Modern Bayesian Inference: Models, Algorithms, and Applications

Bahman Moraffah

Electrical, Computer, and Energy Engineering

Arizona State University

https://bmoraffa.github.io/presentation

Bayesian Modeling and Inference

- Bayesian Nonparametrics (BNP)
    - Bayesian nonparametric modeling for generalized time-varying dynamic systems
        - BNP modeling of generalized time-varying dynamic systems and multiple object tracking
        - Tree-based structure in MOT and quantifying uncertainty about the tree structure
        - BNP modeling for dependent measurements
        - BNP modeling for measurement estimation in high clutter
    - Bayesian nonparametric modeling of reinforcement learning
        - Spectrum sharing and coexistence through Bayesian RL
    - Feature allocations and Bayesian nonparametric modeling in health data
    - Sparse dynamic multigraphs
        - Bayesian edge-exchangeable models for sparse graph and hierarchical model
        - Evolving dynamic BNP model for interaction network to reinforce recent patterns
        - Exchangeable random measures for sparse graphs with overlapping communities
- Approximation Inference
    - Frequentist properties of variational inference
    - Variational Bayes and its advances in signal processing
    - Scalable inference
        - Gradient methods
        - Markov chain Monte Carlo methods (MCMC) e.g., Hamiltonian MC

Non-Bayesian Approaches

- Nonparametric Modeling
  - Predicting bradycardia events in infants
- Probabilistic Graphical Model
  - Estimating high-dimensional graph networks: Forest Bagging method
- High-dimensional Information theory
  - Interactive source coding
- Causal Inference
  - Causal adversarial network to learn observational and interventional distribution

Nonparametric Bayes View:

- Bayesian Statistics:
  - Probabilistic modeling to express all forms of uncertainty and noise
  - … then *inverse probability* rule (i.e. Bayes' Theorem) allows us to infer unknown quantities, learn from data, and make predictions
    - Bayes' theorem:

$$Q(d\theta|X=x) = \frac{dP(X \in \cdot|\theta)}{dP(x \in \cdot)} Q(d\theta)$$

- Bayesian statistics that is not parametric (wait!)
- Bayesian nonparametrics (i.e. not finite parameter, unbounded/ growing/infinite number of parameters)
  - BNP models do not generally satisfy Bayes' theorem since the density cannot exist for all x (undominated models) (not the same as posterior tractability!)
  - Random discrete measures are often undominated.

Bayesian Nonparametric Model

- Why Bayesian nonparamterics?
  - Bayesian : Simplicity (of the framework)
  - Nonparametric : Complexity (of the real world phenomena)
- Definition: A BNP model is a Bayesian Model with an infinite-dimension parameter space and assumes that data distribution cannot be represented in terms of finite et of parameters.

Modeling Goal	Example process
Distribution on functions Distribution on distributions Clustering Hierarchical clustering Distribution on measures …	Gaussian processes Dirichlet processes PY processes CRP / Polya Urn Dirichlet diffusion tree Kingman's coalescent Completely random measure

Time-Varying Dynamic Systems

- **Generalized time-varying dynamic systems** (multiple object tracking)
  - Each object may leave or stay in the FOV with a time-dependent probability
  - Survived objects may transition to the next time with some probability kernel function
  - New objects can join the scene

previous time step *(k-1)*

time step *k*

- Objective: Jointly estimate the cardinality, path, location, and characteristics of objects from sensor data.

[B. Moraffah 2018, 2019]

Main Challenges

- Track unknown time-varying number of objects (cardinality)
- Robustly associate objects at each time
- Objects leave, enter or stay in scene: unknown state label/identity
- Multiple observations from sensing modalities, possibly dependent, and unknown measurement-to-object associations
- Multiple environmental conditions: high noise levels, clutter, interference (uncertainty on parameters)

A simple Nonparametric model such as DP cannot capture all these aspects.

