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 #5 – February 28^{th}, 2012

Lecture #5 Outline

- One interpretation of the role of the Dirichlet Distribution as a model of algorithmic, adaptive learning via local search and interaction
- Learning, Path Dependence, and Spatial Location
  - Compound Poisson Arrivals
  - Logistic "Lotka-Volterra" Frameworks (See Epstein)
- Alternative Interpretation of the Threshold Framework in the Context of Reinforcement Learning (Local Search and Optimization)
- Hierarchical Models
- What is this?

$$p^r \binom{-r}{k} (-q)^k$$

Compound *Poisson* Arrivals

A compound *Poisson* process is a (note: continuous-time) stochastic process with jumps. The jumps arrive randomly according to a *Poisson* process, where the jump size is also a random variable, with a specified probability distribution. A compound Poisson process, parameterised by a rate $\lambda > 0$ and jump size distribution G, is a process $\{Y(t) : t \geq 0\}$ given by:

$$Y(t) = \sum_{i=1}^{N(t)} D_i$$

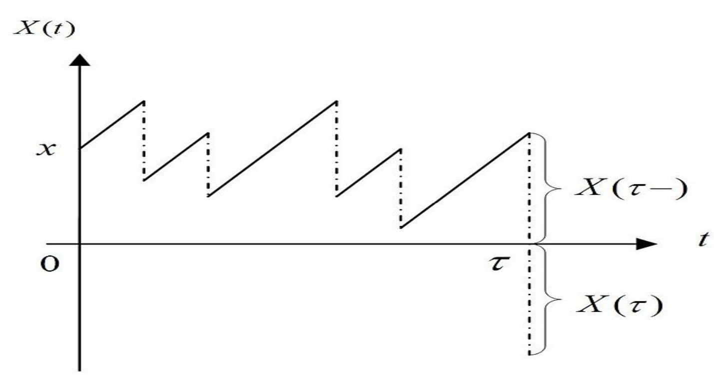
where, $\{N(t) : t \geq 0\}$ is a *Poisson* process with rate λ, and $\{D_i : i \geq 1\}$ are *i.i.d.* random variables, with distribution function G, which are also independent of $\{N(t) : t \geq 0\}$.

In accordance with conditional expectation, the expected value of a compound Poisson process can be calculated as:

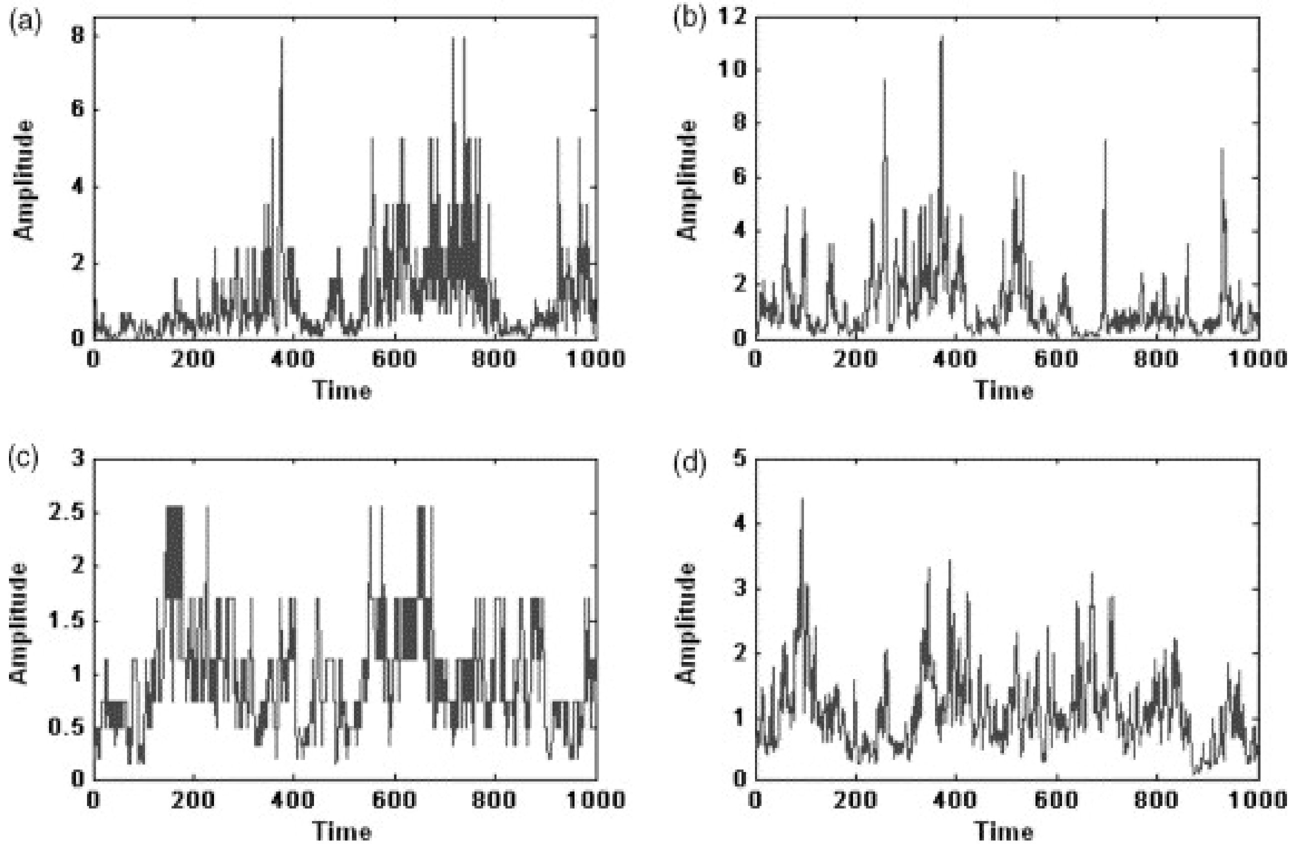
$$E(Y(t)) = E(E(Y(t)|N(t))) = E(N(t)E(D)) = E(N(t))E(D) = \lambda t E(D).$$

