Weighted Chinese restaurant processes

Albert Y. Lo[1]

Department of Information and Systems Management

The University of Science and Technology

Clear Water Bay, Hong Kong

1 Introduction

The Chinese restaurant process (CR) takes its name from a seating procedure observed in a Bay Area Chinese restaurant by L. Dubins and J. Pitman. Imagine customers 1,...,n arrive at a restaurant one after the other. The first customer, labelled one, is seated at an unoccupied table. When the second customer arrives, with some probability he is seated at the table with customer one; otherwise a new table is set up for him. When customer three arrives, he may be seated with customer number one, or with customer number two (or with both, if customer two was seated with customer one), or a third table may be opened for him. The process continues until n customers have been seated. Note that all customers, except for the first one to arrive, are seated randomly.

The seating probability of a CR is defined as: Suppose $j-1$ customers are seated, customer j is seated at an empty table with probability proportional to a constant $c>0$; otherwise the customer is seated at an occupied table with probability proportional to the number of individuals at that table.

The Chinese restaurant process just described is sequential; that is, customers are seated one at a time as they arrive, and once they sit at a table, they remain there. One can also define non–sequential processes, in which customers are moved randomly from table to table. In either case, when the seating probabilities depend on a set of sample data (one observation for each individual in the restaurant), the resulting random partition is called a weighted Chinese restaurant process (WCR).

The weighted Chinese restaurant process can be used as basis of Monte Carlo methods to approximate posterior quantities for Bayesian statistical problems in density and hazard rate estimations, deconvolution problems, image analysis, cluster problems, and Bayesian semi–parametric problems. For such problems, and for priors driven by gamma or Dirichlet processes, posterior quantities are averages of partitions of the form

(1) $$\Sigma_{\mathbf{p}} h(\mathbf{p})\pi(\mathbf{p}),$$

[1]Supported in part by Hong Kong RGC Competitive Earmarked Research Grant 6189/98E

where $\mathbf{p}=\{C_1,...,n(\mathbf{p})\}$ is a partition of the set $\{1,...,n\}$, $h(\mathbf{p})$ is a prescribed function depending on the parameter of interest, and

(2) $$\pi(\mathbf{p}) \propto \Pi_{1\le j\le n(\mathbf{p})} g(C_i)$$

is a probability distribution of $\mathbf{p}$, where g(.) is a known function defined on finite subsets of integers. The Monte Carlo method suggests to sample $\mathbf{p}_1,...,\mathbf{p}_M$ from a distribution $q(\mathbf{p})$, and to use

(3) $$\Sigma_{1\le k\le M} h(\mathbf{p}_k)\pi(\mathbf{p}_k)/q(\mathbf{p}_k)$$

to approximate (1).

In order to have an accurate Monte Carlo sample $\mathbf{p}_1,...,\mathbf{p}_M$, it is preferable that $q(\mathbf{p})$ be proportional to $\pi(\mathbf{p})$; ideally, $\mathbf{p}_1,...,\mathbf{p}_M$ are iid from $\pi(\mathbf{p})$ (i.e., exact sampling). But for a $\pi(\mathbf{p})$ of the form (2), such an exact sampling does not seem to be known. There are two alternatives:

(I) a Gibbs sampler:

$\mathbf{p}_1,...,\mathbf{p}_M$ is a Markov chain sample; the stationary distribution of the chain is $\pi(\mathbf{p})$;

(II) an sequential importance sampler:

$\mathbf{p}_1,...,\mathbf{p}_M$ is an iid sample from $q(\mathbf{p})$, where $q(\mathbf{p})$ is proportional to $\pi(\mathbf{p})$ up to a normalization constant.

Seating probability of the $\pi(\mathbf{p})$ plays a key role in defining both (i) and (ii).

