A subjective tour

around foundations and modern trends

Sonia Petrone

Bocconi University, Milano

with thanks to Sandra Fortini

ISBA 2016, Sardinia, 13-17 June 2016

motivation

increasingly often, I hear Bayesian friends saying things such as

"*I am agnostic about the Bayesian approach*"

"*Bayes is just another tool*"

"*I just use whatever works*"

⋮

> [...] A sampling of views that I have heard is that "Bayes is merely another tool", describing Bayes as a mere technique and conveying the sense that there is nothing to set Bayesian methods apart from, say a multivariate analysis or a time-series analysis; that "Bayes = shrinkage", and quickly sliding to the position that any method that can be vaguely described as producing shrinkage inherits all of Bayes' benefits; and that "Bayes is magic", and from this, the suggestion that every Bayesian analysis, no matter how preposterous, is a good analysis.[...]

(Steve Mac Eachern, ISBA President 2016 – *ISBA Bulletin, March 2016*) ok, we want "Bayes in action"... but ehm!, in a *foundational* lecture..

My point is that the interpretation of prediction we take is not neutral, and changes our viewpoint, focus, interest..

I'd like to underline that through a subjective choice of basic topics

OUTLINE

- premise (probability, utility and decisions)
- basics: Bernoulli trials and Polya urns
- Reinforced processes in Bayesian statistics
- Randomly reinforced processes
- Exchangeable random structures and open issues..

A flash of history

Let's think of the historical environment where people like Ramsey, de Finetti, Savage - among the founders of modern Bayesian statistics - lived and worked .. back to 1920'-50'...

Perhaps not a coincidence that they were mathematicians, logicians, probabilists also involved in actuarial science, economics, information, risk and decision..

F.P. Ramsey (1903-1930)

Frank P. Ramsey (Cambridge, 1903, London 1930)

TRUTH AND PROBABILITY

& "Further Considerations" and "Probability and Partial Belief"

by

Frank Plumpton Ramsey, M.A.
Fellow and Director of Studies in Mathematics at King's College,
Lecturer in Mathematics in the University of Cambridge

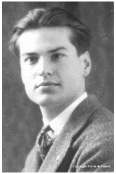
Cambridge of the 1920's, center of knowledge, Camdridge of the economists Piero Sraffa and John Maynard Keynes....

Ramsey starts his "Truth and Probability" arguing on Keynes vision on probability. His interest is in logic as the science of rationa thought...

B. de Finetti (13 June 1906 - 1985)

Bruno de Finetti (Innsbruck 13 June 1906 -- Roma 1985)

He worked at the Italian Central Statistical Institute, until 1931, and at the Assicurazioni Generali insurance company in Trieste



He worked on actuarial problems , life insurance mathematics, credibility theory and theory of risk….

His work in 1940 on the mean-variance approach for portoflio selection, largely anticipates Harry Markowitz's(Nobel Prize in 1990).

IL PROBLEMA DEI « PIENI » *)

B. DE FINETTI.

SUNTO. — Si esamina nei suoi diversi aspetti il problema del *rischio* derivante dalla copertura di un insieme di assicurazioni e, conseguentemente, il problema dei *pieni*, ossia del metodo più opportuno di cedere in riassicurazione una parte di tali assicurazioni per ridurre il rischio entro i limiti voluti colla minima perdita di guadagno. I diversi aspetti considerati sono: quello del rischio entro un singolo esercizio (Cap. I), del rischio per l'intero portafoglio esistente (Cap. II), del rischio relativo all'intero sviluppo futuro dell'impresa (Cap. III). Seguono (Cap. IV) delle considerazioni conclusive.

CAPITOLO PRIMO

IL PROBLEMA NELL'ÀMBITO DI UN ESERCIZIO.

