Computational Statistical Modeling of Dynamic Socioeconomic, Geopolitical and Financial Systems

David K. A. Mordecai
NYU Courant Institute of Mathematical Sciences
Applied Mathematics Advanced Topics Course
Lecture #2 – January 31st, 2012

Social Network Framework of “Epidemics”

- Examples:
  - Financial: Defaults, Foreclosures
  - Socioeconomic: Crime statistics (domestic violence, arson, assault, vandalism, theft, homicide)
  - Demographics: Age, Gender, Income
  - Geopolitical: Voting Blocs
  - Epidemiological: Cardiac Incidents
- Compound (Discrete) Probability Distributions
  - A multinomial distribution with a probability vector distributed according to a Dirichlet distribution => a multivariate Pólya distribution (aka Dirichlet compound multinomial distribution)
  - Applications in document classification and clustering, genetics, economy, combat modeling, …
- Other Relevant Phenomena:
  - Social Contagion

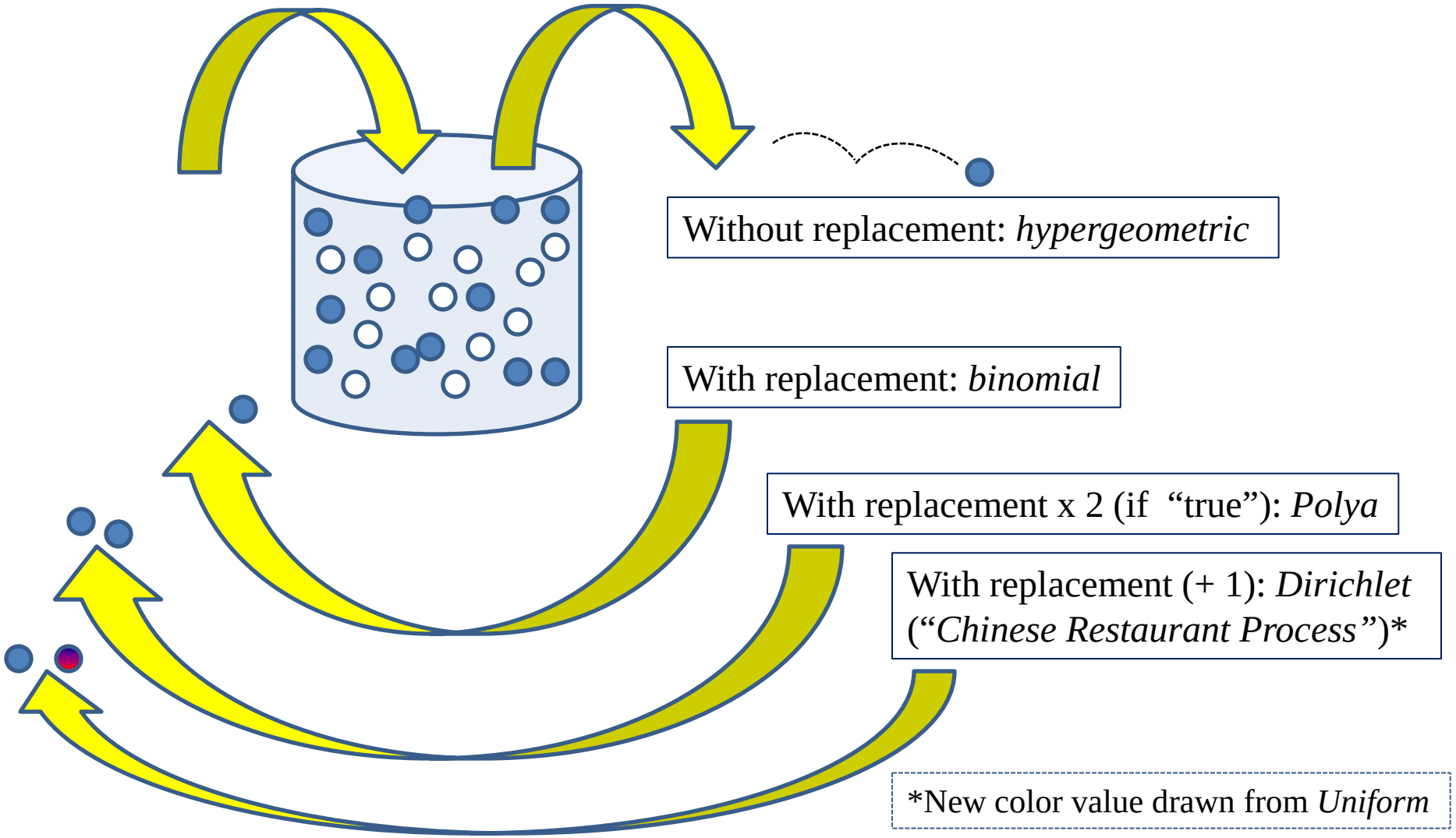
A Regional Population Simulation

- A General Framework for Simulation and Inference in N-player population games: what can be elements and choices?
  - States of nature *X*: House Price Appreciation (Observed); Systematic Economic Condition (Hidden)
  - Agents *N*: Community residents
  - Types *K*: as health subjects; households/homeowners; borrowers; voters; criminals or victims
  - Neighborhoods and/or Cohorts *L*: zip code, county (geographies); house/loan types; resident ages, genders, incomes/occupations; education; ethnicity (groupings)
    - Nodes *D*: Schools, markets, etc.; Broker, Agent, etc.
  - Information or innovations *F*: Discrete Type-Specific Event
  - Actions *A*: Default, Refinance; Arson, Vandalism, Assault, Homicide; Acute Cardiac Conditions (e.g. heart attacks, strokes, etc.); Ballot (choices, strategies)
    - Edges *E*: …
  - Payoffs observable *Y* (returns, outcomes)

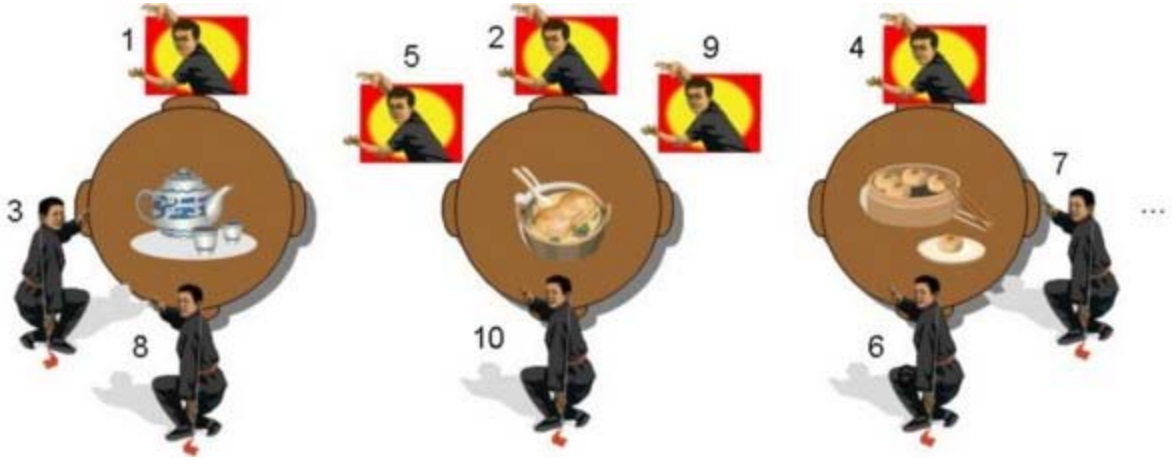