⬇

Dependent Bayesian Nonparametrics

Main Results: DDP Prior

- Introduce a dependent Dirichlet process (DDP) prior to capture Survival, Birth, and Death:

$$G(d\theta_k) = \sum_{j=1}^{\infty} \pi_j(k)\delta_j(d\theta_k)$$

  - Capture full dependency (this introduces a collection of random distributions that are related but not identical)
  - Dependent model to update object cardinality and posterior distribution
  - Adjust the probabilities among the new and survived objects
  - Introduce a constructive way of building the model and provide simple inference algorithms through MCMC and VI
- Achieve high estimation accuracy and lower computational cost at low SNR
- Integrate this model with a hierarchical Bayesian setup to account for multi-sensor measurements

[B. Moraffah 2018, 2019]

Main Theorems and Properties

Theorem 1: Assuming the space of parameters Polish, conditional distribution of DDP at time k given the configurations of DDP at time (k-1) follows a Dirichlet process.

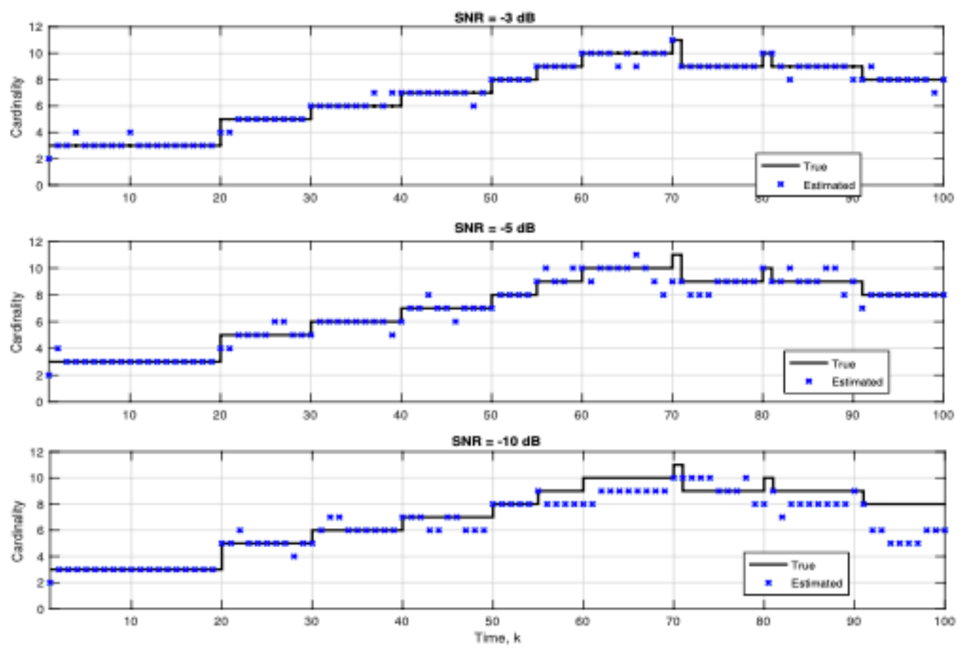
Theorem 2: Given configurations at time (k-1), the induced random partition at time k in infinite exchangeable random partition (EPPF depends only on the size of partitions).

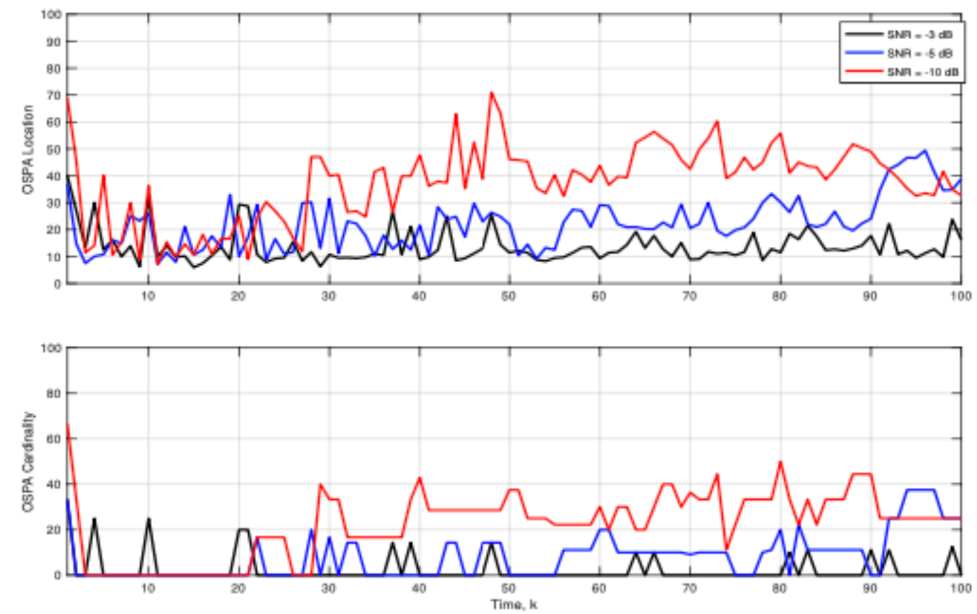
Theorem 3: Posterior distribution is strongly (weekly) consistent.

