Bayesian nonparametric inference for hidden Markov models: an overview and some new insights.

Sonia Petrone
joint work with Sandra Fortini
Bocconi University, Milano

April 12, 2013 – Institute for Statistics and Mathematics
Vienna University of Economics and Business
Hidden Markov models (HMM) are popular tools for time series analysis, speech recognitions, ...

Observable process \((Y_t)\) and latent process \((\theta_t)\). Assume

- \((\theta_t)\) is a Markov chain with state space \(\{\xi_1, \ldots, \xi_k\}\),
- and conditionally on the \(\theta_t\)'s, the \(Y_t\) are independent:

\[
Y_t \mid \theta_t = \xi_i \overset{indep}{\sim} f(y \mid \xi_i)
\]
Hidden Markov models (HMM) are popular tools for time series analysis, speech recognitions, ...

Observable process \((Y_t)\) and latent process \((\theta_t)\). Assume
- \((\theta_t)\) is a Markov chain with state space \([\xi_1, \ldots, \xi_k]\),
- and conditionally on the \(\theta_t\)'s, the \(Y_t\) are independent:
\[
Y_t \mid \theta_t = \xi_i \ind \sim f(y \mid \xi_i)
\]

A general problem in HMMs is **how to choose the number of states**.

A Bayesian nonparametric approach offers an interesting solution, which allows to generate states as the need occurs.

Furthermore, efficient computational tools are available, which exploit the predictive construction of nonparametric priors.
In recent years, there has been an impressive explosion of Bayesian nonparametric methods, in the statistical literature and in other areas, such as machine learning.

This talk is an overview of some of these constructions, and gives theoretical insights on connections across apparently unrelated literature


- many recent proposals in machine learning, such us hierarchical Dirichlet Processes (Teh et al., 2006), infiniteHMM (Beal, Ghahramani, Rasmussen, 2002), sticky infiniteHMM (Fox et al., 2007) Indian Buffet (Griffiths & Ghahramani, 2006), ...)

Here these methods are used for BNP: predictive construction of nonparametric priors for exchangeable and Markov exchangeable sequences.
• Introduction: predictive construction of nonparametric priors
  • Exchangeability.
    Dirichlet process and Hoppe’s urn
  • Markov exchangeability
    Reinforced urn processes.
• Introduction: predictive construction of nonparametric priors
  • Exchangeability.
    Dirichlet process and Hoppe’s urn
  • Markov exchangeability
    Reinforced urn processes.
• Nonparametric priors for HMMs
  • Hierarchical Hoppe’s urns
  • Infinite HMMs
• Introduction: predictive construction of nonparametric priors
  • Exchangeability.
    Dirichlet process and Hoppe’s urn
  • Markov exchangeability
    Reinforced urn processes.
• Nonparametric priors for HMMs
  • Hierarchical Hoppe’s urns
  • Infinite HMMs
• Related processes and extensions
1. Exchangeability

Consider first an exchangeable sequences \((Y_i)\) with probability law \(P\). From de Finetti representation theorem, there exist a unique probability measure (prior) on \(\mathcal{P}(\mathcal{Y})\) such that

\[ Y_i \mid F \overset{i.i.d.}{\sim} F, \quad F \sim \pi. \]

The random d.f. \(F\) is the weak limit of the sequence of empirical distributions

\[ \hat{F}_n = \frac{1}{n} \sum_{i=1}^{n} \delta_{Y_i}, \]

but it is also the limit of the sequence of predictive distributions

\[ P_n(y) = P(Y_{n+1} \leq y \mid y_1, \ldots, y_n), \quad n = 1, 2, \ldots. \]

In a predictive approach, rather than specifying the model and the prior, one wants to construct them starting from predictive assumptions, i.e., assigning a sequence of predictive distributions \(P_n\), that characterize a law \(P\) for \((Y_i)\) that is exchangeable.

