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 #3 – February 7th, 2012

Lecture #3 Outline

- Principal Motivation: Application of Social Computing Paradigm to Models of Adaptive Distributed Learning
  - (Bayesian) Statistical Foundation
  - Focus: Path- and State- Dependence (e.g. Social Contagion)
- Review counting processes and related distributions
  - Compound and mixture distributions
    - Negative Binomial
  - Revisit Urn Processes (Polya Urn Processes and Path-Dependence)
- Dirichlet distribution and related processes (within the context of the probability simplex)
- Gamma function within the context of the area and volume of circles, spheres, hyperspheres, hypercubes
  - Information geometry in terms of densities or "point clouds" (i.e. clusters of probability in time and space)
  - Relationship of Pascal's Triangle to probability simplices (i.e., hyper-tetrahedrons), and polytopes (triangles, tetrahedrons, squares, cubes)

Fundamental notion of counting processes

- Distributions with useful statistical properties for spatio-temporal (state-space) modeling
  - Relevance as a means of characterizing both path-dependence (temporal, i.e. memory) and local interactions (spatial, i.e. network effects)
  - Clustering in time and space; Clustering across (latent) characteristics
- Bernoulli distribution and urn sampling (either with and without replacement)
  - A sequence or set of independent (with replacement) Bernoulli observations:
  - Basis for the Polya family (of binomial, geometric, Pascal, and negative binomial distributions), in which the sequence or set is truncated after either *n* trials of *x* positive (negative) outcomes (or Poisson in which the sequence or set is arbitrarily large, i.e. approaches infinity), e.g. in a case where 'true'=1, ‘not true'=0:

Binomial Distribution (and Some Useful Variants)

Binomial

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

where

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

Geometric

$$(1-p)^k p$$

for $k = 0,1,2,3, \ldots$

ratio of 'false' observations relative to the sequence of *n* total observations before the first 'true' observation

Negative Binomial

$$\binom{k+r-1}{k} (1-p)^r p^k$$

for $k = 0,1,2,3, \ldots$

the ratio of 'false' observations occurring before a specified *i* true observations relative to the sequence of n total observations

Poisson

$$\frac{\lambda^k e^{-\lambda}}{k!}$$

for $k = 0,1,2,3, \ldots$

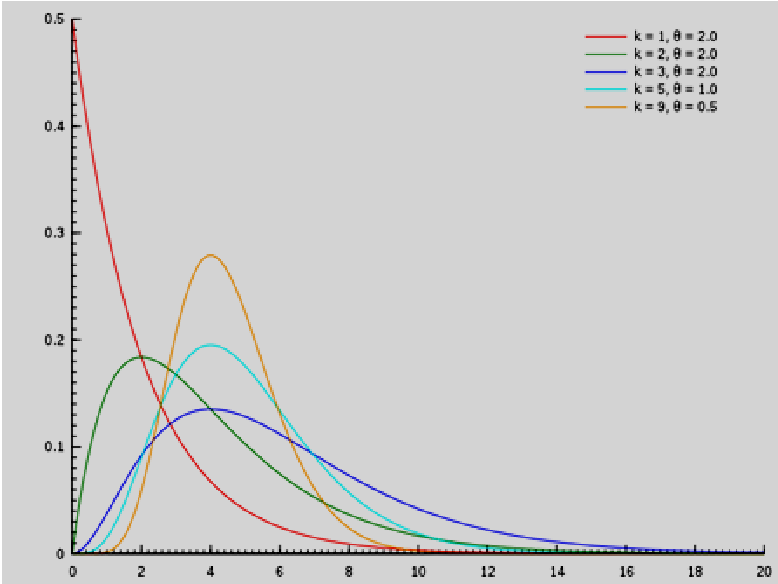
Note: the waiting time between poisson-distributed events is distributed exponential (special case of the Gamma distribution)