1. *Impostazione.* — Il problema del rischio – e quello della determinazione dei pieni, che ne costituisce l'applicazione pratica – si può presentare e considerare sotto così vari aspetti e ha dato luogo

de Finetti Scoops Markowitz

Harry Markowitz
University of California at San Diego

Journal of Investment Management, Vol. 4, No. 3, Third Quarter 2006

Abstract:
In 1940, in the context of choosing optimum reinsurance levels, Bruno de Finetti essentially proposed mean-variance analysis with correlated risks. It was not until 1952 that Markowitz and Roy introduced mean-variance analysis with correlated risks into the financial literature. De Finetti solved the problem of computing mean-variance efficient frontiers for a particular constraint set (one that describes the reinsurance problem) assuming uncorrelated risks. While he understood and explained the importance of the case with correlated risks, he did not provide an algorithm for this case. In fact, one of his conjectures concerning its solution was incorrect. The present article summarizes de Finetti's contribution, presents an algorithm for solving "the de Finetti problem" when risks are correlated, and illustrates these matters with an easily visualized two-policy reinsurance problem.

Keywords: de Finetti, mean-variance analysis, critical line algorithm

L.J. Savage (1917 – 1971)

Leonard Jimmie Savage (1917, Detroit – 1971, New Haven, USA

Savage initially worked at the Institute for Advanced Study at Princeton….. and interacted with von Neuman, Milton Friedman Paul Samuelson..

…..And Abraham Wald…. Decision theory, and Sequential Analysis (1943)

This was just to accompany the change of vision.. from frequencies, to prediction, risk, decisions...

Savage's provides an axiomatization of rational decisions – preferences among actions – and of how they can reveal our beliefs (subjective probability measure) and utility...

This was just to accompany the change of vision.. from frequencies, to prediction, risk, decisions...

Savage's provides an axiomatization of rational decisions – preferences among actions – and of how they can reveal our beliefs (subjective probability measure) and utility...

Of course, the relation between probability and frequency remain fundamental - "*We want our beliefs to be consistent not merely with one another but also with the facts* (Ramsey, 1931)"

– There is active research in economics along Savage's lines (T.J. Sargent, C.A. Sims (Nobel in Economics, 2011), L.P. Hansen,...). It's weird if we, Bayesian statisticians, forget.. (but, recent work on intractable likelihood... Bissiri, Holmes, Walker, 2016; Watson & Holmes, 2016; Mengersen at al, 2016+..)

basics: Bernoulli trials and Pólya urn

Experiment: sampling with replacement from a 2-color urn, with unknown composition

any Bayesian statistician would say

$$X_i \mid \theta \overset{i.i.d}{\sim} \mathit{Bernoulli}(\theta), \quad \text{usually } \theta \sim \mathit{Beta}(\alpha_1, \alpha - \alpha_1).$$

This implies $P(X_1 = 1) = \alpha_1/\alpha$, and for $n \geq 1$

$$P(X_{n+1} = 1 \mid x_1, \ldots, x_n) = \frac{\alpha_1 + \sum_{i=1}^n X_i}{\alpha + n} \equiv Z_n$$

basics: Bernoulli trials and Pólya urn

Experiment: sampling with replacement from a 2-color urn, with unknown composition

any Bayesian statistician would say

$$X_i \mid \theta \overset{i.i.d}{\sim} \textit{Bernoulli}(\theta), \quad \text{usually } \theta \sim \textit{Beta}(\alpha_1, \alpha - \alpha_1).$$

This implies $P(X_1 = 1) = \alpha_1/\alpha$, and for $n \geq 1$

$$P(X_{n+1} = 1 \mid x_1, \ldots, x_n) = \frac{\alpha_1 + \sum_{i=1}^{n} X_i}{\alpha + n} \equiv Z_n$$

But this is the probability law of sampling (X_n) from a Pólya urn!

basics: Bernoulli trials and Pólya urn

Experiment: sampling with replacement from a 2-color urn, with unknown composition

any Bayesian statistician would say

$$X_i \mid \theta \overset{i.i.d}{\sim} Bernoulli(\theta), \quad \text{usually } \theta \sim Beta(\alpha_1, \alpha - \alpha_1).$$

This implies $P(X_1 = 1) = \alpha_1/\alpha$, and for $n \geq 1$

$$P(X_{n+1} = 1 \mid x_1, \ldots, x_n) = \frac{\alpha_1 + \sum_{i=1}^n X_i}{\alpha + n} \equiv Z_n$$

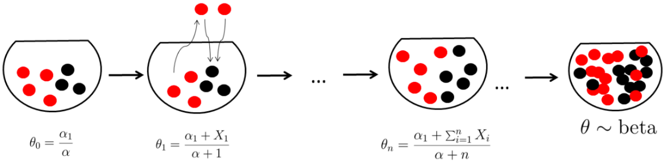
But this is the probability law of sampling (X_n) from a Pólya urn!