Theorem 4: Under some regularity conditions, the contraction rate is $\mathcal{O}(n^{-\frac{\beta}{2\beta+d}})$. Therefore, it matches the minimax rate.

[B. Moraffah 2019]

Experimental Result (I)

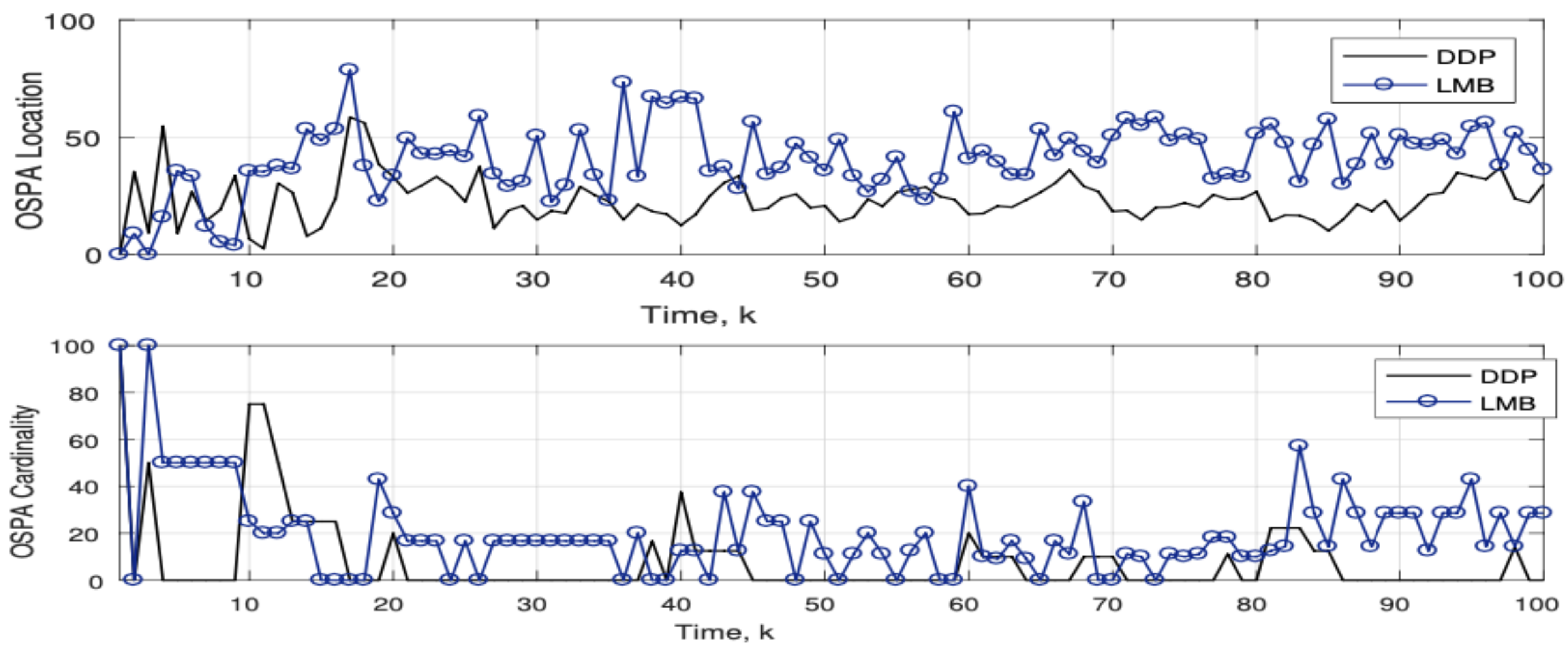




DDP prior for multiple object tracking under various SNRs:
SNR = -3dB, SNR = -5dB, and SNR = -10 dB

[B. Moraffah 2018, 2019]

Experimental Result (II)