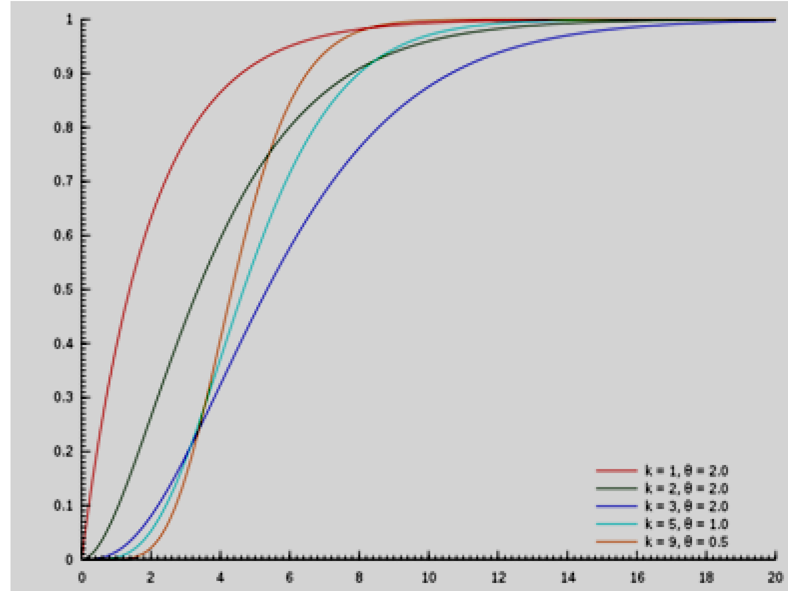
Pascal

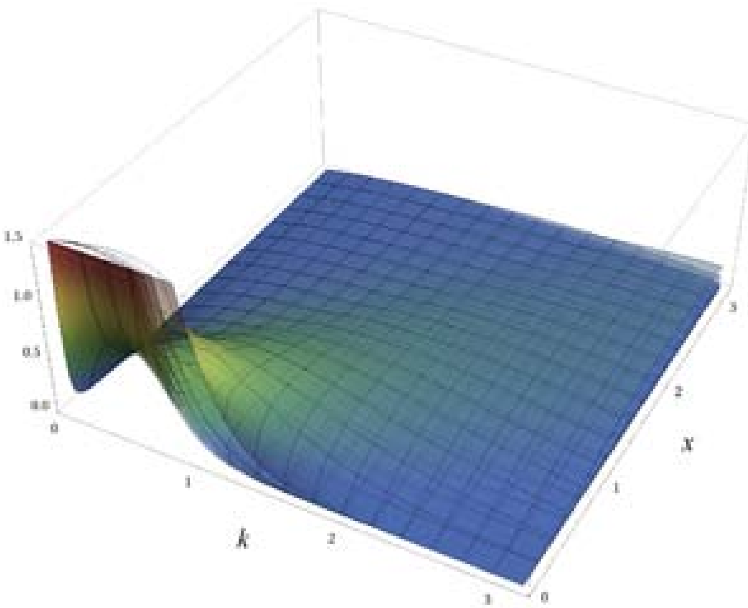
$$\frac{\Gamma(k+r)}{k!\,\Gamma(r)}$$

If *r* is a real-valued integer

(i.e. Binomial Waiting-Time distribution) is a Negative Binomial distribution shifted α units along the x-axis, hence a distribution that runs from α to infinity.







3D Representation of a Gamma Probability Distribution. With superimposed surfaces (each for a different value of θ)

If X_i has a $\Gamma(k_i, \theta)$ distribution for $i = 1, 2, ..., N$ (i.e., all distributions have the same scale parameter θ), then provided all $X_{i'}$ *are i.i.d.*

$$\sum_{i=1}^{N} X_i \sim \mathrm{Gamma}\left(\sum_{i=1}^{N} k_i, \theta\right)$$

The gamma distribution exhibits infinite divisibility.

Note: the computational intensity of computing the median depends upon the α parameter and does not have an easy closed-form solution (in contrast to the median and mode)

$$\frac{1}{\Gamma(a)\theta^a}\int_0^{x_0} x^{a-1}e^{-x/\theta}dx = \frac{1}{2}$$

… and hence is best simulated

Negative Binomial is equivalent to a Poisson(λ) distribution, where λ is itself a random variable, distributed according to *Gamma*(r, $p/(1-p)$)

$$f(k) = \int_0^\infty f_{\mathrm{Poisson}(\lambda)}(k) \cdot f_{\mathrm{Gamma}\left(r,\, \frac{p}{1-p}\right)}(\lambda)\, \mathrm{d}\lambda$$

$$= \int_0^\infty \frac{\lambda^k}{k!} e^{-\lambda} \cdot \lambda^{r-1} \frac{e^{-\lambda(1-p)/p}}{\left(\frac{p}{1-p}\right)^r \Gamma(r)}\, \mathrm{d}\lambda$$

$$= \frac{(1-p)^r p^{-r}}{k!\,\Gamma(r)} \int_0^\infty \lambda^{r+k-1} e^{-\lambda/p}\, \mathrm{d}\lambda$$

$$= \frac{(1-p)^r p^{-r}}{k!\,\Gamma(r)}\, p^{r+k}\, \Gamma(r+k)$$

$$= \frac{\Gamma(r+k)}{k!\,\Gamma(r)}\, (1-p)^r p^k.$$

Also known as the gamma–Poisson (mixture) distribution, a continuous mixture of Poisson distributions where the mixing distribution of the Poisson rate is a gamma distribution