If we don't care about the subjective interpretation of probability, why do we use a Pólya urn to describe a Bernoulli urn?

Pólya urn

The Pólya urn is the basic example of a process with reinforcement.

An urn (population) initially contains α balls, of which α_1 white. At each step, a ball is drawn, its color is recorded, and the ball is returned in the urn, along with one ball of the same color.



Evolving process (spread of contagion). Interest in the urn composition (fraction of white balls Z_n)

$$Z_n \to \theta, \quad \text{a random limit, } \theta = \theta(\omega)$$

$$\theta \sim Beta(\alpha_1, \alpha - \alpha_1).$$

- **Evolutionary process**. e.g. Spread of contagion
- The sequence of colors (X_n) is exchangeable.

  By the representation theorem, the evolving process (X_n) is probabilistically equivalent to a static experiment:

  - pick an urn composition θ from $Beta(\alpha_1, \alpha - \alpha_1)$,
  - then sample with replacement.

- **subjective learning rule.**

  The experiment is static, but we learn from experience. Observing white ball *reinforces* our prediction that next ball is white.

  $\rightarrow$ The predictive rule describes the learning process.

  By exchangeability, it follows

  $X_i \mid \theta \overset{i.i.d}{\sim} Bernoulli(\theta), \quad \theta \sim Beta(\alpha_1, \alpha - \alpha_1).$

probability and frequency

- Spread of contagion: interest is in the urn composition Z_n. We know $Z_n \to$ random $\theta \sim Beta(\alpha_1, \alpha - \alpha_1)$.
  CLT for rescaled differences $\sqrt{n}(Z_n - \theta)$

  $$\frac{\sqrt{n}(Z_n - \theta)}{\theta(1-\theta)} \to \mathrm{N}(0,1), \quad \text{a.s.-}P$$

  a random rescaling – i.e. for large n, $Z_n \mid \theta \approx \mathrm{N}(\theta, \frac{\theta(1-\theta)}{n})$.
- subjective learning rule: interest in comparing probability and frequency, i.e. $\bar{X}_n = \sum_{i=1}^{n} X_i/n$. We know $\bar{X}_n \to \theta$ a.s. and CLT for rescaled differences $(Z_n - \bar{X}_n)$ or prediction errors $(Z_n - X_{n+1})$.

We want our beliefs to be consistent not merely with one another but also with the facts (Ramsey, 1931)

Results in this direction could be a 'Bayesian counterpart' to frequentist asymptotics

Results in this direction could be a 'Bayesian counterpart' to frequentist asymptotics

(But) they are a.s. with respect the exchangeable P.

Diaconis & Freedman (1990) give results (for a k-dice) that reformulate the usual statement "the posterior piles up around the true θ_0, a.s. $P_{\theta_0}^{\infty}$" as a finite-sample result, without exceptional null sets or "true values" of parameters.

Results in this direction could be a 'Bayesian counterpart' to frequentist asymptotics

(But) they are a.s. with respect the exchangeable P.

Diaconis & Freedman (1990) give results (for a k-dice) that reformulate the usual statement "the posterior piles up around the true θ_0, a.s. $P_{\theta_0}^{\infty}$" as a finite-sample result, without exceptional null sets or "true values" of parameters.

For coin tossing, they develop an explicit inequality: the posterior must concentrate near the observed fraction of heads. For ϕ-positive priors, posterior odds $R(n, \bar{X}_n, k)$ of the interval $[\bar{X}_n - h, \bar{X}_n + h]$ satisfy

$$R(n, \bar{X}_n, k) \geq \phi(h, \epsilon, \phi) \exp\{n(1-\epsilon)g(h)\}.$$

The inequality replaces the asymptotics and eliminates the null set; observed fraction stands in for the true parameter.

I feel there is still work to do on finite sample properties of Bayesian procedures.

