UNIVERSITY OF CAMBRIDGE

An Infinite Product of Sparse Chinese Restaurant Processes

Yarin Gal • Tomoharu Iwata • Zoubin Ghahramani

yg279@cam.ac.uk

The Chinese restaurant process (CRP)

- Distribution over partitions of N natural numbers $\mathcal{P}_N$.
- Marginal distribution of the Dirichlet process.
- Following a Chinese restaurant metaphor –
  - Restaurant = partition $R \in \mathcal{P}_N$,
  - Table = partition block $T \in R$,
  - Customer sitting at a table = element in a block $i \in T$.
  - **Restaurant configuration** = sequence of block sizes $\mathbf{n} = (n_1, n_2, ..., n_K)$, in order of appearance

CRP – quick recap

UNIVERSITY OF CAMBRIDGE

Follows a recursive construction

- Given discount parameter $\mathbf{d} \in [0, 1)$ and concentration parameter $\mathbf{c} > -\mathbf{d}$,
- Given restaurant R^N with configuration $\mathbf{n} = (n_1, ..., n_K)$ with K tables and $\sum_{i=1}^{K} n_i = N$ customers,
- Probability of customer $N + 1$ to sit at an *existing table* T_i:

$$p(N + 1 \in T_i | R^N) = \frac{n_i - \mathbf{d}}{N + \mathbf{c}}$$

- Or at a *new table* T_{K+1}:

$$p(N + 1 \in T_{K+1} | R^N) = \frac{\mathbf{c} + K\mathbf{d}}{N + \mathbf{c}}.$$

An interesting question

We assumed discount $\mathbf{d} \in [0, 1)$ and concentration $\mathbf{c} > -\mathbf{d}$ with conditional probabilities

$$\frac{n_i - \mathbf{d}}{N + \mathbf{c}}, \frac{\mathbf{c} + K\mathbf{d}}{N + \mathbf{c}}.$$

What happens when...

- $\mathbf{c}$ is fixed and $\mathbf{d} \nearrow 1$?
- $\mathbf{d}$ is fixed and $\mathbf{c} \nearrow \infty$?
- $\mathbf{c}$ is fixed and $\mathbf{d} \searrow 0$?
- $\mathbf{d}$ is fixed and $\mathbf{c} \searrow -\mathbf{d}$?

We assumed discount $\mathbf{d} \in [0,1)$ and concentration $\mathbf{c} > -\mathbf{d}$ with conditional probabilities

$$\frac{n_i - \mathbf{d}}{N + \mathbf{c}}, \frac{\mathbf{c} + K\mathbf{d}}{N + \mathbf{c}}.$$

What happens when...

- ~~$\mathbf{c}$ is fixed and $\mathbf{d} \nearrow 1$?~~
- ~~$\mathbf{d}$ is fixed and $\mathbf{c} \nearrow \infty$?~~
- ~~$\mathbf{c}$ is fixed and $\mathbf{d} \searrow 0$?~~ $\leftarrow$ **one-parameter CRP**
- $\mathbf{d}$ is fixed and $\mathbf{c} \searrow -\mathbf{d}$? $\leftarrow$ **very interesting indeed**

UNIVERSITY OF CAMBRIDGE

Discount d is fixed and concentration c $\searrow -\mathbf{d}$.

- Let $\gamma > 0$ and $M \in \mathbb{N}$.
- Given restaurant R_M^N with configuration $(n_1, ..., n_K)$ (sum to N),
- Let R_M^{N+1} follow a CRP distribution with concentration parameter $-\mathbf{d} + \gamma/M$ and discount parameter $\mathbf{d} \in [0, 1)$.
- Probability of customer $N+1$ to sit at an *existing table* $T_i \in R_M^N$:

$$p(N+1 \in T_i | R_M^N) = \frac{n_i - \mathbf{d}}{N - \mathbf{d} + \gamma/M} \overset{M\to\infty}{\to} \frac{n_i - \mathbf{d}}{N - \mathbf{d}},$$

- Or at a *new table* T_{K+1}:

$$p(N+1 \in T_{K+1} | R_M^N) = \frac{-\mathbf{d} + \gamma/M + K\mathbf{d}}{N - \mathbf{d} + \gamma/M} \overset{M\to\infty}{\to} \frac{(K-1)\mathbf{d}}{N - \mathbf{d}}.$$

A CRP parametrisation

UNIVERSITY OF CAMBRIDGE

Discount d is fixed and concentration c $= -\mathbf{d} + \gamma/M$**. We have** $R_M \overset{M\to\infty}{\Rightarrow} R_\infty$**.**

- First customer sits at table T_1,
- Probability of second customer to sit at *existing table* $T_1 \in R^1_\infty$:

$$p(2 \in T_1 | R^1_\infty) = \frac{1 - \mathbf{d}}{1 - \mathbf{d}} = 1,$$

- Or at a *new table* T_2: $p(2 \in T_2 | R^1_\infty) = \frac{(1-1)\mathbf{d}}{1-\mathbf{d}} = 0$,
- For all $i \in \mathbb{N}$: $p(i \in T_1 | R^{i-1}_\infty) = \frac{i-1-\mathbf{d}}{i-1-\mathbf{d}} = 1$.
- A.s. resulting configuration with N customers all sitting at T_1:

$$(N)$$

— referred to as a **degenerate** configuration.