Given:

$$\sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} \quad \text{where} \quad \binom{n}{k} = \frac{n!}{k!\,(n-k)!}$$

If the upper bound of the summation is infinite (Newton's Binomial Theorem):

$$\binom{n}{k} = \frac{n!}{k!\,(n-k)!} \Longrightarrow \binom{n}{k} = \frac{n(n-1)(n-2)\ldots(n-k+1)}{k!}$$

Then (if r>0 and is an integer), by employing a negative exponent:

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

And, the probability that the number of not true before the rth true statement

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

Beta-Binomial

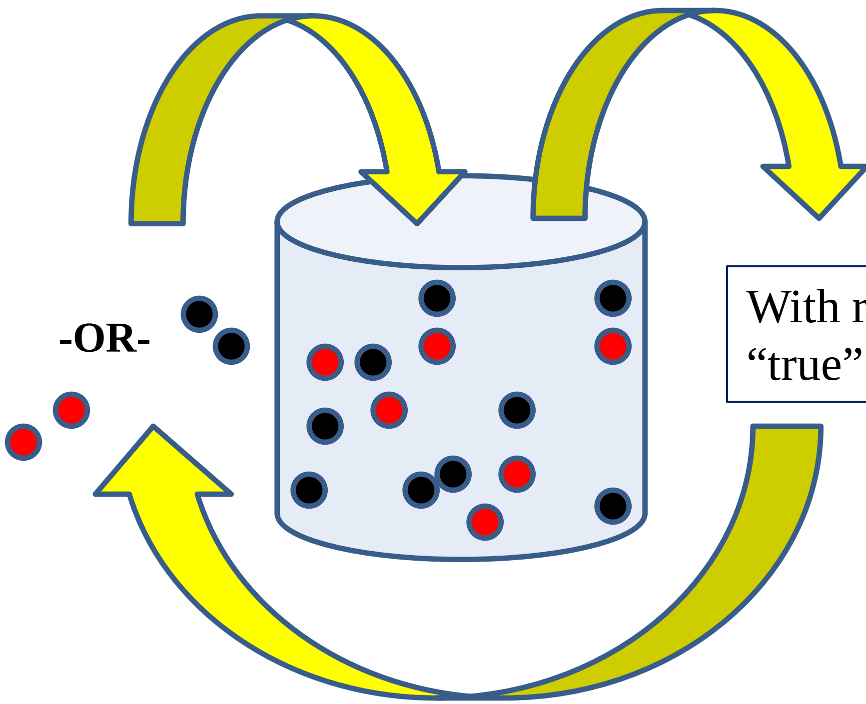
$$\sum_{k=0}^{n}\binom{n}{k}p^{k}(1-p)^{n-k}$$

Where: *$p \sim Beta(\alpha,\beta)$*

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*

Beta-Binomial PMF

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

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

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

Negative Multinomial $\Gamma\left(\sum_{i=0}^{m} k_i\right)\frac{p_0^{k_0}}{\Gamma(k_0)}\prod_{i=1}^{m}\frac{p_i^{k_i}}{k_i!}$

Sampling from a multinomial distribution

First, reorder the parameters $p_1, \cdots p_k$ such that they are sorted in descending order (this is only to speed up computation and not strictly necessary). Now, for each trial, draw an auxiliary variable *X* from a uniform (0, 1) distribution. The resulting outcome is the component

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

i.e., a sample for the multinomial distribution with n = 1. A sum of independent repetitions of this experiment is a sample from a multinomial distribution with n equal to the number of such repetitions.

Note: Pair-wise *NegMN* correlations are always positive, whereas the correlations between multinomial counts are always negative. As the parameter k_0 increases, pairwise correlations =>0 The multinomial counts X_i behave as *independent Poisson* random variables with respect to their means, where the marginal distribution of each of the X_i variables is distributed negative binomial, as the X_i count (true statement) is measured against all the other observations (not true statements).

$$\left(\mu_i = k_0\frac{p_i}{p_0}\right)$$

Jointly, the distribution of is negative multinomial, $X = \{X_1, \cdots, X_m\}$ i.e., $X \sim NM(k_0, \{p_1, \cdots, p_m\})$.