Making similar use of the law of total variance, the variance can be calculated as:

$$\begin{aligned}
\operatorname{var}(Y(t)) &= E(\operatorname{var}(Y(t)|N(t))) + \operatorname{var}(E(Y(t)|N(t))) \\
&= E(N(t)\operatorname{var}(D)) + \operatorname{var}(N(t)E(D)) \\
&= \operatorname{var}(D)E(N(t)) + E(D)^2 \operatorname{var}(N(t)) \\
&= \operatorname{var}(D)\lambda t + E(D)^2 \lambda t \\
&= \lambda t(\operatorname{var}(D) + E(D)^2) \\
&= \lambda t E(D^2).
\end{aligned}$$



Compound *Poisson* Arrivals (cont'd)



Compound *Poisson G* vs Log Normal *G*

Logistic "Lotka-Volterra" Frameworks

In *Lotka–Volterra* equations for predation in that the equation for each species has one term for self-interaction and one term for the interaction with other species. In the equations for predation, the base population model is exponential. For the competition equations, the logistic equation is the basis.

The logistic population model, when used by ecologists often takes the following form:

$$\frac{dx}{dt} = rx\left(1 - \frac{x}{K}\right).$$

Where x is the population size at time t, r is the per-capita growth rate and K is the environmental carrying capacity

For two populations, x_1 and x_2, with logistic dynamics, the Lotka–Volterra framework includes an additional term to account for the aggregated species' local interactions, resulting in the competitive *Lotka–Volterra* equations:

$$\frac{dx_1}{dt} = r_1 x_1 \left(1 - \left(\frac{x_1 + \alpha_{12} x_2}{K_1}\right)\right)$$

$$\frac{dx_2}{dt} = r_2 x_2 \left(1 - \left(\frac{x_2 + \alpha_{21} x_1}{K_2}\right)\right).$$

Here α_{12} represents the effect species 2 has on the population of species 1 and α_{21} represents the effect species 1 has on the population of species 2. These values do not have to be equal. Because this is the competitive version of the model, and therefore, all interactions being competitive, all α-values are positive. Note: each species may have its own growth rate and carrying capacity.

Logistic "*Lotka-Volterra*" Framework (cont'd)

This model can be generalized to any number of competing populations, for which populations and growth rates are represented as vectors, and the interactions α's as a matrix. Hence, the equation for any species *i* becomes:

$$\frac{dx_i}{dt} = r_i x_i \left(1 - \frac{\sum_{j=1}^{N} \alpha_{ij} x_j}{K_i}\right)$$

or, if the carrying capacity is incorporated into the interaction matrix (changing how the interaction matrix is defined),

$$\frac{dx_i}{dt} = r_i x_i \left(1 - \sum_{j=1}^{N} \alpha_{ij} x_j\right)$$

where *N* is the total number of interacting species. To simplify, all self-interacting terms α_{ii} are often set to 1.