Comparison between dependent Bayesian nonparametric modeling and generalized labeled multi-Bernoulli

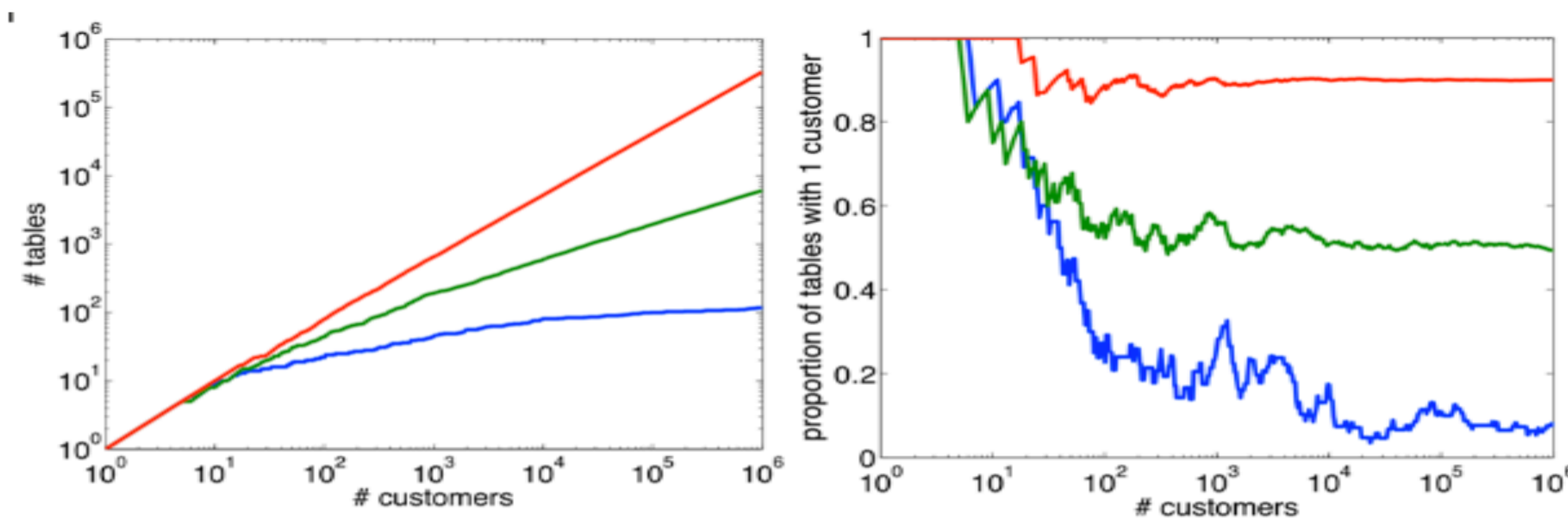
[B. Moraffah 2018, 2019]

Pitfall of this model for MOT

Objects benefit from a larger number of available partitions (clusters) to capture full dependency

⬇

More plausible to have less popular partitions (clusters) → Two-Parameter Poisson-Dirichlet Process