Negative multinomial distribution: the generalization of *NegBin*(r, p) to more than two outcomes.

For a simulation that generates $m+1 \geq 2$ possible outcomes, $\{X_0,\ldots,X_m\}$, each occurring with non-negative probabilities $\{p_0,\ldots,p_m\}$ respectively, sampling proceeds until X_0 reaches the predetermined value k_0, then the distribution of the m-tuple $\{X_1,\ldots,X_m\}$ is *negative multinomial*.

Multinomial Visualization (as slices of generalized Pascal's triangle)

- Binomial distribution as (normalized) 1D slices of Pascal's triangle,
  - the multinomial distribution as 2D (triangular) slices of Pascal's pyramid, or 3D, 4D, (pyramid-shaped) slices of higher-dimensional analogs of Pascal's triangle.
- An interpretation of the range of the distribution: discretized equilaterial "pyramids" in arbitrary dimension -- i.e. a simplex with a grid.

As polynomial coefficients: interpret the binomial distribution as the expansion of the polynomial coefficients of $(px_1 + (1 - p)x_2)^n$, and hence the multinomial distribution as the expansion coefficients of $(p_1x_1 + p_2x_2 + p_3x_3 + \ldots + p_kx_k)^n$.

In each case, the coefficients must sum to 1.

$$\sum_{i=1}^{k} p_i = 1$$

Expected Value

$$\mathrm{E}(X_i) = np_i.$$

Variance (Diagonals)

$$\mathrm{var}(X_i) = np_i(1 - p_i).$$

Covariances (Off-Diagonals)

$$\mathrm{cov}(X_i, X_j) = -np_ip_j$$

Correlations

$$\rho(X_i, X_j) = -\sqrt{\frac{p_ip_j}{(1 - p_i)(1 - p_j)}}.$$

$$f(x_1, \ldots, x_k; n, p_1, \ldots, p_k) = \Pr(X_1 = x_1 \text{ and } \ldots \text{ and } X_k = x_k)$$

$$= \begin{cases} \dfrac{n!}{x_1! \cdots x_k!} p_1^{x_1} \cdots p_k^{x_k}, & \text{when } \sum_{i=1}^k x_i = n \\ 0 & \text{otherwise,} \end{cases}$$

The random variables X_i indicate the number of times outcome number i was observed over the n trials, such that the vector $X = (X_1, ..., X_k)$ follows a multinomial distribution with parameters n and $\mathbf{p}$, where $\mathbf{p} = (p_1, ..., p_k)$, for negative integers $x_1 \ldots, x_k$.

For the multinomial, the analog of the Bernoulli distribution is the categorical distribution, where each trial results in exactly one of some fixed finite number k of possible outcomes, with probabilities $p_1, ..., p_k$ (so that pi ≥ 0 for $i = 1, ..., k$), and n i.i.d. trials.

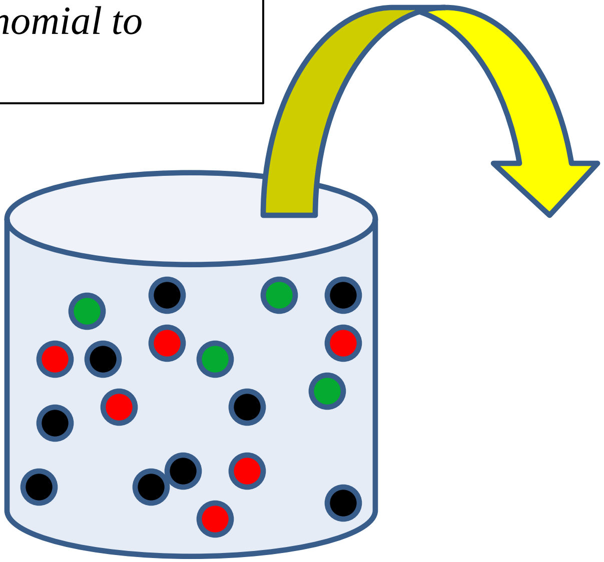
Note: The categorical distribution is equivalent to a multinomial over a single observation, i.e. a vector (1-of-K) with one element equal to 1 and all other elements equal to 0. Hence, in natural language processing the terms are used interchangeably.