2 Seating probability of a WCR

This section discusses seating probabilities defining various WCR distributions of the form (2). Let $\mathbf{p}$ be a partition of $\{1,...,n\}$. Delete an integer r (r=1,...,n) in $\mathbf{p}$ to form $\tilde{\mathbf{p}}=\{\tilde{C}_i,\ i=1,...,n(\tilde{\mathbf{p}})\}$ which partitions $\{1,...,r-1,r+1,...,n\}$; $\tilde{C}_0$ denotes an empty table of $\tilde{\mathbf{p}}$. The partitions $\mathbf{p}$ and $\tilde{\mathbf{p}}$ differ at only one table. The joint distribution of $(\tilde{\mathbf{p}},\mathbf{p})$ is identified through the conditional (i.e., predictive) distribution of $\mathbf{p}|\tilde{\mathbf{p}}$ and the marginal distribution of $\tilde{\mathbf{p}}$.

Lemma 3.1 Assume a distribution on the partition $\mathbf{p} \sim \pi(\mathbf{p}) \propto \Pi_{1\le i\le n(\mathbf{p})} g(C_i)$, and remove an integer r (r=1,...,n) in $\mathbf{p}$ to form $\tilde{\mathbf{p}}=\{\tilde{C}_i,\ i=1,...,n(\tilde{\mathbf{p}})\}$.

(i) Given $\tilde{\mathbf{p}}$, the (conditional) probability that r sits at table $\tilde{C}_s$ is

$$\pi_s(\mathbf{p}|\tilde{\mathbf{p}}) \propto g(\{r\}+\tilde{C}_s)/g(\tilde{C}_s) \text{ for } s=0,...,n(\tilde{\mathbf{p}});\ g(\tilde{C}_0)\equiv 1.$$

(ii) $$\pi(\tilde{\mathbf{p}}) \propto \Pi_{1\le i\le n(\tilde{\mathbf{p}})} g(\tilde{C}_i).$$

Proof Define $\omega(\mathbf{p}) = \Pi_{1\le i\le n(\mathbf{p})} g(C_i)$; define $\omega(\tilde{\mathbf{p}})$ similarly. Note that $\mathbf{p}$ and $\tilde{\mathbf{p}}$ differ only at the table containing r, say, table $C_s=\{r\}+\tilde{C}_s$.

(a) If $C_s=\{r\}+\tilde{C}_s$ and $\tilde{C}_s$ is not empty, $n(\mathbf{p})=n(\tilde{\mathbf{p}})$ and

$$\omega(\mathbf{p}) = \Pi_{1\le i\le n(\mathbf{p})} g(C_i)$$
$$= [\Pi_{1\le i\le n(\mathbf{p}),i\ne s}\ g(C_i)] \times g(C_s)$$
$$= [\Pi_{1\le i\le n(\tilde{\mathbf{p}}),i\ne s}\ g(\tilde{C}_s)] \times g(\{r\}\cup\tilde{C}_s)$$
$$= [\Pi_{1\le i\le n(\tilde{\mathbf{p}})}\ g(\tilde{C}_s)] \times g(\{r\}\cup\tilde{C}_s)/g(\tilde{C}_s)$$
$$= \omega(\tilde{\mathbf{p}}) \times g(\{r\}\cup\tilde{C}_s)/g(\tilde{C}_s).$$

(b) If $C_s=\{r\}+\tilde{C}_0$, i.e., r sits on an empty table of $\tilde{\mathbf{p}}$, $\omega(\mathbf{p}) = \omega(\tilde{\mathbf{p}}) \times g(\{r\})$.

(i) follows. (ii) also follows since given the joint distribution, the marginal distribution is unique. ||

Remarks

(i) Lemma 3.1(ii) implies that WCRs of the form (for n=1,2,..., and with an abuse of notation) $\pi(\mathbf{p}) \propto \Pi_{1\le i\le n(\mathbf{p})} g(C_i)$ defines a "consistent" family of finite dimensional distributions on the space of partitions.

(ii) The seating probability [Lemma 3.1(i)] also defines a sequential seating procedure which seats customers {1,...., n} IN THE ORDER WRITTEN. The product rule of probability dictates that the resulting distribution on $\mathbf{p}$ has an numerator identical to $\Pi_{1\le i\le n(\mathbf{p})} g(C_i)$. However, each step of the sequential seating requires a normalization constant $1/\Sigma_{1\le i\le n(\tilde{\mathbf{p}})}\ g(\{r\}+\tilde{C}_s)/g(\tilde{C}_s)$ for the seating probability, and the denominator is a product of these renormalization constants. Numerical examples seems to suggest that both the Gibbs and the importance samplers work well.