Brian Arthur AER (1991)*

- Designing Economic Agents that Simulate Human Agents
  - Purposive and Boundedly (Procedurally) Rational
  - DKAM Conjecture #1: Perhaps Dirichlet Distribution as a compound distribution represents intergenerational evolution via local search?
- Strong Connection between Increasing Returns and ML
  - DKAM Conjecture #2: Can we think of Technology Adoption as "Imitative Learning" (*Lock-in* and *Reinforcement*)?
  - DKAM Conjecture #3: How can we characterize local Interactions in terms of local search and/or local optimization?
  - DKAM Conjecture #4: What are the implications for Path Dependence of "Spatial Location" (i.e. Proximity)

*Note: Arrow and Sargent were both acknowledged as discussants on this paper

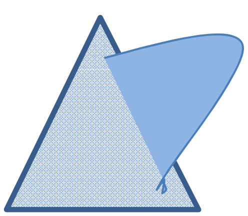
$\theta \sim \mathfrak{D}(\alpha)$ *Dirichlet: Is It A Multivariate Generalization of Beta?*

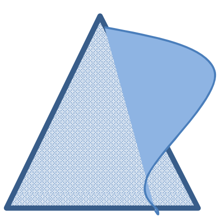
$$p(\theta) \sim \mathrm{D}(\alpha_1, \ldots, \alpha_k) = \frac{\Gamma(\sum_k \alpha_k)}{\prod_k \Gamma(\alpha_k)} \prod_k p_k{}^{\alpha_k - 1} \text{ where } \theta_k > 0 \text{ and } \sum_k p_k = 1$$

$$\beta(\alpha)^{-1} = \frac{\Gamma(\sum_i \alpha_i)}{\Gamma(\alpha_1) * \ldots * \Gamma(\alpha_n)} = \frac{\Gamma(\alpha_0)}{\prod_i \Gamma(\alpha_i)} \text{ where } \alpha_0 = \sum_i \alpha_i$$

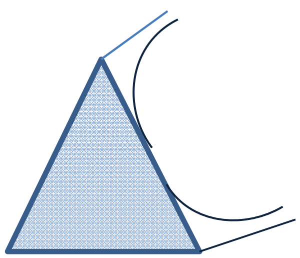
$$\theta = \{\theta_1, \ldots, \theta_n\}, \alpha = \{\alpha_1, \ldots, \alpha_n\} \text{ and } \alpha > 0$$

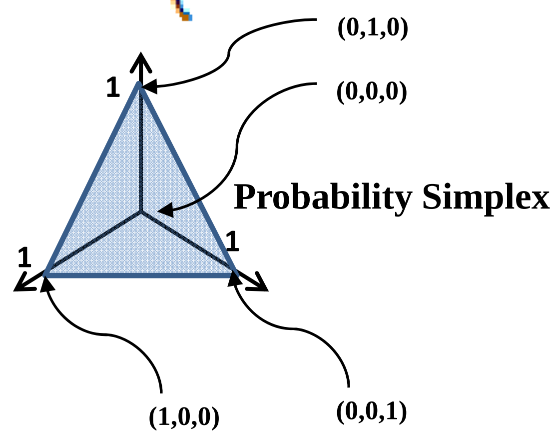
$$p(\theta) = \beta(\alpha)^{-1} \prod_k \theta_i{}^{\alpha_i - 1} \mathrm{I}(\theta \epsilon \mathcal{S}) \text{ where } \mathcal{S} = \left\{ x \epsilon \mathbb{R}^n : x_i \geq 0, \sum_i^n x_i = 1 \right\}$$