Example: Assuming an assignment among three types for a large population of agents (example votes), type A comprises 20%, type B comprises 30%, and type C comprises 50% of the population, respectively. If six agents are sampled randomly, what is the probability that there will be exactly one agent of type A, two agents for type B and three agents of type C in the sample?

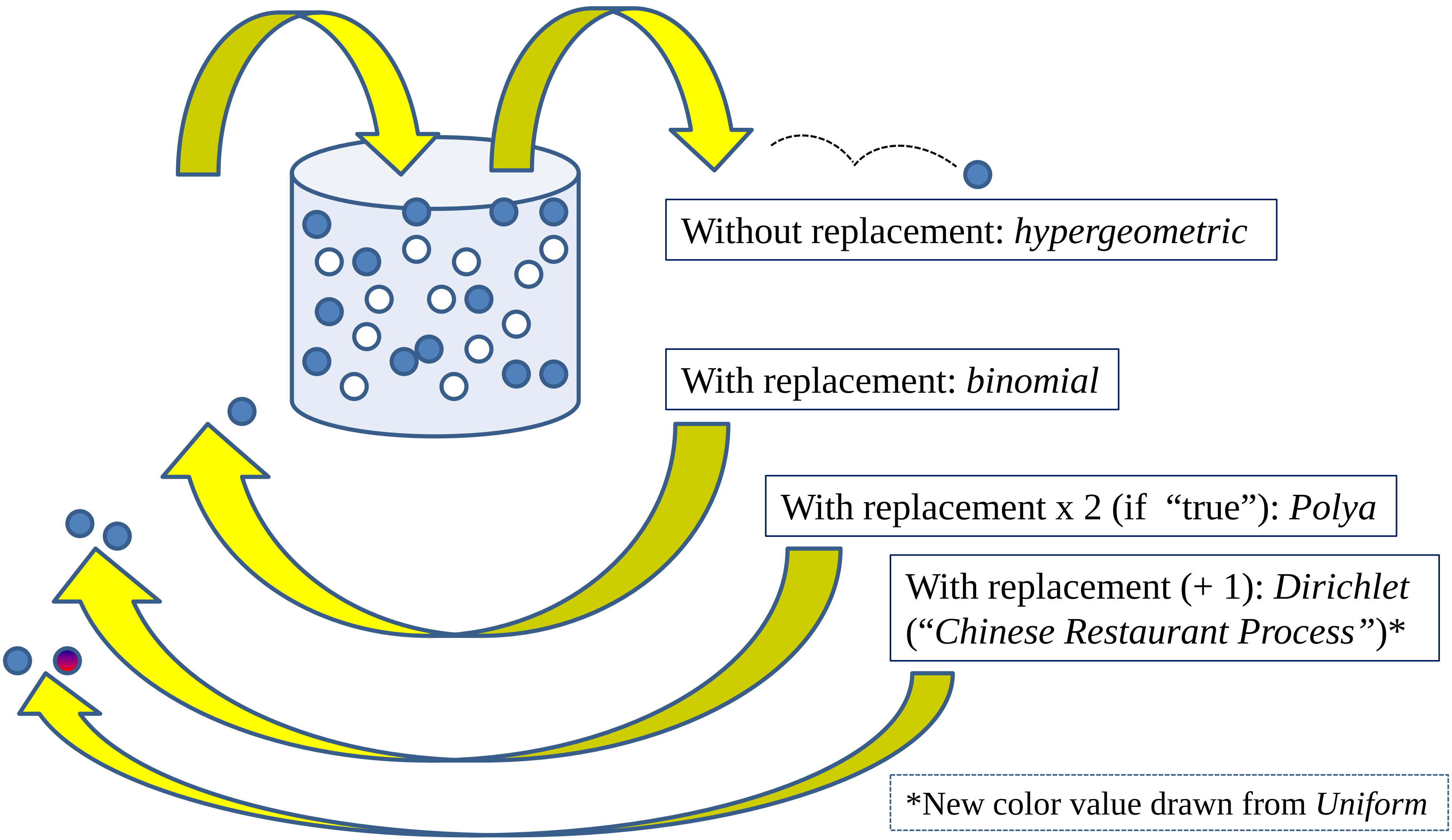
Example: Assuming an assignment among three types for a large population of agents (example votes), type A comprises 20%, type B comprises 30%, and type C comprises 50% of the population, respectively. If six agents are sampled randomly, what is the probability that there will be exactly one agent of type A, two agents for type B and three agents of type C in the sample?

$$\Pr(A=1, B=2, C=3) = \frac{6!}{1!2!3!}(0.2^1)(0.3^2)(0.5^3) = 0.135$$

Note: Assuming that the population is large, can treat the probabilities as unchanging once a agent is selected for sample, and hence technically, this represents sampling without replacement. The correct distribution would be the multivariate hypergeometric distribution, but these two distributions converge (multinomial to multivariate hypergeometric) as the population grows large.



