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 #7 – March 20th, 2012

Lecture #7 Outline

- *Branching Processes* and the Role of the *Beta* function in the *Yule-Simon process* (*linear preferential attachment)* of *cumulative advantage*
  - *Preferential attachment* exhibits which type of stochastic urn process?
  - Role of preferential attachment in scale-free networks (under suitable circumstances)
    - Species Propagation; Urban population densities; Income/wealth distributions
    - Links on the world wide web; cultural/linguistic propagation; technology adoption
    - Lotka's law of scientific productivity
    - Bradford's law of citations
- Introducing *Incomplete Beta* Function
  - Application of Regularization (in statistical model specification)
    - In statistics and machine learning used to prevent overfitting (ridge regression, lasso, L2-norm in SVMs) by imposing additional informational structure (e.g Bayesian prior)
    - Cross-validation (regression on subsamples)
    - Regularization introduces a second factor which weights the penalty against more complex models with an increasing variance in the data errors. This gives an increasing penalty as model complexity increases.
    - For example, see Ridge regression, which imposes penalty function on regression coefficients

Evaluating *the Beta* Distribution

PDF: $f(x;\alpha,\beta) = \dfrac{x^{\alpha-1}(1-x)^{\beta-1}}{\int_0^1 u^{\alpha-1}(1-u)^{\beta-1}\,du}$

$$= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\, x^{\alpha-1}(1-x)^{\beta-1}$$

Where $\Gamma(z)$ is the *Gamma* function

$$= \frac{1}{\mathrm{B}(\alpha,\beta)}\, x^{\alpha-1}(1-x)^{\beta-1}$$

and ,where: $B(\alpha,\beta) = \dfrac{(\alpha-1)!\,(\beta-1)!}{(\alpha+\beta-1)!}$

$$\frac{x^{\alpha-1}(1-x)^{\beta-1}}{\dfrac{(\alpha-1)!\,(\beta-1)!}{(\alpha+\beta-1)!}}$$

So what if : $\alpha, \beta \to \infty$?

When: $0 < x < 1$

$x > 1$

$x < 0$

The *Yule-Simon Process and the Beta* Function

- in the theory of the *preferential attachment process*, the *Yule-Simon process* (linear preferential attachment) of *cumulative advantage*
  - a type of *stochastic urn process*
  - under suitable circumstances, can generate power law distributions from *branching processes*
- Commonly observed within species distributions, income and wealth distributions, various *scale-free* networks
  - Speciation; Urban population densities; Income/Wealth Distributions
  - Cultural and Language Propagation
  - Technology Adoption/Innovation; Financial Contagion (?)
  - Lotka's law of scientific productivity, Bradford's law of citations
  - World Wide Web

Beta-Binomial

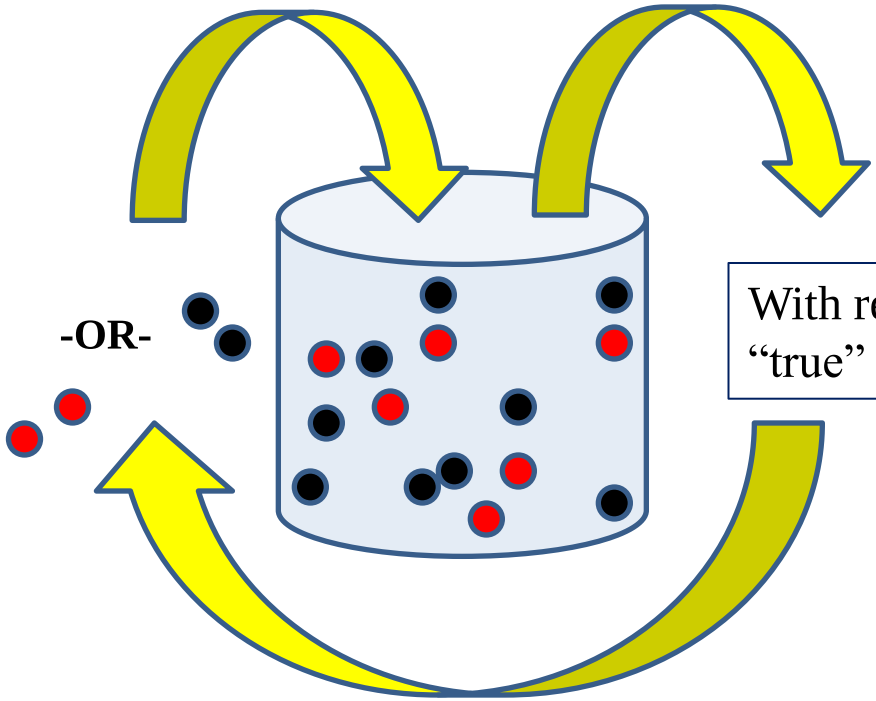
$$\sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k}$$