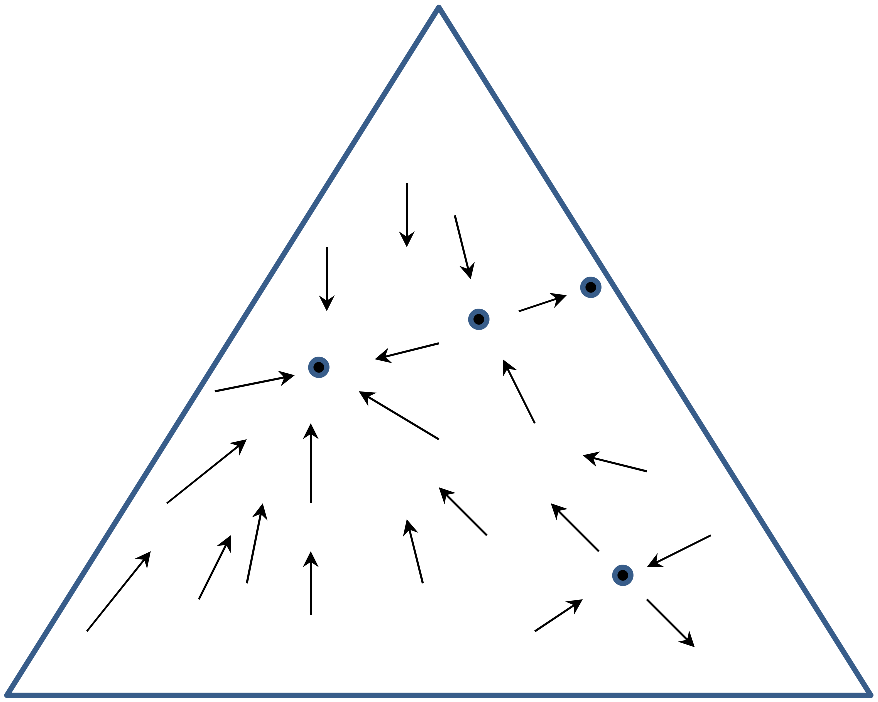


$$E[\theta_i] = \frac{\alpha_i}{\alpha_0}$$





Path Dependence and Spatial Location



<u>Expected Motions for a Locational Probability Function</u>

A Hierarchical Model Example

$\theta_i \mid x_i$

$x_i \sim N(\theta_i, \sigma^2)$ $\theta_i \sim N(\varphi, \tau^2)$

What if : $\varphi_i \sim N(0,1)$, *where* $\tau^2 \sim U(0,1)$?

Beta-Binomial

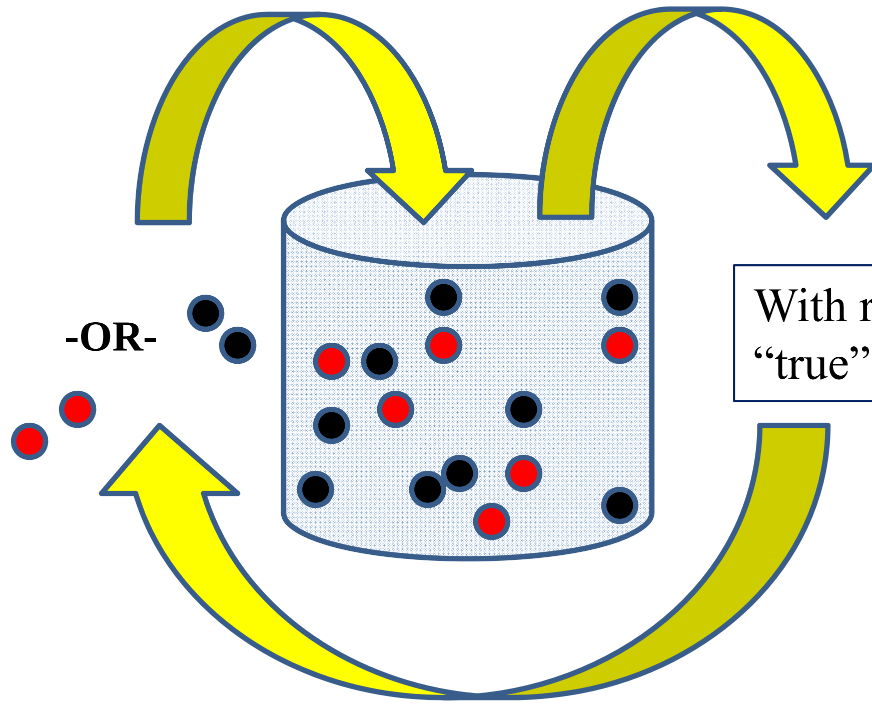
$$\sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k}$$

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*

Estimating A Hierarchical Model From Data

Beta-Binomial PMF

$$\binom{n}{k}\frac{B(k+\alpha, n-k+\beta)}{B(\alpha,\beta)}$$

$$p(\theta|x) \propto p(x|\theta)p(\theta)$$

$$p(\theta, \varphi|x) \propto p(x|\theta)p(\theta|\varphi)p(\varphi).$$

$\varphi|\psi$ *where* $\psi \sim$ ...