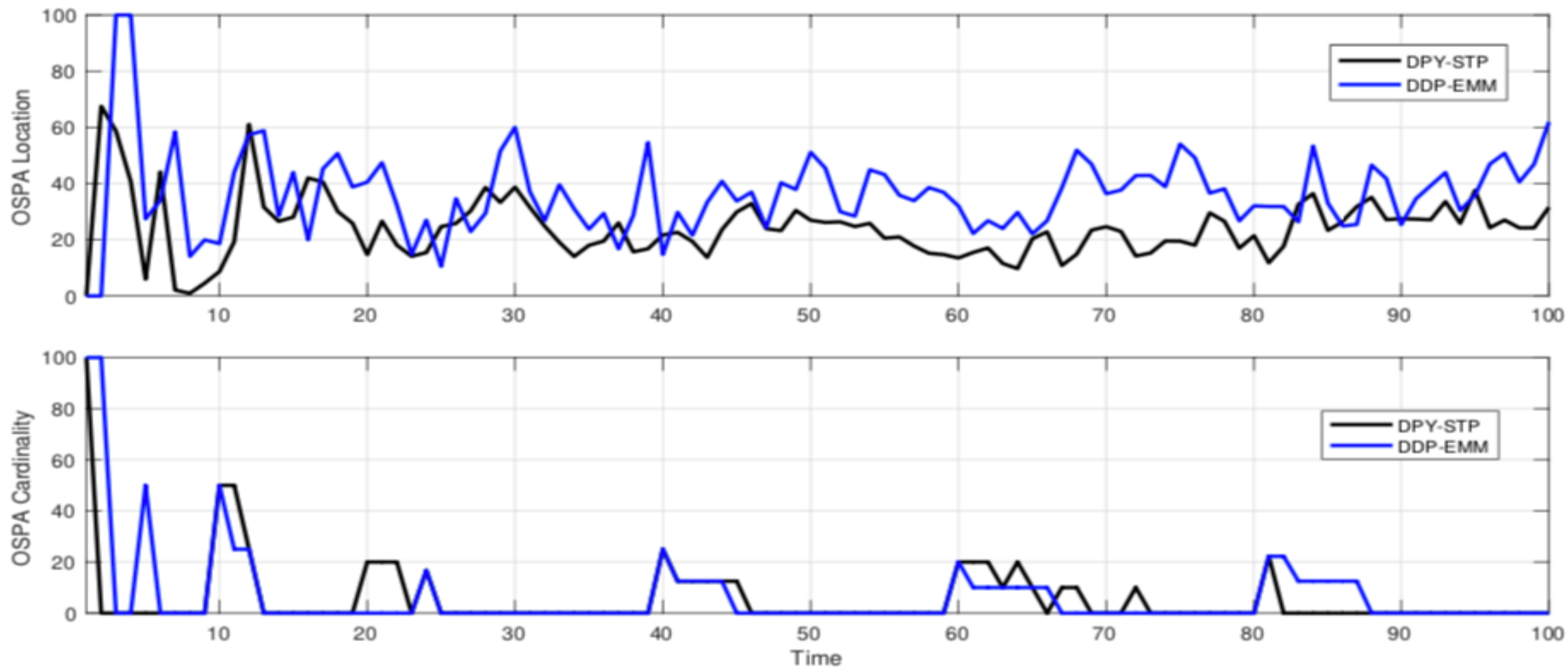
$\alpha = 10$ and for $d = 0.9$, $d = 0.5$, and $d = 0$

Main Results: DPY Prior

- Introduce a dependent two-parameter Poisson-Dirichlet prior to capture Survival, Birth, and Death
  - This model provides the same flexibility as DDP prior modeling and more
  - There is a constructive way of building the model and its relationship to the Pitman-Yor Process.
  - Provide simple inference algorithms through MCMC and VI
- Achieve high estimation accuracy and lower computational cost at low SNR
- Integrate this model with a hierarchical Bayesian setup to account for multi-sensor measurements

[B. Moraffah 2019]

Experimental Results



OSPA Comparison between DPY prior based model and DDP prior based model

[B. Moraffah 2018, 2019]

Model Comparison

- Both DDP prior and DPY prior result in induced dependent partitions (clusters) that are infinitely exchangeable.
- Simple inference algorithms for both methods.
- The expected number of clusters is

	DP	PY
Number of unique cluster in N points	$\mathcal{O}(\alpha \log N)$	$\mathcal{O}(\alpha N^b)$

- Posterior distribution under DDP prior is consistent under mild condition where posterior distribution under DPY prior is only consistent under a very restrictive conditions (They fall into the category of Gibbs prior and their consistency follows that.).
- They both work well in practice!