A Contagion Game?

- Apply The Framework for Simulation and Inference
  - Local House Price Changes and Income Distribution as State Variables
  - Agents as Community residents
  - Types health subjects; households/homeowners; borrowers; voters; criminals or victims
  - Neighborhoods and/or Cohorts *L*: zip code, county (geographies); house/loan types; resident ages, genders, incomes/occupations; education; ethnicity (groupings for conditioning information)
    - Nodes *D*: Schools, markets, hospitals, etc.; Broker, Agent, etc.
  - Information or innovations *F*: Discrete Type-Specific Event
  - Actions *A*: Defaults, Refinancings; Acute Health Emergencies; Arson, Vandalism, Assault, Homicide; Ballot (Crime Stats, Hospital Stats)
    - Edges *E*: …
  - What Payoffs may be observable? (returns, outcomes)

Contagion and the Polya Urn Model



Chinese Restaurant Process



$$G|G_0 \sim \mathrm{DP}(G_0)$$

$$\theta_1, \ldots, \theta_n | G \sim G$$

$$\Longrightarrow \theta_n | \theta_1, \ldots, \theta_{n-1} \sim CRP(G_0)$$

Chinese Restaurant Process

- If allocated to an occupied table, must order the same dish as those currently seated, OR receive a randomly assigned dish, if allocated to a new table
  - Related to the Polya Urn sampling scheme for finite Dirichlet distributions.
- The probability of an observation taking on a specific value is directly proportional to the number of times that value has already been seen (i.e. a popularity contest).
  - More generally, a random sample from a stochastic process whose sample path is also a (*Dirichlet*) probability distribution is a (finite-dimensional) Pitman–Yor distribution
  - Pitman–Yor process is useful for modeling data that exhibit *power-law* tail properties (e.g. wealth and income distributions)
- Note: The one-dimensional version of the multivariate *Polya* distribution is commonly known as the *Beta-binomial* distribution