Then, at least in principle, one has characterized the random \(F\), i.e. the model and the prior.
Suppose \((Y_i)\) is exchangeable, and assume that \(Y_1 \sim F_0\) and for any \(n \geq 1\)

\[
P_n(y) \equiv P(Y_{n+1} \leq y \mid y_1, \ldots, y_n) = \frac{\alpha}{\alpha + n} F_0(y) + \frac{n}{\alpha + n} \sum_{i=1}^{n} \frac{1}{n} \delta y_i,
\]

a weighted average of a prior guess \(F_0\) and the empirical d.f. \(\frac{1}{n} \sum_{i=1}^{n} \delta y_i\).
Dirichlet process

Suppose \((Y_i)\) is exchangeable, and assume that \(Y_1 \sim F_0\) and for any \(n \geq 1\)

\[
P_n(y) \equiv P(Y_{n+1} \leq y \mid y_1, \ldots, y_n) = \frac{\alpha}{\alpha + n} F_0(y) + \frac{n}{\alpha + n} \sum_{i=1}^{n} \frac{1}{n} \delta_{y_i},
\]

a weighted average of a prior guess \(F_0\) and the empirical d.f. \(\frac{1}{n} \sum_{i=1}^{n} \delta_{y_i}\).

Blackwell and McQueen (1973) showed that the Polya sequence \((Y_n)\) so defined is exchangeable, more specifically

- \(P_n \Rightarrow F\), a.s.;
- \(F\) is discrete a.s. and \(F \sim DP(\alpha F_0)\)
- \(Y_i \mid F \stackrel{i.i.d.}{\sim} F\).

One can show that \(F = \sum_{j=1}^{\infty} w_j \delta_{\xi_j}\) where the atoms \(\xi_j \stackrel{i.i.d.}{\sim} F_0\) and the weights \((w_j)\) have a stick-breaking prior \((\alpha)\), independently on the \((\xi_j)\).
The discrete nature of the DP is at the basis of many applications to BNP mixture models and clustering. For understanding the implications of the predictive rule in terms of random partitions, and thus the potentiality in clustering and more, it is useful to use an urn representation, proposed by Hoppe (1984).
The discrete nature of the DP is at the basis of many applications to BNP mixture models and clustering. For understanding the implications of the predictive rule in terms of random partitions, and thus the potentiality in clustering and more, it is useful to use an urn representation, proposed by Hoppe (1984).

**Hoppe’s urn** (Hoppe, 1984). Urn with $\alpha$ black balls. Pick a ball: if black, return it together with a ball of new color; if colored, return 2 balls of the same color.

Initially, we pick a black ball, generate a color and set $S_1 = 1$. Then, we pick a ball: if colored, we set $S_2 = 1$; if black, generate a new color, and set $S_2 = 2$, and so on.
Hoppe’s urn

By construction, $S_1 = 1$,

$$S_2 \mid S_1 = 1 \sim \begin{cases} 
  2 & \text{with prob. } \frac{\alpha}{\alpha+1} \\
  1 & \text{with prob. } \frac{1}{\alpha+1} 
\end{cases}$$

and for $n \geq 1$, denoting by $k$ the number of colors discovered in $(S_1, \ldots, S_n)$ and by $n_j$ the frequency of color $j$

$$S_{n+1} \mid S_1, \ldots, S_n = \begin{cases} 
  k+1 & \text{with prob. } \frac{\alpha}{\alpha+n} \\
  j & \text{with prob. } \frac{n_j}{\alpha+n}, j = 1, \ldots, k 
\end{cases}$$
Hoppe’s urn

By construction, \( S_1 = 1 \),

\[
S_2 \mid S_1 = 1 \sim \begin{cases} 
2 & \text{with prob. } \frac{\alpha}{\alpha+1} \\
1 & \text{with prob. } \frac{1}{\alpha+1}
\end{cases}
\]

and for \( n \geq 1 \), denoting by \( k \) the number of colors discovered in \((S_1, \ldots, S_n)\) and by \( n_j \) the frequency of color \( j \)

\[
S_{n+1} \mid S_1, \ldots, S_n = \begin{cases} 
k+1 & \text{with prob. } \frac{\alpha}{\alpha+n} \\
j & \text{with prob. } \frac{n_j}{\alpha+n}, j = 1, \ldots, k
\end{cases}
\]

