1}{n}\sum_{i=1}^{n} M_i L_i &\rightarrow \delta(\eta-\theta_x\theta_y), \end{aligned}$$

where $\theta_x = \lim \frac{\sum_{i=1}}{n} X_i$, $\theta_y = \lim \frac{\sum_{i=1}}{n} Y_i$, $\eta = \lim \frac{\sum_{i=1}^{n} X_i Y_i}{n}$, and $\delta = \frac{\mathrm{Var}(W_1)}{\mathbb{E}(W_1)^2}$.
By the central limit theorem for partially $\mathcal{G}$-cid sequences:

$$\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n} X_i - \mathbb{E}[X_{n+1}|\mathcal{G}_n], \frac{1}{n}\sum_{i=1}^{n} Y_i - \mathbb{E}[Y_{n+1}|\mathcal{G}_n]\right) \overset{stably}{\rightarrow} \mathcal{N}\left(0, \begin{bmatrix} \delta\theta_x(1-\theta_x) & \delta(\eta-\theta_x\theta_y) \\ \delta(\eta-\theta_x\theta_y) & \delta\theta_y(1-\theta_y) \end{bmatrix}\right)$$

Example 2: randomly reinforced processes

Generalized species sampling sequences, or Ottawa sequences have been proposed by Bassetti, Crimaldi, Leisen (2010) as a generalization of species sampling models.

A natural extension is as follows.

$$\begin{aligned} P(X_{n+1} \in \cdot \mid X_{1:n}, Y_{1:n}, G_n) &= \frac{\alpha_x G_n(\cdot) + \sum_{i=1}^n W_i \delta_{X_i}(\cdot)}{\alpha_x + \sum_{i=1}^n W_i} \\ P(Y_{n+1} \in \cdot \mid X_{1:n}, Y_{1:n}, G_n) &= \frac{\alpha_y G_n(\cdot) + \sum_{i=1}^n W_i' \delta_{Y_i}(\cdot)}{\alpha_y + \sum_{i=1}^n W_i'} \end{aligned}$$

Remark. Even with $W_n = W_n' = 1$, the sequence (X_n, Y_n) would not be partially exchangable, unless $\alpha_x = \alpha_y$. This case reduces to (X_n, Y_n) being exchangeable, with a Dirichlet process directing measure.
Yet, for appropriate G_n, if

$$\begin{aligned} X_{n+1} &\perp\!\!\!\perp W_{n+1}, Y_{1:n+1}, W'_{1:n+1} \mid X_{1:n}, W_{1:n}, G_n \\ Y_{n+1} &\perp\!\!\!\perp W'_{n+1}, X_{1:n+1}, W_{1:n+1} \mid Y_{1:n}, W'_{1:n}, G_n \end{aligned}$$

then (X_n, Y_n) is partially c.i.d. with respect to the filtration $\mathcal{G}_n = \sigma(X_{1:n}, Y_{1:n}, W_{1:n}, W'_{1:n}, G_n)$. Thus it is also partially c.i.d. with respect to a smaller filtration.

The sequences (X_n) and (Y_n) can be regarded as dependent generalized species sampling models. If G_0 is diffuse, the two population have distinct species, but the reinforcement in one population may depend on samples from the other species; or, both reinforcements may depend on common (latent or observable) variables Z_n.

EXAMPLES

Randomly reinforced bivariate DP.
Species sampling sequences are characterized by the predictive rule, where new species are sampled from a diffuse G_0.

Consider independent weights (W_n) and (W'_n). We can set $G_n = G_0$, a known distribution, having a discrete component and a diffuse component, preserving the partially c.i.d. property.

In this case, the two random probability measures have common atoms (generated by the discrete component of G_0) and distinct atoms (generated by the diffuse component of G_0).

For $W_n = W'_n = 1$, and G_0 an appropriate discrete distribution, the process reduces to the bivariate Dirichlet process by Muliere, Walker (2003). Extension could lead to weighted versions of the the dependent DP by Caron, Davy, Doucet (2007).

Weighted Hierarchical DP
Dependent randomly reinforced hierarchical Dirichlet processes (TEH, JORDAN, BEAL, BLEI (2006)) can be obtained by setting

$$G_n = \frac{\alpha'' \nu(\cdot) + \sum_{i=1}^n W''_i \delta_{Z_i}(\cdot)}{\alpha'' + \sum_{i=1}^n W''_i}$$

with (Z_n) a randomly reinforced process with $Z_n \perp\!\!\!\perp X_{1:n}, Y_{1:n} \mid Z_{1:n}, W''_{1:n}$. Here, the two populations share all atoms.

DISCUSSION AND SOME OPEN PROBLEMS

Problems of of interest in many areas.

Recent applications include a weighted Indian Buffet process (Berti, Crimaldi, Pratelli, Rigo, (*Ann. Prob.*, 2015),

and weighted preferential attachment rules (Caldarelli *et al.*, 2013), where edges associated to weights with higher expected value tend to dominate.

* Given the predictive rule, finding the explicit distribution of the random directing measure is difficult. The random directing measure is different for c.i.d. and *quasi-c.i.d.* sequences.

* Somehow opposite, one may want to determine a predictive scheme (weights) that converges to a target (non-random?) f_0 (importance sampling algorithm?)

* How are limit theorems for c.i.d., and partial c.i.d., connected with Bernstein von-Mises results, and consistency in sup-norm ? (e.g. Castillo+ Nickl, *Ann. Statist.*, 2013 and 2014).

More work to be done in case of multiple interacting systems...