There are well known results on admissibility of Bayesian rules, and on *frequentist accuracy* (Efron, 2015)..

but we'd want Bayesian criteria, based on information, and efficiency; and calibration, evaluation of forecasters (discussion in de Groot & Feinberg, 1982), scoring rules .. (recent discussion Dawid & Musio, BA 2015).. merging of opinion...

$\rightarrow$ see Stefano Rizzelli's and Sara Wade & Michele Peruzzi's posters

Predictive approach

The Pólya urn is a basic example of a predictive construction of an exchangeable sequence.

- Probability on observable facts – here, (X_n) exchangeable;
- express the predictive distributions P_n: the sequence P_n, $n \geq 1$ defines the law P of (X_n), and under conditions, an *exchangeable* P.
- by the representation theorem, $(X_n) \sim P$ is equivalent to "model+prior"
  - $P_n \Rightarrow \tilde{F}$ (and empirical dist $\hat{F}_n \Rightarrow \tilde{F}$) directing measure
  - $X_i \mid \tilde{F} = F \overset{i.i.d}{\sim} F$.

Notions of predictive sufficiency: $P(X_{n+1} \in \cdot \mid X_{1:n}) = P(X_{n+1} \in \cdot \mid T_n)$ may lead to a parametric directing measure (model) indexed by $\tilde{\theta} = \lim T_n$ (e.g. Fortini et al, 2000).

Processes with reinforcement

Important predictive constructions are based on processes with reinforcement (see Pematle 2007), the Pólya urn being a basic example.

- *k* colors Polya urn. Johnson's sufficiency postulate (Zabell, 1982): if $p_n(j) \equiv P(X_{n+1} = j \mid X_1, \ldots, X_n) = f_i(n_i)$, then

$$p_n(j) = \frac{\alpha_j + n_j}{\alpha + n}$$

  and $X_i \mid (\pi_1, \ldots, \pi_k) \overset{i.i.d}{\sim} \mathrm{Dir}(\alpha_1, \ldots, \alpha_k)$.

- Pólya sequences. extending to countable colors, $X_1 \sim F_0$ and

$$P_n(\cdot) = \frac{\alpha F_0(\cdot) + \sum_{i=1}^n \delta_{X_i}}{\alpha + n}.$$

  Then (X_n) is exchangeable and

  - $P_n \rightarrow_{TV} \tilde{F}$, a.s.
  - $\tilde{F} = \sum_{j=1}^{\infty} \pi_j \delta_{x_j^*}$ and $\tilde{F} \sim DP(\alpha F_0)$.

random partitions

the predictive rule that characterize a DP

$$X_{n+1} \mid X_{1:n} \text{ is an 'old value' } X_j^* \text{ with probability } n_j/(\alpha+n)$$
$$\text{or a new value } X_{k_n+1} \text{ with probability } \alpha/(\alpha+n)$$

implies ties: a sample $(X_{1:n})$ from $\tilde{F} \sim DP(\alpha F_0)$ is described by the random partition ρ_n and by the 'species' values x_j^*.

The probability law of ρ_n is given by Antoniak (1974)

$$P(\rho_n = (A_1, \ldots, A_k)) = \frac{\alpha^k}{\alpha^{[n]}} \prod_{j=1}^{k} (n_j - 1)!$$

where $\alpha^{[n]} = \alpha(\alpha+1)\cdots(\alpha+n-1)$ and n_j is the number of elements of A_j, $j = 1, \ldots, k$. When considering the allelic partition, it corresponds to the celebrated Ewens sampling formula (Ewens, 1972).

rich developments..

The predictive approach is the basis of many developments in Bayesian nonparametrics (a review in Fortini & P., 2012)

- it has been perhaps one of the main factors of the impetuous growth, in the last decade, of sound BNP constructions from the machine learning community (just an example is the Indian Buffet Process (Griffiths & Ghahramani, 2005) for feature selection, whose de Finetti measure was later obtained (Thibaux and Jordan, 2007).
- the Ewens sampling formula is a beautiful example of interdisciplinary cross-fertilization (Ewens & Tavare', 1998), from evolutionary molecular genetics, population dynamics, neutral theory of biodiversity, combinatorial stochastic processes...