The sequence \((S_1, \ldots, S_n)\) defines a random partition of \(\{1, 2, \ldots, n\}\). For example, for \( n = 6 \), \((S_1, \ldots, S_n) = (1, 2, 1, 1, 3, 2)\) gives the partition \((\{1, 3, 4\}, \{2, 6\}, \{5\})\).
colored Hoppe’s urn

The sequence \((S_n)\) is not exchangeable

However, if we color \((S_n)\) with colors \(\xi_i \sim_{i.i.d} p_0\), the resulting sequence of colors \((X_n)\) is a Polya sequence

\[
Y_{n+1} \mid y_1, \ldots, y_n \sim \frac{\alpha}{\alpha + n} p_0 + \frac{1}{\alpha + n} \sum_{j=1}^{k} n_j \delta_{y_j^*}
\]

thus it is exchangeable and its de Finetti measure is a \(DP(\alpha p_0)\).
Applications in mixture models

This clustering property is often used at the second stage of Bayesian hierarchical models. Consider

\[ Y_i \mid \theta_i \overset{\text{indep}}{\sim} f(y \mid \theta_i) \]

\[ \theta_i \mid G \overset{i.i.d}{\sim} G \]

\[ G \sim \pi \]

Integrating the \( \theta_i \) out, we obtain a mixture model

\[ Y_i \mid G \overset{i.i.d}{\sim} \int f(y \mid \theta) dG(\theta). \]
Applications in mixture models

This clustering property is often used at the second stage of Bayesian hierarchical models. Consider

\[ Y_i \mid \theta_i \overset{\text{iid}}{\sim} f(y \mid \theta_i) \]
\[ \theta_i \mid G \sim G \]
\[ G \sim \pi \]

Integrating the $\theta_i$ out, we obtain a mixture model

\[ Y_i \mid G \overset{\text{iid}}{\sim} \int f(y \mid \theta)dG(\theta). \]

If we assume $G \sim DP(\alpha G_0)$, then

- $G$ is a.s. discrete, $G = \sum_{j=1}^{\infty} w_j \delta_{\xi_j}$, thus the model reduces to a countable mixture

\[ Y_i \mid G \overset{\text{iid}}{\sim} \sum_{j=1}^{\infty} w_j f(y \mid \xi_j). \]

- the predictive structure implies a prior on the random clustering of $(\theta_1, \ldots, \theta_n)$. 
Markov exchangeability

Let us now consider Bayesian inference for a Markov chain \((Y_t)\), with state space \(\{1, \ldots, k\}\).
We assume that \((Y_t)\) is a Markov chain conditionally on the unknown transition matrix \(\Pi = \{\pi_{i,j}\}\) and we should assign a prior on \(\Pi\).

Again, we can take a predictive approach: define predictive distributions \(P_n\) such that \((Y_n)\) is Markov exchangeable and recurrent. This characterizes the prior on the transition matrix \(\Pi\).
Markov exchangeability

Let us now consider Bayesian inference for a Markov chain \((Y_t)\), with state space \(\{1, \ldots, k\}\). We assume that \((Y_t)\) is a Markov chain conditionally on the unknown transition matrix \(\Pi = \{\pi_{i,j}\}\) and we should assign a prior on \(\Pi\).

Again, we can take a predictive approach: define predictive distributions \(P_n\) such that \((Y_n)\) is Markov exchangeable and recurrent. This characterizes the prior on the transition matrix \(\Pi\).

Fortini and Petrone (2013) give necessary and sufficient conditions on the predictive rules \(P_n, n \geq 1\) so that they characterize a law \(P\) for \((Y_n)\) such that \((Y_n)\) is recurrent and Markov exchangeable.
Markov exchangeability

exchangeability $\iff$ mixture of i.i.d.

$?? \iff$ mixture of Markov chains

If $(Y_n)$ is recurrent, “??” is Markov exchangeability (Diaconis+Freedman, 1980).
Markov exchangeability

exchangeability ⇔ mixture of i.i.d.
?? ⇔ mixture of Markov chains

If \((Y_n)\) is recurrent, “??” is Markov exchangeability (Diaconis+Freedman, 1980).