Contagion and the Polya Urn Model



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

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

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)

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

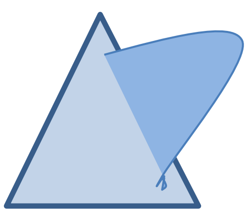
$$p(\theta) \sim \mathrm{D}(\alpha_1, \dots, \alpha_k) = \frac{\Gamma(\sum_k \alpha_k)}{\prod_k \Gamma(\alpha_k)} \prod_k p_k^{\alpha_k - 1} \text{ where } \theta_k > 0 \text{ and } \sum_k p_k = 1$$

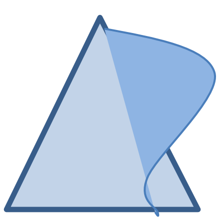
$$\beta(\alpha)^{-1} = \frac{\Gamma(\sum_i \alpha_i)}{\Gamma(\alpha_1) * \dots * \Gamma(\alpha_n)} = \frac{\Gamma(\alpha_0)}{\prod_i \Gamma(\alpha_i)} \text{ where } \alpha_0 = \sum_i \alpha_i$$

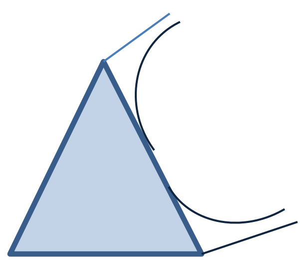
$$\theta = \{\theta_1, \dots, \theta_n\}, \alpha = \{\alpha_1, \dots, \alpha_n\} \text{ and } \alpha > 0$$

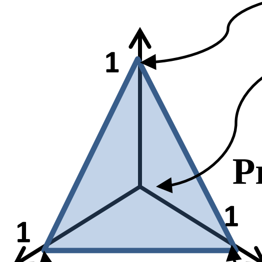
$$p(\theta) = \beta(\alpha)^{-1} \prod_k \theta_i^{\alpha_i - 1} \mathrm{I}(\theta \epsilon \mathcal{S}) \text{ where } \mathcal{S} = \left\{ x \epsilon \mathbb{R}^n : x_i \geq 0, \sum_i^n x_i = 1 \right\}$$