..and beyond the Dirichlet process

- The clustering structure of the DP is quite parametric and lacks power law behavior (Kingman, 1978). Extensions: 2-parameter Poisson-Dirichlet, completely random measures.. see De Blasi, Favaro, Lijoi, Mena, Pruenster, Ruggiero (2013).
- Walker & Muliere (1999) give a characterization of neutral to the right processes via an extension of Johnson's sufficientness postulate
- For continuous data, Johnson's sufficientness postulate is unsatisfactory: in the predictive rule, one would like to spread the information carried by $n_A =$ frequency of observations $\in A$ to neighborhood sets. This motivates smoothing of the DP, e.g. by Bernstein polynomials (Diaconis & Ylvisaker, 1985) and mixture models.

Partial exchangeability

Predictive construction based on processes with reinforcement are the basis of important constructions of partially exchangeable processes. Partial exchangeability is invariance under a class of permutations.

- de Finetti partial exchangeability (de Finetti, 1937)
  consider $[X_{n,i}]_{n \geq 1; i \in I}$ (e.g., multiple experiments).
  $[X_{n,i}]$ is partially exchangeable if its probability law is invariant under separate finite permutations of the columns

$$[X_{k,i}]_{k=1,\ldots,n; i \in J} \stackrel{d}{=} [X_{\pi_i(k),i}]_{k=1,\ldots,n; i \in J}$$

- Markov exchangeability (Diaconis & Freedman, 1980)
- Row-column exchangeability (Hoover, 1979; Aldous, 1981; Kallenberg, 2005)

reinforced processes & partial exchangeability

* Edge reinforced random walks on a graph (Coppersmith & Diaconis, 1987; Pemantle, 1988) give predictive characterization of a conjugate prior for the transition matrix of a reversible Markov chain (Diaconis & Rolles, 2006). Extensions in Bacallado (2011) and Bacallado, Favaro, Trippa (2013).

* Walker & Muliere (1997), Muliere, Secchi & Walker (2000) characterize the Beta-Stacy process for Bayesian nonparametric survival analysis as the de Finetti measure of the lengths of x_0-blocks of a Markov exchangeable RUP

* Beal, Ghahramani & Rasmussen (2002) propose the infinite HMM through a hierarchical Hoppe's urn construction, which characterizes a Hierarchical Dirichlet process (Teh, Jordan, Beal, Blei, 2006) prior on the infinite random transition matrix.

The HDP, and extensions e.g. Hierarchical Pitman Yor (Wood, Gasthaus, Teh, 2008) are widely used for topics and language models and shared clustering. And many proposals have originated by Mc Eachern (1999)'s notion of Dependent Dirichlet Processes...

However, for partially exchangeable data, there seem to be no conjugate prior that is characterized by a simple predictive distribution.

$(X_n, Y_n) \mid F \overset{i.i.d}{\sim} F$ and $F \sim DP(\alpha F_0(\cdot, \cdot)$ would be conjugate, but implies$F_x \sim D(\alpha F_{0,x})$ and $F_y DP(\alpha F_{0,y})$; a too rigid, global, clustering. An enriched conjugate prior is proposed by Wade, Mongelluzzo & P. (2011).

But many natural generative models fail to be partially exchangeable.

interacting randomly reinforced processes

Consider a system of interacting sequences $(X_{n,i}, n \geq 1)$, $i \in I$, evolving according to a randomly reinforced scheme.

$X_{1,i} \sim \nu_i(\cdot)$, for a distribution ν_i on $\mathcal{X}$, and for any $n \geq 1$,

$$\mathbb{P}[X_{n+1,i} \in \cdot \mid X_{1:n}, W_{1:n}] = \frac{w_{0,i}\nu_i(\cdot) + \sum_{k=1}^{n} W_{k,i}\delta_{X_{k,i}}(\cdot)}{w_{0,i} + \sum_{k=1}^{n} W_{k,i}}.$$

where $w_{0,i}$ is a constant and $W_{1,i}, W_{2,i}, \ldots$ are population specific weights, generally random.