\((Y_n, n \geq 0), Y_i \in I\) countable.

\((Y_n)\) is recurrent if \(P(Y_n = Y_0 \text{ for infinitely many } n) = 1\).
Markov exchangeability

exchangeability ⇔ mixture of i.i.d.

⇔ mixture of Markov chains

If \((Y_n)\) is recurrent, “??” is Markov exchangeability (Diaconis+Freedman, 1980).

\((Y_n, n \geq 0), Y_i \in I\) countable.

\((Y_n)\) is recurrent if \(P(Y_n = Y_0 \text{ for infinitely many } n) = 1\).

Two sequences \(y = (y_1, \ldots, y_n)\) and \(x = (x_1, \ldots, x_n)\) are equivalent, \(y \sim x\) iff they start from the same state and have same transitions counts e.g.: \(y = (1, 3, 2, 1, 2)\) and \(x = (1, 2, 1, 3, 2)\).

\((Y_n)\) is Markov exchangeable if \(y \sim x\) implies

\[ P(Y_1 = y_1, \ldots, Y_n = y_n) = P(Y_1 = x_0, \ldots, Y_n = x_n). \]
**Theorem**

*Suppose* $(Y_n)$ *is recurrent. Then* $(Y_n)$ *is Markov exchangeable iff it is a mixture of Markov chains,*

$$P(Y_0 = y_0, \ldots, Y_n = y_n \mid Y_0 = y_0) = \int \prod_{i=1}^{n} \Pi(y_i \mid y_{i-1}) \, d\mu(\Pi \mid y_0).$$

*The prior* $\mu$ *is uniquely determined.*

In other words: $(Y_n) \mid \Pi$ *is Markov, with state space* $I$ *and transition matrix* $\Pi$, *and* $\Pi \sim \mu$. 
An equivalent definition


**successors matrix** $S$: $i$th row $(X_{i,n}, n \geq 1)$ successors of state $i$.

**Theorem**

$(X_n)$ is a mixture of Markov chains iff each state that is visited has an infinite number of successors, and the successor matrix $S$ is partially exchangeable by rows.

The prior on the $i$th row of the transition matrix is the de Finetti measure of the exchangeable sequence of the successors of state $i$. 
An equivalent definition


**successors matrix** $S$: $i$th row $(X_{i,n}, n \geq 1)$ successors of state $i$.

**Theorem**

$(X_n)$ is a mixture of Markov chains iff each state that is visited has an infinite number of successors, and the successor matrix $S$ is partially exchangeable by rows.

The prior on the the $i$th row of the transition matrix is the de Finetti measure of the exchangeable sequence of the successors of state $i$.

Remark: this is useful in urn schemes where draws from urn $U_i$ represent the successors of state $i$. If recurrent and Markov exchangeable, one can also characterize the prior and the dependence across the rows of $\Pi$. 
Interpretation of $\Pi$

- From Diaconis and Freedman (1980), $\Pi$ is the limit of the matrix of transition counts
  
  $$
  \frac{T_{i,j}(X_1, \ldots, X_n)}{T_{i+}(X_1, \ldots, X_n)} \to \Pi_i(j), \quad \text{a.s.} \, P
  $$

- From Fortini et al. (2002), it is the limit of the sequence of empirical distributions of the successors:
  
  $$(\frac{\sum_{k=1}^{n} \delta_{X_{1,k}}}{n}, \frac{\sum_{k=1}^{n} \delta_{X_{2,k}}}{n}) \to (\Pi_1, \Pi_2) \sim \mu_{1,2}(\cdot, \cdot)$$
Interpretation of $\Pi$

- From Diaconis and Freedman (1980), $\Pi$ is the limit of the matrix of transition counts

$$\frac{T_{i,j}(X_1, \ldots, X_n)}{T_{i+}(X_1, \ldots, X_n)} \rightarrow \Pi_i(j), \quad \text{a.s.} \ P$$