As 2-Stage Hierarchical Model

$$k_i \sim \mathrm{Bin}(n_i, \theta_i)$$

$$\theta_i \sim \mathrm{Beta}(\mu, M), \text{ i.i.d.}$$

Conditional Probability Expressed In Terms of Properties of Beta Distribution

$$f(k|\alpha,\beta) = \frac{\Gamma(n+1)}{\Gamma(k+1)\Gamma(n-k+1)}\frac{\Gamma(\alpha+k)\Gamma(n+\beta-k)}{\Gamma(\alpha+\beta+n)}\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}.$$

So What is This?

$$p^r \binom{-r}{k} (-q)^k$$

As a sum of geometric random variables, if $Y_r \sim NegBin(r, p)$, with support {0, 1, 2, ...}, then Y_r is a sum of r independent variables following the geometric distribution (on { 0, 1, 2, 3, ... }) with parameter $1 - p$ ➔ CLT: Y_r (properly scaled and shifted) is therefore approximately normal for sufficiently large r.

$$
\begin{aligned}
\Pr(Y_r \leq s) &= 1 - I_p(s+1, r) \\
&= 1 - I_p((s+r) - (r-1), (r-1)+1) \\
&= 1 - \Pr(B_{s+r} \leq r-1) \\
&= \Pr(B_{s+r} \geq r) \\
&= \Pr(\text{after } s+r \text{ trials, there are at least } r \text{ successes}).
\end{aligned}
$$

Note: Another useful property is that *NegBin* is infinitely divisible such that for any positive integer n, there exist independent identically distributed random variables $Y_1, ..., Y_n$ whose sum has the same distribution as does Y.

Also, if B_{s+r} is a random variable following the binomial distribution with parameters $s + r$ and $1 - p$, then the *NegBin* distribution is the "inverse" of the binomial distribution, and the sum of independent negative-binomially distributed random variables r_1 and r_2 with the same value for parameter p is negative-binomially distributed with the same p but with "r-value" $r_1 + r_2$.

$$f(k) = \frac{-p^k}{k \ln(1-p)}, \qquad k \in \mathbb{N}.$$

$$X = \sum_{n=1}^{N} Y_n$$

NegBin(r, p) can be also represented as a *compound (overdisperse) poisson*: Let $\{Y_n, n \in \mathbb{N}_0\}$ denote a sequence of *i.i.d*, each one having the logarithmic distribution *log*(p), with probability mass function and let N be a random variable (independent of the sequence). If N has a Poisson distribution with parameter $\lambda = -r \ln(1 - p)$. Then the random sum is NB(r,p)-distributed.

The binomial theorem implies that $Y \sim \text{Bin}(n,p)$ and $p + q = 1$ with $p, q \geq 0$.

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

Which by Newton's binomial theorem can be expressed as the following (for which the upper bound of summation is infinite)

$$(p+q)^n = \sum_{k=0}^{\infty} \binom{n}{k} p^k q^{n-k},$$

In certain cases, the binomial coefficient can be defined when n is a real number, (vs. a positive integer), or the binomial distribution can be zero when k > n.

$$\binom{n}{k} = \frac{n(n-1)(n-2)\cdots(n-k+1)}{k!}.$$

If one supposes r > 0 and uses a negative exponent, Then all of the terms are positive, and the probability that the number of failures before the rth success is equal to k, provided r is an integer.

$$1 = p^r \cdot p^{-r} = p^r(1-q)^{-r} = p^r \sum_{k=0}^{\infty} \binom{-r}{k} (-q)^k. \Longrightarrow p^r \binom{-r}{k} (-q)^k$$

The Newton-Raphson method in one variable: Given a function $f(x)$ and its derivative $f\,'(x)$, we begin with a first guess x_0 for a root of the function and iterate (provided the function is reasonably well-behaved) to find approximation x_1 … x_n

$$x_1 = x_0 - \frac{f(x_0)}{f'(x_0)}.$$

Geometrically, x_1 is the intersection with the x-axis of a line tangent to f at $f\,(x_0)$. Repeat until a sufficiently accurate value is reached:

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}.$$

To simulate a multinomial:

(1) reorder the parameters such that they are sorted in descending order (this is only to speed up computation and not strictly necessary).

(2) for each trial, draw X from a uniform (0, 1). The resulting outcome is the component for the multinomial distribution with n = 1.

$$j = \arg\min_{j'=1}^{k} \left(\sum_{i=1}^{j'} p_i \geq X \right).$$

Note: Some Other Useful Distributions Related to the Multinomial:

- When k = 2, the multinomial distribution is the Binomial distribution.
- The continuous analogue is Multivariate Normal distribution.
- Categorical distribution (for k = 2 is Bernoulli)
- The Dirichlet distribution (the conjugate prior of the multinomial in Bayesian statistics)
- Multivariate Pólya distribution.
- Beta-binomial model.