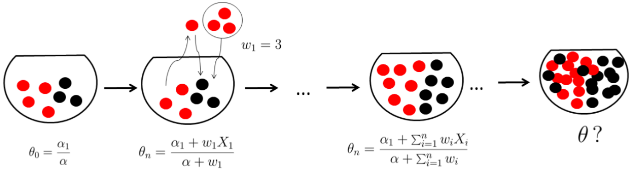
This can be a natural construction.. But not stationary, and not partially exchangeable.

Let's start from the univariate case.

Processes with random reinforcement

Let's consider again with a 2-color Pólya urn, but now suppose that the reinforcement, i.e. number of additional balls at each step, may vary along time (Pemantle, 1990) or be random:

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



(X_n) is no longer stationary: time cannot be ignored.

This is a special family of Generalized Polya Urns, with diagonal replacement matrix. Application include clinical trials, economics and finance, information science, network theory...

Processes with random reinforcement

In a predictive approach, we may want to weight differently past observations.

Consider the predictive rule: $X_1 \sim F_0$ and for $n \geq 1$

$$P_n(\cdot) = P(X_{n+1} \in \cdot \mid X_{1:n}) = \frac{\alpha F_0(\cdot) + \sum_{i=1}^{n} W_i \delta_{X_i}(\cdot)}{\alpha + \sum_{i=1}^{n} W_i}$$

A weighted CRP where the probability of joining a table [a node in a network] depends not only on the number of customers [degree of the node] but also on their weight (influence, authority, ...).

These are generalized species sampling models (Bassetti, Crimaldi, Leisen, 2010); specifications include generalized Ottawa sequences and beta-GOS (Airoldi, Costa, Bassetti, Leisen, Guindani, 2014; Bassetti, Leisen†, Airoldi, Guindani, 2015+)

c.i.d. sequences

If, in the above predictive rule, $W_n \perp\!\!\!\perp X_n$, the process (X_n) satisfies

$X_{n+1} \mid X_{1:n} \overset{d}{=} X_{n+k} \mid X_{1:n}$, for any k conditionally identically distributed

c.i.d. is a "marginal spreadability" property [spreadability: invariance under selection of indexes, and equivalent to exchangeability];
(X_n) c.i.d. and stationary iff (X_n) is exchangeable (Kallenberg, 1988).

A c.i.d. process is not, in general, stationary: so, no equivalence with a static process is possible. Rather, $X_n \mid F_n \overset{indep}{\sim} F_n$ and F_n has an "unpredictable" evolution, $E(F_{n+h} \mid X_{1:n}) = E(F_{n+1} \mid X_{1:n})$, for any $h \geq 1$.

asymptotic behavior

Although generally non-exchangeable, a c.i.d. sequence preserves limit properties of exchangeable sequences

- the predictive P_n is a measure-valued martingale, and (using stable convergence), $\lim P_n = \lim \hat{F}_n = \tilde{F}$
- Convergence of P_n implies (e.g. Aldous, 1985) that (X_n) is asymptotically exchangeable (informally, for m large, $(X_{m+1}, X_{m+2}, \ldots) \approx (Z_1, Z_2, \ldots)$ for (Z_n) exchangeable), and the directing measure of (Z_n) is $\tilde{F}$
- SLLN, CLT and stronger limit theorems are available (Berti, Pratelli & Rigo, 2004).

remark: If W_n and X_n are not independent (i.e. the weight depends on the color extracted), the process (X_n) is not c.i.d., but asymptotic results, including asymptotic exchangeability, can still hold.

limit law

The probability law of $\tilde{F}$ is generally difficult to obtain explicitly. Some results are given for specific constructions.

An interesting result (Berti, Crimaldi, Pratelli, Rigo, 2011,2012) is that, in the multicolor randomly reinforced urn, where weights W_j are associated to color j, if there is a set of *dominating colors*

$$\mathcal{D} = \{j : E(W_j) = \mu > 0\}, \quad \text{with } E(W_j) < \mu \text{ for all } j \notin \mathcal{D},$$

then the predictive probability $p_n(j)$ of color j converges to zero for colors $j \notin \mathcal{D}$, while it converges to a random variable $Z_j \in (0, 1]$ a.s. for $j \in \mathcal{D}$ and $\sum_{j \in \mathcal{D}} Z_j = 1$.