- From Fortini et al. (2002), it is the limit of the sequence of empirical distributions of the successors:

$$\left( \frac{\sum_{k=1}^{n} \delta_{X_1,k}}{n}, \frac{\sum_{k=1}^{n} \delta_{X_2,k}}{n} \right) \rightarrow (\Pi_1, \Pi_2) \sim \mu_{1,2}(\cdot, \cdot)$$

- **Theorem** $(X_n)$ recurrent and Markov exchangeable. Then it is a mixture of Markov chains, where the mixing distribution $\mu$ is the law of the limit of the sequence of predictive distributions.
Aim: Construct a process \((Y_n)\), with \(Y_i \in I = \{1, \ldots, k\}\), \(k \leq \infty\), through a collection of Hoppe's urns.

We first consider Bayesian inference for Markov chains, then for hidden Markov models.
Reinforced Hoppe’s urns

state space $I = \{1, \ldots, k\}, k \leq \infty$.
For any $i$, associate a Hoppe urn $U_i$ with $\alpha$ black balls, and discrete color distribution $p_0$ on $I$. 
Reinforced Hoppe’s urns

state space $I = \{1, \ldots, k\}$, $k \leq \infty$.
For any $i$, associate a Hoppe urn $U_i$ with $\alpha$ black balls, and discrete color distribution $p_0$ on $I$.

Fix $Y_0 = y_0$, and pick a ball from urn $U_{y_0}$. Since it is black, a color is sampled from $p_0$ and added in the urn, together with the black ball.
If $y_1$ is the sampled color, we set $Y_1 = y_1$ and move to Hoppe’s urn $U_{y_1}$, and so on.
Reinforced Hoppe’s urns

state space $I = \{1, \ldots, k\}, \ k \leq \infty$.
For any $i$, associate a Hoppe urn $U_i$ with $\alpha$ black balls, and discrete color 
distribution $p_0$ on $I$.

Fix $Y_0 = y_0$, and pick a ball from urn $U_{y_0}$.
Since it is black, a color is sampled from $p_0$ and added in the urn, together 
with the black ball.
If $y_1$ is the sampled color, we set $Y_1 = y_1$ and move to Hoppe’s urn $U_{y_1}$, and 
so on.

Thus at time $n$

$$P_n(j) \equiv P(Y_{n+1} = j \mid y_1, \ldots, y_n = i) = \frac{\alpha_i p_{0,i}(j) + T_{i,j}(y_{1:n})}{\alpha_i + T_{i,+}(y_{1:n})}.$$ 

Call the process of colors $(Y_n)$ so defined a reinforced Hoppe’s urn process 
(Hoppe RUP).
The Hoppe RUP \((Y_n, n \geq 0)\) is not Markov. It is defined through the predictive rule.

We can prove that:

- \((Y_n)\) is Markov exchangeable.
The Hoppe RUP \((Y_n, n \geq 0)\) is not Markov. It is defined through the predictive rule.

We can prove that:

- \((Y_n)\) is Markov exchangeable.
- Infinitely many black balls are drawn, thus we sample an infinite sequence of colors \(\xi_i \sim p_0\).
The Hoppe RUP \((Y_n, n \geq 0)\) is not Markov. It is defined through the predictive rule.

We can prove that:

- \((Y_n)\) is Markov exchangeable.
- Infinitely many black balls are drawn, thus we sample an infinite sequence of colors \(\xi_i \overset{i.i.d.}{\sim} p_0\).
- \((Y_n)\) is recurrent.
The Hoppe RUP \((Y_n, n \geq 0)\) is not Markov. It is defined through the predictive rule.

We can prove that:

- \((Y_n)\) is Markov exchangeable.
- Infinitely many black balls are drawn, thus we sample an infinite sequence of colors \(\xi_i \overset{i.i.d.}{\sim} p_0\).
- \((Y_n)\) is recurrent.
- The draws from urn \(U_i\) are the successors of state \(i\), and they are independent through the different urns.

Each urn is visited infinitely often. By construction, the draws from urn \(U_i\) are sampled through a Hoppe scheme, thus they are exchangeable and their de Finetti measure is a \(DP(\alpha p_0)\).
Hoppe RUP

The Hoppe RUP \( (Y_n, n \geq 0) \) is not Markov. It is defined through the predictive rule.