$$\Theta = \{\theta_1, \theta_2, \ldots, \theta_m\}$$

$$\Theta \sim \text{Dirichlet}(\alpha_1, \alpha_2, \ldots, \alpha_m)$$

Distribution over possible parameter vectors for a multinomial distribution (generalization of the binomial for more than two outcomes).

- Beta distribution: special case of a Dirichlet for 2 dimensions.
- A distribution over distributions.

Remember: Multinomial can be interpreted as 2-D (triangular) slices of Pascal's pyramid (i.e. the 3-D, 4D, … (pyramid-shaped) slices of higher-dimensional analogs of Pascal's triangle. Hence the "range" or "support" of the distribution can be characterized by discrete equilaterial "pyramids" in arbitrary dimension (i.e. a *simplex* with a grid)

Indian Buffet Process

For the general case $\alpha>0$, the expected number of occupied tables

$$\frac{\Gamma(\theta+n+\alpha)\Gamma(\theta+1)}{\alpha\Gamma(\theta+n)\Gamma(\theta+\alpha)}-\frac{\theta}{\alpha}.$$

Indian Buffet Process

- Adaptation of the Chinese Restaurant Process
  - Each data point is no longer uniquely associated with a class, with any combination of the classes.
- Analogous to process in which each diner samples from a buffet some subset of an infinite selection of dishes on offer.
  - The probability that a particular diner samples a particular dish is proportional to the popularity of the dish among diners so far
  - The diner may also sample from the unsampled dishes.
- Useful for inferring latent features in data

To Consider: Some Other Related Distributions that are Particularly Useful:
• When k = 2, the multinomial distribution is the Binomial distribution.
• The continuous analogue is Multivariate Normal distribution.
• Categorical distribution (for k = 2 is Bernoulli)
• Beta-binomial model.

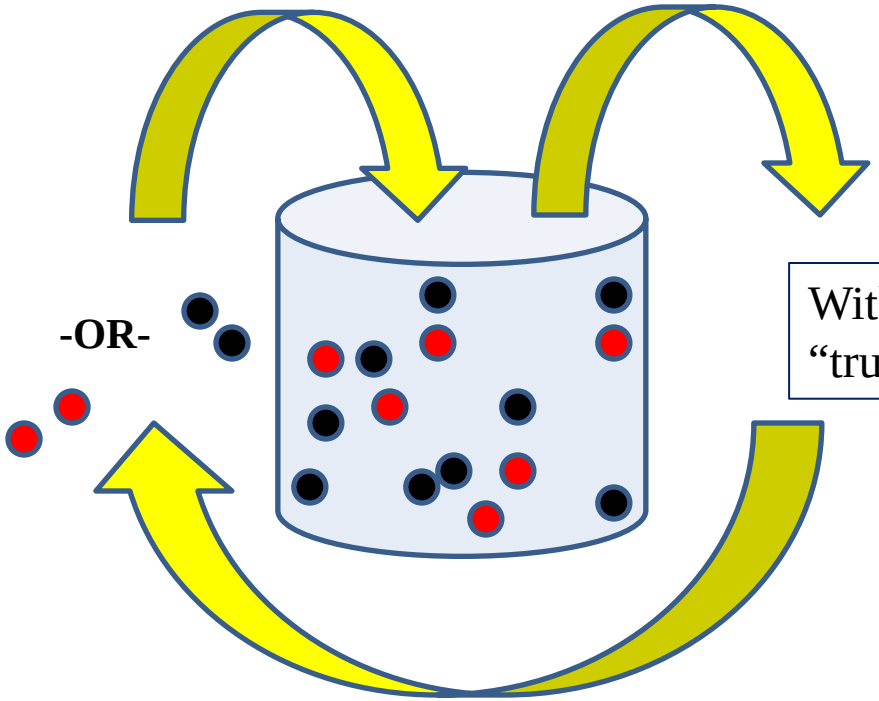
Beta-Binomial

$$\sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k}$$

Where: *p ~ Beta(α,β)*
and

$$\binom{n}{k} = \frac{n!}{k!\,(n-k)!}$$

For positive integer values *α , β:*



With replacement x 2 (if either "true" or "not true"): *Beta-Binomial*