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*

Linear Preferential Attachment Processes

- Linear preferential attachment processes in which the number of urns increases are known to produce a distribution of balls over the urns following the so-called
- Yule-Simon distribution (also a continuous mixture of geometric distributions, e.g. exponential-geometric mixture).
- In the most general form of the process, balls are added to the system at an overall rate of m new balls for each new urn.
- Each newly created urn starts out with k_0 balls and further balls are added to urns at a rate proportional to the number k that they already have plus a constant $a > -k_0$.

The fraction $P(k)$ of urns having k balls in the limit of long time is given by:

$$P(k) = \frac{\mathrm{B}(k+a,\gamma)}{\mathrm{B}(k_0+a,\gamma-1)},$$

with $\Gamma(x)$ being the standard gamma function, and

$$\mathrm{B}(x,y) = \frac{\Gamma(x)\Gamma(y)}{\Gamma(x+y)},$$

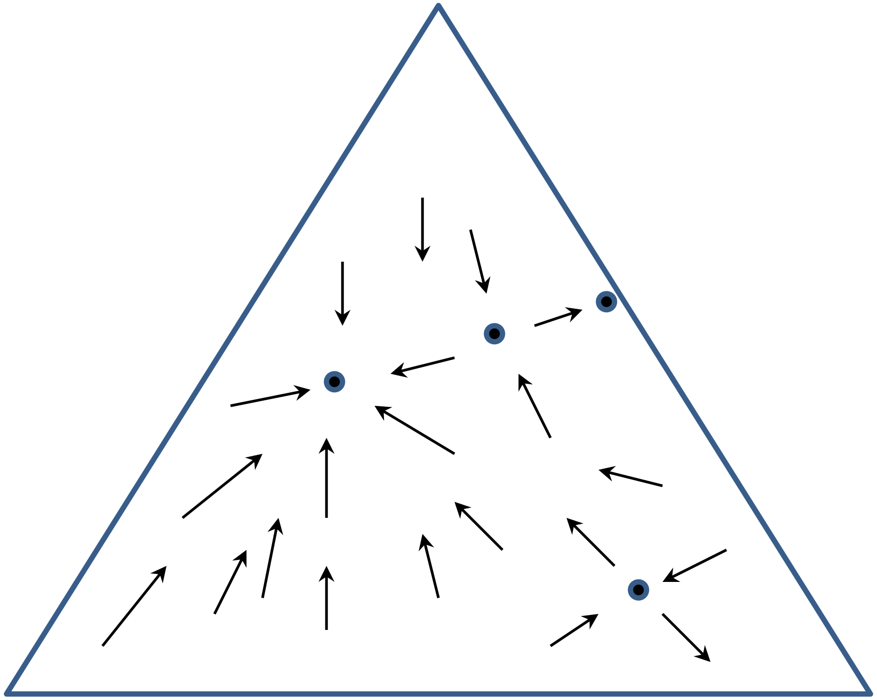
for $k \geq k_0$ (and zero otherwise), where $\mathrm{B}(x, y)$ is the Euler beta function:

$$\gamma = 2 + \frac{k_0+a}{m}.$$

The beta function behaves asymptotically as $\mathrm{B}(x, y) \sim x^{-y}$ for large x and fixed y, which implies that for large values of k, the preferential attachment process generates a "long-tailed" distribution (a realization of Zipf's law) following a Pareto distribution (i.e. power law in the tail of the distribution):

$$P(k) \propto k^{-\gamma}.$$

Preferential Attachment, Path Dependence, and Spatial Location



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

For example, in the case of species propagation, new "urns" are added to a class in a taxonomy whenever a newly appearing species is considered sufficiently different from its predecessors that it does not belong in any of the current classes. New species "balls" are added as prior generations branch out (i.e. speciate). Assuming that new species belong to the same genus as their parent (except for those that start new genera), the probability that a species is added to a new genus will be proportional to the number of species the genus already has. This process of speciation (*assortative mixing* process) first studied by Yule, is a linear preferential attachment process, since the rate at which classes accrue new species is linear in their prior number.

Hence:

$$\Gamma(x)\,\Gamma(y) = \Gamma(x+y)\mathrm{B}(x,y).$$

$$\Gamma(x)\Gamma(y) = \left(\int_{\mathbb{R}} f(u)\,du\right)\left(\int_{\mathbb{R}} g(u)\,du\right) = \int_{\mathbb{R}} (f*g)(u)\,du = \mathrm{B}(x,y)\,\Gamma(x+y)$$

$$\frac{\partial}{\partial x}\mathrm{B}(x,y) = \mathrm{B}(x,y)\left(\frac{\Gamma'(x)}{\Gamma(x)} - \frac{\Gamma'(x+y)}{\Gamma(x+y)}\right) = \mathrm{B}(x,y)(\psi(x) - \psi(x+y)),$$

where $\psi(x)$ is the digamma function defined as the logarithmic derivative of the gamma function

$$\psi(x) = \frac{d}{dx}\ln\Gamma(x) = \frac{\Gamma'(x)}{\Gamma(x)}.$$

But, remember from this 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 Bayesian *conjugate prior* of the multinomial)**
- **Multivariate Pólya distribution.**
- **Beta-binomial model.**

Note: Stirling's approximation gives the asymptotic formula:

$$\mathrm{B}(x,y) \sim \sqrt{2\pi}\frac{x^{x-\frac{1}{2}}y^{y-\frac{1}{2}}}{(x+y)^{x+y-\frac{1}{2}}}$$

For large $\{x, y\}$. Note: if $\{x, y\}$ fixed, then:

$$\mathrm{B}(x,y) \sim \Gamma(y)\, x^{-y}.$$

A generalization of the *Beta* function is the *incomplete Beta* function

$$\mathrm{B}(x;\, a, b) = \int_0^x t^{a-1}\,(1-t)^{b-1}\,dt.$$

Note: if x=1, then the *incomplete Beta* function corresponds to the (complete) *Beta* function

Note: The regularized incomplete beta function is the CDF of the Beta distribution, and is related to the CDF of a random variable from a binomial distribution, where the "probability of success" is p and the sample size is n.

Note: Algorithms with "GammaLn" and "(cumulative) *Beta* distribution functions can be used to calculate:

Complete Beta Value = Exp((GammaLn(a) + GammaLn(b) - GammaLn(a + b))

and,

Incomplete Beta Value = BetaDist(x, a, b) * Exp((GammaLn(a) + GammaLn(b) - GammaLn(a + b)),

which result from rearranging the form for the Beta distribution, and the incomplete beta and complete beta functions (when defined as the ratio of the logs)

So, Is Dirichlet A Multivariate Generalization of Beta?

The relationship between the *Beta* and *Incomplete Beta* functions is like that between the *Gamma* function and its generalization the *incomplete Gamma* function.

$$\theta \sim \mathcal{D}(\alpha)$$

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

The regularized *incomplete Beta* function (i.e. *regularized Beta* function) is defined in terms of the *incomplete Beta* function and the (complete) beta function.

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

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

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

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