We can prove that:

- \((Y_n)\) is Markov exchangeable.
- Infinitely many black balls are drawn, thus we sample an infinite sequence of colors \(\xi_i \sim p_0\).
- \((Y_n)\) is recurrent.
- The draws from urn \(U_i\) are the successors of state \(i\), and they are independent through the different urns.

Each urn is visited infinitely often. By construction, the draws from urn \(U_i\) are sampled through a Hoppe scheme, thus they are exchangeable and their de Finetti measure is a \(DP(\alpha p_0)\).

Thus we have proved

**Proposition.** The reinforced Hoppe RUP \( (Y_n) \) is conditionally Markov, given the transition matrix \(\Pi\), and the prior on \(\Pi\) is such that \(\Pi_i \sim DP(\alpha p_0)\).
Consider a hierarchical extension of the Hoppe RUP.

- **(1) Known colors.**
  Colors are drawn from an oracle Hoppe urn, with $\gamma$ black balls and discrete, known distribution $q$ on \{1, 2, \ldots\}.

- **(2) Unknown state space (HMM).**
  As before, colors are drawn from an oracle Hoppe urn, with $\gamma$ black balls. But now we have a diffuse color distribution $q$. 
Hierarchical Hoppe RUP - 1. known colors

If the color distribution of the oracle urn $q$ is discrete on $\{1, 2, \ldots, k\}, k \leq \infty$, only these colors can be drawn. Thus the state space of $(Y_n)$ is $\{1, 2, \ldots, k\}$. 
Hierarchical Hoppe RUP - 1. known colors

If the color distribution of the oracle urn \( q \) is discrete on \( \{1, 2, \ldots, k\}, k \leq \infty \), only these colors can be drawn. Thus the state space of \((Y_n)\) is \( \{1, 2, \ldots, k\} \).

**Lemma.** The oracle urn is visited infinitely many times, a.s..

Thus the colors \((\xi_i)\) from the oracle urn are an *infinite*, exchangeable sequence, thus \( \xi_i \mid p \sim^{i.i.d} p \), with \( p \sim DP(\gamma q) \).
Hierarchical Hoppe RUP - 1. known colors

If the color distribution of the oracle urn $q$ is discrete on $\{1, 2, \ldots, k\}$, $k \leq \infty$, only these colors can be drawn. Thus the state space of $(Y_n)$ is $\{1, 2, \ldots, k\}$.

**Lemma.** The oracle urn is visited infinitely many times, a.s..

Thus the colors $(\xi_i)$ from the oracle urn are an *infinite*, exchangeable sequence, thus $\xi_i \mid p \overset{i.i.d.}{\sim} p$, with $p \sim DP(\gamma q)$.

Therefore, conditionally on $p$, we are back to the previous case, and we have

$$(Y_n) \mid P, p \text{ is a Markov chain with state space } I \text{ and transition matrix } \Pi; \text{ the prior on } \Pi \text{ is such that the rows are conditionally independent, with a hierarchical DP prior, namely}$$

$$
\Pi_i \mid p \overset{\text{indep}}{\sim} DP(\alpha p) \\
p \sim DP(\gamma q)
$$
2. diffuse $q$ – infinite HMM

If $q$ is 

diffuse, each time we visit the oracle urn we create a new color

$\xi_i \sim q$.

Thus the state space of $(Y_n)$ is $\{\xi_1^*, \xi_2^*, \ldots\}$. We have $Y_1 = \xi_1^*$, then $Y_2$
can be of color $\xi_1^*$ or a new color $\xi_2^*$, etc: 

colors are created when needed!
2. diffuse $q$ – infinite HMM

If $q$ is diffuse, each time we visit the oracle urn we create a new color $\xi_i^* \sim q$.