Not that interesting...

A sparse CRP (sCRP) parametrisation

But if we had M restaurants with concentration $\mathbf{c} = -\mathbf{d} + \gamma/M$...

Theorem (Sparse parametrisation of the CRP)

- *Let $\gamma > 0$ be a sparsity parameter and $m \leq M \in \mathbb{N}$.*
- *Let $R^N_{M,m}$ follow a CRP distribution with concentration $-\mathbf{d} + \gamma/M$ and discount $\mathbf{d} \in [0, 1)$ with N customers.*
- *Denote by $(\mathbf{n}_m)_{m=1}^M$ the random sequence of configurations of the restaurants $(R^N_{M,m})_{m=1}^M$.*

The following holds true as $M \to \infty$:

- *the expected count of* ***degenerate configurations*** *(N) is* ***unbounded****,*
- *the expected count of* ***non-degenerate configurations*** *is given by $\gamma \mathbf{H_d}(\mathbf{N-1})$ with $H_{\mathbf{d}}(N) = \sum_{i=1}^{N} \frac{1}{i-d}$.*

A sparse CRP (sCRP) parametrisation

So, in the infinite sequence of sparse CRPs (sCRPs)...

- almost all restaurants have all customers sitting next to a single table,
- but finitely many restaurants have more than a single table.

In fact,

- These have exchangeable partition probability function (EPPF):

$$p(n_1, n_2, ..., n_K) = \frac{\prod_{i=1}^{K} \left(\prod_{j=1}^{n_i - 1} (j - d) \right) \prod_{i=1}^{K-2} (id)}{\prod_{i=1}^{N-1} (i - d)}.$$

- We can analytically write the probability of a sequence of non-degenerate configurations —

Infinite product of sparse CRPs

For an infinite product of sparse CRPs the probability of configurations $\{\mathbf{n}_m\}_{m=1}^{K^+}$ follows,

Theorem (Infinite product of sparse CRPs)

The probability of K^+ non-degenerate configurations $\{(n_1^m, ..., n_{k^m}^m) | m \in [K^+]\}$ is given by

$$\frac{\gamma^{K^+}}{K^+!} e^{-\gamma H_{\mathbf{d}}(N-1)} \prod_{m=1}^{K^+} \left(\frac{\prod_{i=1}^{k^m} \left(\prod_{j=1}^{n_i^m - 1} (j - d) \right) \prod_{i=1}^{k^m - 2} (id)}{\prod_{i=1}^{N-1} (i - d)} \right)$$

- *Obtained by setting $\mathbf{c} = -\mathbf{d} + \frac{\gamma}{M}$ in a product of M sparse CRPs with $M \to \infty$.*

Infinite product of sparse CRPs

- We can re-write this equation as a product of Poisson densities and the EPPFs above...

$$\underbrace{\frac{\gamma^{K^+}}{K^+!} e^{-\gamma H_{\mathbf{d}}(N-1)}}_{\text{Poissons}} \underbrace{\prod_{m=1}^{K^+} \left(\frac{\prod_{i=1}^{k^m} \left(\prod_{j=1}^{n_i^m - 1} (j-d) \right) \prod_{i=1}^{k^m - 2} (id)}{\prod_{i=1}^{N-1} (i-d)} \right)}_{\text{EPPFs}}$$

- And obtain a recursive construction.
- Instead of extending the Chinese restaurant metaphor, we use an urn scheme:

UNIVERSITY OF CAMBRIDGE

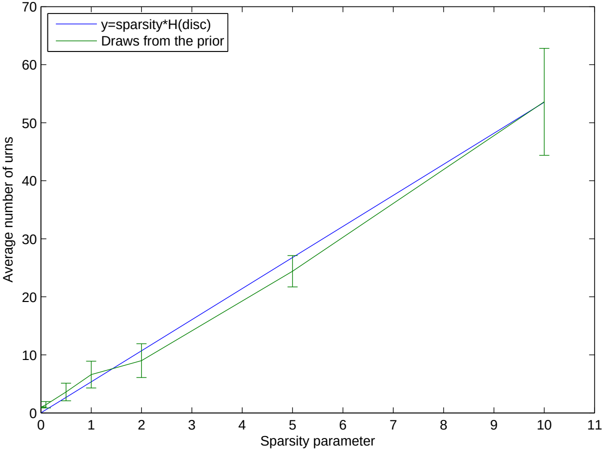
Time permits...

Definition (Urn scheme)

- *Create* $Poi(\frac{\gamma}{1-d})$ *urns; for each urn*
  - *add two balls with distinct colours.*
- *At the l'th step* ($l \geq 3$)
  - *For each existing urn with balls of* k *colours,* n_i *balls of colour* i
    - *select colour* i *with probability* $\frac{n_i - d}{l-1-d}$ *and add another ball of the same colour,*
    - *or add a ball of a new colour with probability* $\frac{(k-1)d}{l-1-d}$.
  - *Create* $Poi(\frac{\gamma}{l-1-d})$ *new urns; for each urn*
    - *add a ball of a new colour,*
    - *add* $l-1$ *balls of a distinct colour.*

This suggests an MCMC scheme.