extensions

Consider an array of r.v.'s

$$[X_{n,i}]_{n\geq 1, i\in I} = \begin{bmatrix} X_{1,1} & \cdots & X_{1,k} \\ X_{2,1} & \cdots & X_{2,k} \\ \vdots & \vdots & \vdots \\ X_{n,1} & \cdots & X_{n,k} \\ \vdots & \vdots & \vdots \end{bmatrix}$$

Paralleling de Finetti's extension from exchangeability to partial exchangeability, a natural question is how the notion of c.i.d. sequences can be extended to partially c.i.d. arrays (dependent, or interacting, processes).

partially c.i.d. arrays

A notion of partially c.i.d. arrays is given by Fortini, P. and Sporysheva (2016), so that it is equivalent to partial exchangeability under stationarity,

A partially c.i.d. process is asymptotically partially exchangeable
The directing measure is a vector of random probability measures $[\tilde{F}_1, \ldots, \tilde{F}_k]$ (in general, dependent), that is the a.s. limit of the sequence of the *joint* predictive, and empirical, distributions.
In principle, partially c.i.d. generative models define novel priors for dependent random measures.

SLLN and CLT for the the joint predictive errors $(E(X_{n+1,i} \mid \mathcal{G}_{\backslash}) - \mathcal{X}_{\backslash,\rangle})$ and for the rescaled differences $(\bar{X}_{n,i} - E(X_{n+1,i} \mid \mathcal{G}_{\backslash})$.

Exchangeable random structures

Consider an array of r.v.'s $[X_{i,j}]_{i,j \in N}$;

e.g., i denotes users, j movies; or, i, j nodes of a graph..

- $[X_{i,j}]$ is separately exchangeable if $[X_{i,j}] \stackrel{d}{=} [X_{\pi(i),\sigma(j)}]$ for any pair of permutations π and σ of $\mathbf{N}$.
- $[X_{i,j}]$ is jointly exchangeable if $[X_{i,j}] \stackrel{d}{=} [X_{\pi(i),\pi(j)}]$.

Aldous-Hoover representation theorem gives

$$[X_{i,j}] = \stackrel{d}{=} [h(\alpha, U_1, U_j, U_{i,j}]$$

where $\alpha, U_i, U_j, U_{i,j}$ are i.i.d. $U(0,1)$.

$(X_{i,j} \mid \theta_{i,j} \stackrel{indep}{\sim} Bernoulli(\theta_{i,j})$ and $\theta_{i,j} = w(\alpha, U_i, U_j))$.

Popular statistical models for networks are based on the notion of RCE; see Goldenberg, Zheng, Fienberg, Airoldi (2010) for a review, and Hoff .

Exchangeability and graph limits

A random graph $G = (V, E)$ is exchangeable if its distribution is invariant under every permutation of the vertexes, or, equivalently, if its adjacency matrix $[X_{i,j}]$ is jointly RCE.

A beautiful theory on large graphs and graph limits has been developed by Lovatz and collaborators (Lovatz, 2013) and has strict connections with Aldous-Hoover-Kallenberg representation theorem for RCE arrays. This is well explained by Diaconis & Janson (2007); also Orbanz & Roy (2015).

Exchangeability and sparsity

A debated issue is that exchangeable graphs are either dense or empty, while real network tend to be sparse and show power law behavior....

Recent proposals:

Caron & Fox (2014), based on completely random measures; Veitch & Roy (2015); Broderick & Cai (2015); Crane (2016), based on new notions of edge exchangeability ..

(just a conjecture/ongoing work)

Randomly reinforced preferential attachment can model competition, selection and non-stationary graph evolution.

For random graphs, there is analogue of convergence of empirical distributions (Kallenberg, 1999).

If we have an analogue of convergence of the predictive distributions,

then a randomly reinforced preferential attachment, with a c.i.d. or quasi c.i.d. evolution, might lead to an exchangeable limit graphon on the set of dominating edges $\mathcal{D}$, possibly sparse in $\mathcal{V}$.

(perhaps more at the Isaac Newton Programme, Cambridge, 2016.....)

final remarks

Bayes is not just another tool.

The exercise of thinking about the role of probability, when approaching problems and modern challenges, is useful, at least for clarity... and for much more.

Thank you