Thus the state space of $(Y_n)$ is \{\(\xi_1^*, \xi_2^*, \ldots\)\}. We have $Y_1 = \xi_1^*$, then $Y_2$ can be of color $\xi_1^*$ or a new color $\xi_2^*$, etc: **colors are created when needed!**

But this is what we need as a prior for HMMs!

in HMMs, the Markov process is latent, thus also the state space \{\(\xi_1^*, \xi_2^*, \ldots\)\} is unknown.
Hierarchical Hoppe RUP with diffuse $q$

**Lemma.** The oracle urn is visited infinitely many times, a.s..

Thus the colors $(\xi_i)$ from the oracle urn are an infinite, exchangeable sequence, thus $\xi_i \mid p \overset{i.i.d.}{\sim} p$, with $p \sim DP(\gamma q)$. 
Hierarchical Hoppe RUP with diffuse $q$

**Lemma.** The oracle urn is visited infinitely many times, a.s..

Thus the colors $(\xi_i)$ from the oracle urn are an infinite, exchangeable sequence, thus $\xi_i \mid p \overset{i.i.d.}{\sim} p$, with $p \sim DP(\gamma q)$.

Therefore the above results hold conditionally on $p$, and we have

$$(Y_n) \mid \Pi, p$$ is a Markov chain with state space given by the support $\xi_1^*, \xi_2^*, \ldots$ of $p$, and transition matrix $\Pi$. The prior on $\Pi$ is such that the rows are conditionally independent, with a hierarchical DP prior:

$$
\pi_{\xi^*_j} \mid p = \sum_{j=1}^{\infty} w_j \delta_{\xi^*_j} \overset{indep}{\sim} DP(\alpha p) \\
p \sim DP(\gamma q)
$$
In a HMM \((Y_n, \theta_n)\), we can construct the latent sequence \((\theta_n)\) as a hierarchical Hoppe reinforced urn process.

The results show that \((\theta_n)\) is conditionally Markov, and the prior on the unknown transition matrix is a hierarchical Dirichlet process.

The predictive scheme used to construct the prior is usefully exploited for computations.
Further examples and developments

- **Sticky infinite HMMs**: each Hoppe urn $U_i$ has $\alpha$ black balls, and $M$ balls of color $i$, so we put more mass on $\pi_{i,i}$ (Fox et al.).

- The **Indian buffet process** (Griffiths and Ghahramani, 2006) is also based on an urn scheme.

- Other examples: **Hoppe RUP with random reinforcement**. The number $Y_n$ of balls added at time $n$ is random, with $Y_n$ independent on $(X_1, \ldots, X_{n-1})$.

- **Covariate-dependent transition matrix**
  Urns indexed by color and covariate value.
  $(X_n)$ is partially Markov exchangeable.

- mixture of Markov chain observed at random time points

- ...
summary

This was an overview of some predictive constructions, based on urn schemes, for characterizing mixtures of Markov chains. Hoppe RUPs and hierarchical Hoppe’s RUPs shed light on theoretical properties and connections between RUP, infinite HMM and hierarchical DP.

The predictive construction is exploited for efficient computational algorithms, with a huge number of applications....
This was an overview of some predictive constructions, based on urn schemes, for characterizing mixtures of Markov chains. Hoppe RUPs and hierarchical Hoppe’s RUPs shed light on theoretical properties and connections between RUP, infinite HMM and hierarchical DP.

The predictive construction is exploited for efficient computational algorithms, with a huge number of applications....

Developments:
Further nonparametric constructions for Markov exchangeable sequences (e.g. using extensions of Hoppe’s urn (Feng and Hoppe, 1998; two-parameters Poisson-Dirichlet) and extensions to more general exchangeability structures (row-column exchangeability, network data).
This was an overview of some predictive constructions, based on urn schemes, for characterizing mixtures of Markov chains. Hoppe RUPs and hierarchical Hoppe’s RUPs shed light on theoretical properties and connections between RUP, infinite HMM and hierarchical DP.

The predictive construction is exploited for efficient computational algorithms, with a huge number of applications....

Developments:
Further nonparametric constructions for Markov exchangeable sequences (e.g. using extensions of Hoppe’s urn (Feng and Hoppe, 1998; two-parameters Poisson-Dirichlet) and extensions to more general exchangeability structures (row-column exchangeability, network data)

thank you for your attention!


References