Infinite product of sCRPs – properties

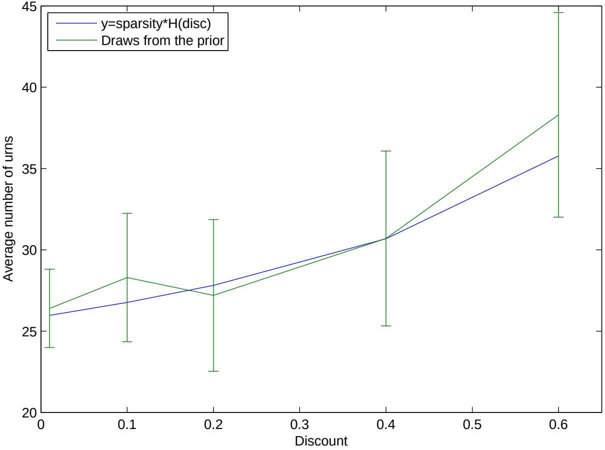


Expected number of urns as a function of the sparsity parameter γ

$$y = \gamma H_{\mathbf{d}}(N-1)$$

with $\mathbf{d} = 0.1$ and $N = 100$.

Infinite product of sCRPs – properties

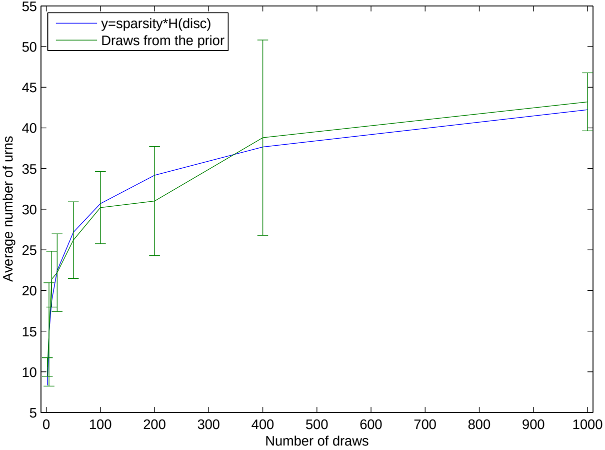


Expected number of urns as a function of the discount parameter **d**

$$y = \gamma H_{\mathbf{d}}(N-1)$$

with $\gamma = 5$ and $N = 100$.

Infinite product of sCRPs – properties

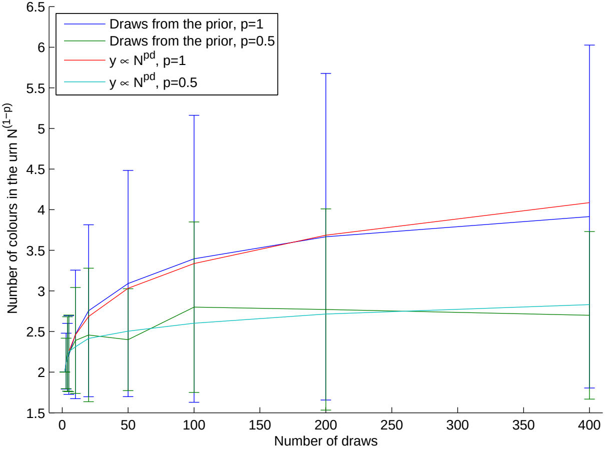


Expected number of urns as a function of the number of draws N

$$y = \gamma H_{\mathbf{d}}(N-1)$$

with $\gamma = 5$ and $\mathbf{d} = 0.4$.

Infinite product of sCRPs – properties



Expected number of colours in the first urn ($p = 1$, blue) and in the $N^{0.5}$'th urn ($p = 0.5$, green) as a function of the number of draws N

$$y \propto N^{p\mathbf{d}}.$$

Relation to existing research

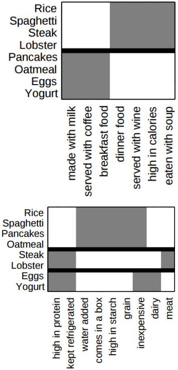
- The process induces a **categorical feature allocation**
  - Restaurants = features
  - Tables = feature values
  - i'th customer in each restaurant = data point i
  - i'th customer sitting next to table T in restaurant R = data point i takes value T for feature R
- Relates to existing literature in the field:
  - IBP (Griffiths and Ghahramani, 2011) – binary feature allocation
  - Continuum of urns (Roy, 2014) – a product of urns with balls of two colours characterising the IBP

Finally - future research

What can we do with this process?

- Multi-view clustering; principled extension to –
  - Cross-Cat (Mansinghka, Jonas, Petschulat, Cronin, Shafto, Tenenbaum, 2009),
  - Infinite latent attribute model (Palla, Knowles, Ghahramani, 2012),
  - etc.
- Bayesian representation learning
  - Features as data representation
- Currently running experiments!

Cross-Cat:



Thank you