

# Discussion

Peter Orbanz

Cambridge University and Columbia University

## Is generalization possible?

- ▶ Exchangeability is not just a simplifying assumption
- ▶ Prediction without assumptions is not possible (Hume's induction problem)

## In context: De Finetti's theorem

- ▶ Prediction is possible if underlying pattern does not change over time

## Non-exchangeable data

- ▶ Requires careful justification of what prediction/generalization means
- ▶ Substitute sequential exchangeability by other form of permutation invariance/symmetry

## Again: De Finetti

$$\mathbb{P}(X_1 \in A_1, X_2 \in A_2, \dots) = \int_{\mathcal{P}} \left( \prod_{i=1}^{\infty} P(X_i \in A_i) \right) Q(dP)$$

Idea (Blackwell, Ferguson, ...):

- ▶ Holds for *some* random probability measure  $P$
- ▶ Model distribution  $Q$  of  $P$  directly, e.g.  $Q =$  Dirichlet process

## In the 1990s

Universal inference idea:

- ▶ Define prior with “large support”
- ▶ Combined with Gibbs sampler: Universal inference engine

D. Draper, discussion of Walker & Damien & Laud & Smith, JRSS B, 1999

## Today

- ▶ Express  $P$  on some parameter space (functions, partitions, binary matrices, ...)

# WHAT DO WE KNOW ABOUT OUR MODELS?

## Consistency and Convergence

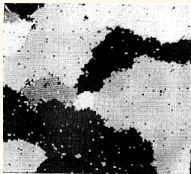
Machine learning:

- ▶ Emphasizes different models than mathematical statistics
- ▶ Machine learning-specific problems have not been addressed
- ▶ E.g., are models consistent w.r.t. latent variables?

## Gibbs sampling

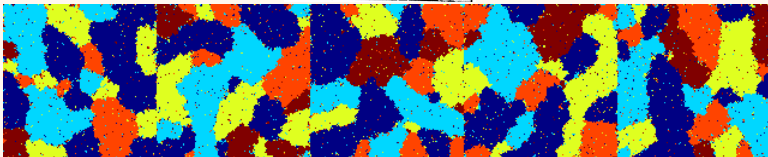
- ▶ Hardly any systematic work
- ▶ How many models have we actually seen samples from?
- ▶ Hierarchies drastically blow up state space

# Geman & Geman, 1984



128 x128 grid  
8 nearest neighbor edges  
K = 5 states  
Potts potentials:  $\beta = 2/3$

200 Iterations



10,000 Iterations