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







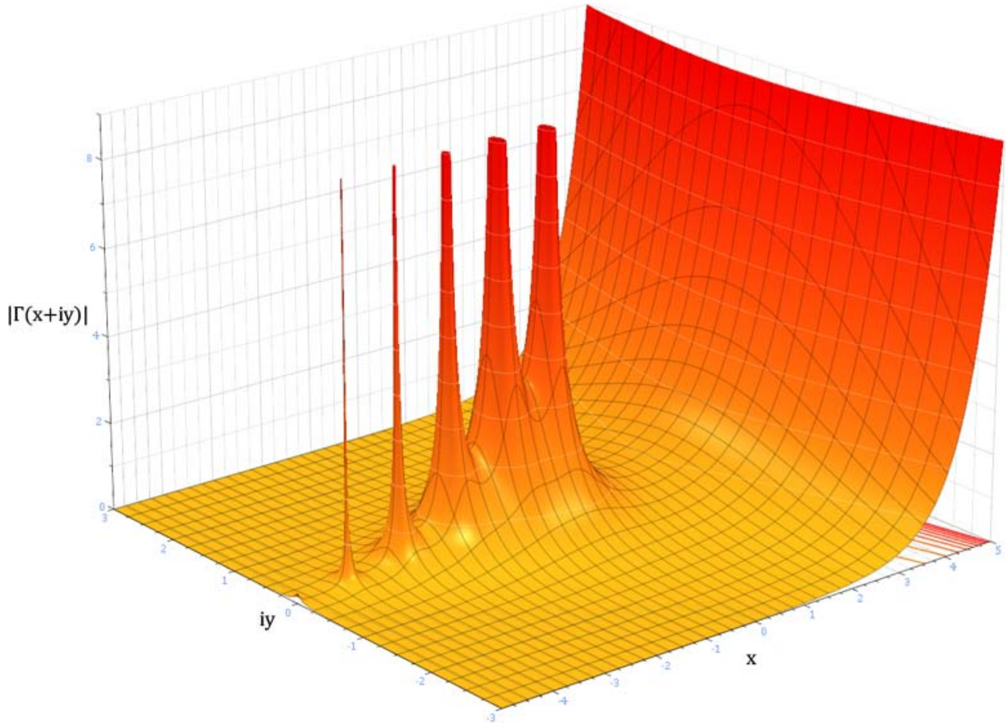


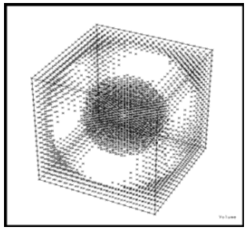
The gamma function, an extension of the factorial function (with its argument shifted down by 1), is a solution to a interpolation problem –to find smooth curve that connects the points (x, y) given by y = (x − 1)! at the positive integer values for x – applied to real and complex numbers, such that if *n* is a positive integer, then:

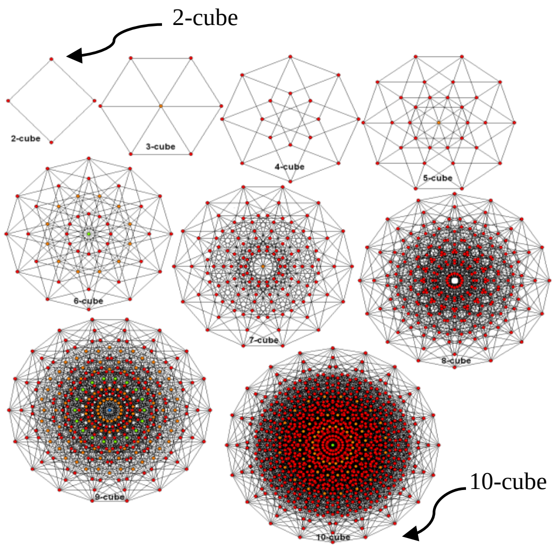
$$\Gamma(n) = (n-1)!$$

The absolute value of the gamma function on the complex plane: $\Gamma(z)$ for an increasing positive variable is simple: it grows quickly — faster than an exponential function. Asymptotically as , $z \to \infty$

$$\Gamma(z+1) \sim \sqrt{2\pi z\left(\frac{z}{e}\right)^{z}}$$







- In topology, an *n-simplex* is equivalent to an *n-ball*, and hence, every *n-simplex* is an *n-dimensional manifold with boundary* (covered?), where the volume of the *standard n-simplex* equals 1/(n+1)! and the volume of a *regular n-simplex* equals $\sqrt{n+1}/n!\sqrt{2^n}$

- Probabilistically, the points of the *standard n-simplex* in *(n + 1)-space* are equivalent to the space of possible parameters <u>(probabilities) of the categorical distribution on n+1 possible outcomes</u>. Particularly relevant to numerical applications of probability theory in terms of *a projection* onto the *standard simplex (information geometry)*:
  - In algebraic topology, simplices are used as building blocks to construct an interesting class of topological spaces called *simplicial complexes* (i.e., simplices glued together in a combinatorial fashion).
  - An *n-simplex* when interpreted as a corner of the *n-cube*, it is a *standard orthogonal simplex*. In the related context, the concept of an *orthogonal corner* means there is a vertex at which all adjacent hyperfaces are pairwise orthogonal (relevant for dimension reduction in statistics). <u>As generalizations of right angle triangles, these motivate a n-dimensional version of the Pythagorean theorem</u>:
  - The sum of the squared (n-1) dimensional volumes of the hyperfaces adjacent to the orthogonal corner equals the squared (n-1) dimensional volume of the hyperface opposite of the orthogonal corner.

- Mapping a hypercube's vertices to each of an n-simplex's elements can efficiently enumerate the simplex's face lattice (since more general related algorithms tend to be more computationally intensive).

The space enclosed by an n-sphere is called an (n + 1)-ball. An (n + 1)-ball is closed if it includes the n-sphere, and it is open if it does not include the n-sphere.

o1-ball, a line segment, is the interior of a (0-sphere).
o2-ball, a disk, is the interior of a circle (1-sphere).
o3-ball, an ordinary ball, is the interior of a sphere (2-sphere).
o4-ball, is the interior of a 3-sphere, etc.