Research Roadmap

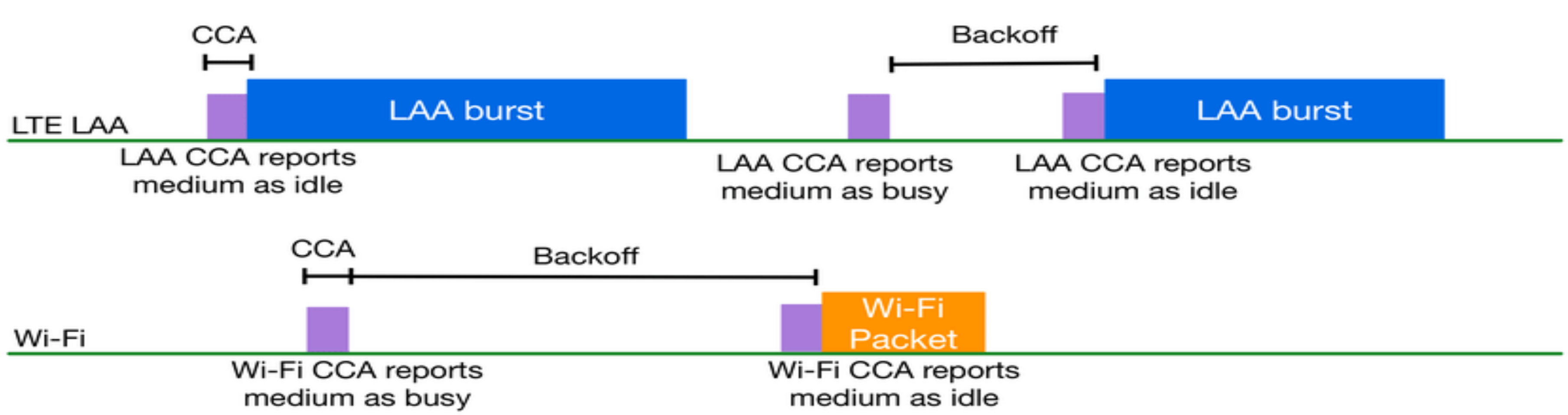
- Approximation inference
- Bayesian reinforcement learning and its application in signal processing
- Bayesian causal inference
- Bayesian nonparametric modeling of health data

Approximation Inference

- Variational Inference
  - Frequentist properties of variational models, e.g. consistency, contraction rate, asymptotic normality [B. Moraffah 2020, Y. Wang 2017]
  - Applications to Bayesian Stochastic Block Model [Wang 1987, Abbe 2015]
  - Stochastic variational inference in time-series and non-conjugate inference [Willsky 2014]
  - Advances in VI models in Bayesian deep learning such as doubly stochastic variational inference [Salimbeni 2017]
- Gradient-based MCMC
  - Hamiltonian Monte Carlo
    - Geometric ergodicity of dynamic HMC
      - ❑ Can one identify obstructions to geometric ergodicity for dynamic implementations?
    - Improving HMC for multimodal distributions
    - Generalizing HMC to non-smooth topological spaces
      - ❑ Can one identify group constructions that provide the foundations for generalizing Hamiltonian Monte Carlo to novel spaces?
      - ❑ Can we generalize HMC for discrete distributions?
  - Langevin Dynamics
    - Convergence of the algorithms for non-convex optimization
    - SDE in stochastic gradient MCMC to characterize its property and applications in DL

Bayesian Reinforcement Learning

- Model spectrum sharing between LTE-LAA and Wi-Fi networks as decentralized partially observable Markov decision process for adopting reinforcement learning
- Optimize sharing policies, including channel selection, back-off counter, and channel occupancy time, for LTE-LAA and Wi-Fi nodes in dynamic spectrum environment using reinforcement learning to minimize interference
- Adopt Bayesian approach on both policy structure and parameters to encode prior knowledge for facilitating policy-based reinforcement learning
- Introduce a nonparametric model over policy prior to incorporate unbounded number of potential wireless nodes and heterogeneous policy requirements
- Utilize variational inference to approximate the sophisticated and large posterior models efficiently



LTE-LAA and Wi-Fi coexistence scenario

[P. Shih, B. Moraffah 2020]

Bayesian Causal Inference

- How to incorporate Bayesian statistics to understand the causal relationships?
  - Bayesian feature selection in causal inference
    - What class of prior distributions can benefit us in feature selection?
    - What are the efficient methods to do inference when we have high-dimensional data?
    - What are the advantages of Bayesian modeling over the frequentist counterpart?

Bayesian Nonparametric Modeling of Health Data

- High-dimensional biological data
  - Learn patterns in biosequences, such as sequences obtained from biological samples to indicate some target phenotype.
  - Robust and scalable algorithms with theoretical guarantees for feature modeling/selection and convergence rate stipulate more complex unsupervised settings (feature allocation).
  - Learning casual relationship between features (learning causal Bayesian network) → allows for predicting a molecular system response to different interventions (perfect/general interventions)
  - Determining causal relationships in pattern discovery can be used in multiple sequence alignments, protein structure, function prediction, characterization of protein families, signal detection, and other areas
  - Integrate Bayesian nonparametric modeling with causal discovery methods

[B. Moraffah 2020, NIH]