Example 3.1 If $\pi(\mathbf{p})$ is a discrete uniform distribution, $\pi_s(\mathbf{p}|\tilde{\mathbf{p}})=1/[n(\mathbf{p})+1]$, $s=0,\dots,n(\tilde{\mathbf{p}})$.

Example 3.2 If $\pi(\mathbf{p}) \propto c^{n(\mathbf{P})}\Pi_{1\le i\le n(\mathbf{p})}[(e_i-1)!]$, i.e., the CR distribution with parameter $c>0$, $\pi_s(\mathbf{p}|\tilde{\mathbf{p}}) \propto \tilde{e}_s$, $s=0,\dots,n(\tilde{\mathbf{p}})$; $c=\tilde{e}_0$. To see this, set

$$g(C) = c \times \text{factorial of } [(\#\text{ of elements in } C) -1],$$

and $g(\tilde{C}_0)\equiv c$ in Lemma 3.1. The CR distribution is also called the Polya urn distribution.

Example 3.3 Let $x_1,\dots,x_n$ have a joint density $f(x_j;\ j\in\{1,\dots,n\})$.{Bayesian mixture models coresponds to the case that $f(x_j;\ j\in\{1,\dots,n\})$ is an exchangeable density.} The data x_i's are regarded as given and fixed (as in Bayesian and bootstrap methods). Define a WCR on partitions $\pi(\mathbf{p}) \propto \Pi_{1\le i\le n(\mathbf{p})} f(x_j;\ j\in C_i)$. It follows then

$$\pi_s(\mathbf{p}|\tilde{\mathbf{p}}) \propto f(x_j;\ j\in\{r\}+\tilde{C}_s)/f(x_j;\ j\in\{r\}+\tilde{C}_s) \equiv m(x_r|\ x_j;\ j\in\tilde{C}_s).$$

That is, the seating probability of "r" to the table $\tilde{C}_s$ is proportional to the value of the predictive density of x_r given $\{x_j;\ j\in\tilde{C}_s\}$.

Articles in the following reference section give more examples.

Reference

Aldous, D.J. (1985). Exchangeability and Related Topics. Lecture Notes in Mathematics. 1117 Springer–Verlag.

Brunner, L.J. and and Lo, A.Y. (2000). Bayesian classification. Research report. Research Report, Department of Statistics, University of Toronto, Canada

Blackwell, D. and MacQueen, J. B. (1973). Ferguson distributions via Pólya urn schemes. *Annals of Statistics* **1**, 353–355.

Ferguson, T.S. (1973). A Bayesian analysis of some nonparametric problems. *Annals of Statistics*, **1**, 209–230.

Ishwaran, H. and James, L.F. (2000). Generalized weighted Chinese restaurant processes for species sampling mixture models. Preprint.

Kuo, L. (1986). Computations of mixtures of Dirichlet processes. *SIAM J. Sci. Statist. Comput.,* **7**, 60–71.

Lo, A.Y. (1984). On a class of Bayesian nonparametric estimates: I. Density estimates. *The Annals of Statistics,* **12**, 351–357.

Lo, A.Y., Brunner, L.J. and Chan, A.T. (1996) Weighted Chinese restaurant processes and Bayesian mixture models. Research Report (Revision 1998). Department of Information and Systems Management, University of Science and Technology, Hong Kong.

Lo, A.Y. and Weng, C.S. (1989). On a class of Bayesian nonparametric estimates: II. Hazard rate estimates. *Ann. Instit. Statist. Math.,* **41**, 227–245.

MacEachern, S.N. (1994). Estimating normal means with a conjugate style Dirichlet process prior. *Communications in Statistics.–Simulation,* **23**, 727–741.

MacEachern, S.N., Clyde, M and Liu, J.S. (1999). Sequential importance sampling for nonparametric Bayes models: the next generation, *Canadian Journal of Statistics* **27**, 251–267.

Propp, J.G. and Wilson, D.B. (1996). Exact sampling with coupled Markov chains and applications to statistical mechanics. *Random Structures and Algorithms* **9**, 223–252.