**Summary**

• We present a new family of Bayesian hierarchical models based on the nested Chinese restaurant process, and show that every completely exchangeable hierarchical model can be represented as a member of this family.

• We do this by giving a criterion (the martingale criterion) that allows substantial generalization of the nested Chinese restaurant process beyond topical models.

• Using this criterion, we construct infinitely deep hierarchical Dirichlet and beta processes.

• Our construction circumvents issues present in the true-structured stick-breaking model.

**Motivation**

• Priors over tree structures are crucial for performing Bayesian hierarchical modeling.

• To date, all proposals for priors over discrete trees have undesirable properties.

• True-structured stick-breaking has a constant depth under the prior.

• Nested Chinese restaurant processes are hard to extend beyond topical models.

• Hierarchical diffusion trees are designed for continuous, non-discrete data.

• To flexibly learn the structure of models such as hierarchical Dirichlet and beta processes, we need something better.

• One solution: build machinery to extend the nCRP to these models.

**Example: An Infinite Random Walk**

• Suppose that each node v contains a real number xv, and that for a child cv of v, the distribution for xc given xv is N(0, cv).

• Therefore, this model is not well-defined.

**Example: An Infinite Hierarchical Dirichlet Process**

• Suppose that each node v contains a probability vector µv over 0-1 vertices (φv), and that for a child cv of v, the distribution for µc given µv is Dirichlet(αv, βv, φv).

• Then the marginal distribution for x given its path depends only on the latent parameters along the path.

**The Martingale Criterion**

• For both the random walk and the hierarchical Dirichlet process, we have E[θv | θv-1] = θv-1, where θ is the collection of parameters at node v.

• This condition is called the martingale criterion with general, ask that E[θv-1 | θv-2] = θv-2 for some f.

• Theorem (Doob): All non-negative martingales sequences have a limit with probability 1.

• Corollary: The infinite HDP converges.

• Universality

• A hierarchical model is completely exchangeable if, for a node c with parent v, the distribution for xc depends only on c and the depth of c in the tree.

• Theorem: for any completely exchangeable hierarchical model, there exists an alternate set of latent parameters θ such that E[θv | θv-1] = θv.

• Therefore, every completely exchangeable model can be realized using our construction.

• But the reparameterization in terms of θ might be inconvenient computationally.

**Comparison to Tree-Structured Stick Breaking**

• The main alternative proposal for Bayesian hierarchies is tree-structured stick-breaking.

• To demonstrate the desirability of our construction, we perform an empirical comparison of the nCRP and TSSB.

• A theoretical analysis is given in the paper.

• Comparison 1: depth of the tree as a function of data size.

• Note that the depth of the nCRP grows with the data, but the depth of TSSB does not.

• Comparison 2: samples from the prior for (Data=100).