The n-volume of an n-sphere of radius R or, equivalently, the surface area of an (n + 1)-ball of radius R is:

$$S_n(R) = \frac{2\pi^{\frac{n+1}{2}}}{\Gamma(\frac{n+1}{2})} R^n$$

The n-volume of a n-ball of radius R:

$$V_n(R) = \frac{\pi^{\frac{n}{2}}}{\Gamma(\frac{n}{2}+1)} R^n$$

1-sphere (circle with radius R) circumference:

$$S_1(R) = 2\pi R.$$

2-ball (sphere with radius R) area:

$$S_2(R) = 4\pi R^2$$

2-ball (disk with radius R) area:

$$V_2(R) = \pi R^2.$$

3-ball (with radius R) volume:

$$V_3(R) = \frac{4}{3}\pi R^3.$$

In general, the n-ball of radius R volume is (proportional to) the nth power of the R:

$$V_n(R) = C_n R^n \quad \text{where} \quad C_n = \frac{\pi^{\frac{n}{2}}}{\Gamma(\frac{n}{2}+1)}$$

C is constant of proportionality

$$C_n = \frac{\pi^{\frac{n}{2}}}{\left(\frac{n}{2}\right)!},$$

Unit n-ball even volume:

and, since $\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$,

for n-odd $C_n = \dfrac{2^{\frac{n+1}{2}}\pi^{\frac{n-1}{2}}}{n!!}$

$$S_{n-1} = \frac{dV_n}{dR} = \frac{nV_n}{R} = \frac{2\pi^{n/2}R^{n-1}}{\Gamma(\frac{n}{2})} = nC_nR^{n-1}.$$

n − 1 dimensional volume of the n−1 sphere at the boundary of the n-ball

The Degree Distribution of a Network

Networks may be *directed* or *undirected*

The *degree distribution* (i.e. the number of connections or edges a given node has to other nodes) is very important for studying networks

The simplest (*random*)graphs have binomial degree distributions

Binomial

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

where

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

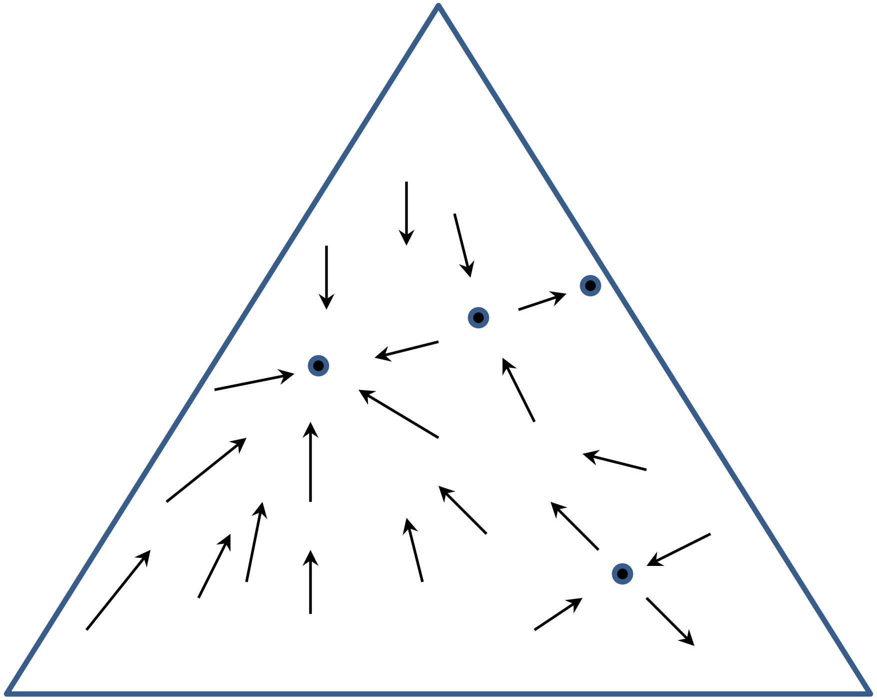
Exponential

$$p(k) = e^{-k/c}$$

Scale-free graphs have *power-law* degree distributions (or by extension *multi-fractal distributions*), the fraction P(k) of nodes in the network having k connections to other nodes becomes large as:

$$P(k) \sim c k^{-\gamma}$$

Path Dependence and Spatial Location



Expected Motions for a Locational Probability Function