A gentle introduction to combinatorial stochastic processes II

Enrico Scalas

Department of Mathematics, University of Sussex, UK

Stochastic Models for Complex Systems
10 June - 7 July, 2021

Outline

Discrete stochastic processes

Consider a finite sequence of discrete random variables $X_1, \ldots, X_n$ assuming values in g categories; the label n can be any positive integer such that $n \geq 1$ and the index has usually but not necessarily the meaning of time or time ordering (consider tossing n coins simultaneously vs tossing a single coin n times). The event $\{X_1 = x_1, \ldots, X_n = x_n\}$ is a *realisation* of the discrete stochastic process $(X_n)_{n \geq 1}$.

Remark

Note that it is possible to start n from 0*, etc..*

Remark

$\{X_1 = x_1, \ldots, X_n = x_n\} := \{X_1 = x_1\} \cap \cdots \cap \{X_n = x_n\}.$

Finite dimensional distributions

The process is fully characterised if, for any n, we know the so-called finite dimensional distributions

$$p(x_1, \dots, x_n) = \mathbb{P}(X_1 = x_1, \dots, X_n = x_n). \tag{1}$$

An important case are processes built with dichotomous random variables ($g = 2$). It is customary to choose one of the outcomes as a *success* and use the label 1 and the other outcome as a *failure* and use the label 0. In other words, if the outcome T denotes a success and its complement (say H) a failure, $X_i = \mathbb{1}_T$. For example, for the toss of 3 coins with $\mathbb{P}(\{T\}) = p$, one has

$$p(1,0,1) = \mathbb{P}(X_1 = 1, X_2 = 0, X_3 = 1) = p \cdot (1-p) \cdot p = p^2(1-p).$$

Compatibility conditions

The functions $p(x_1, \ldots, x_n)$ obey the following compatibility conditions

1. symmetry: $p(x_{i_1}, \ldots, x_{i_n}) = p(x_1, \ldots, x_n)$, where $i_1, \ldots, i_n$ denotes any of the $n!$ permutations of the indices $1, \ldots, n$.
2. marginalisation: $p(x_1, \ldots, x_{n-1}) = \sum_{x_n \in \{1,\ldots,g\}} p(x_1, \ldots, x_n)$.

The first property is a consequence of the symmetry of the intersection of sets/events, the second is a consequence of the finite additivity axiom of Kolmogorov (or of infinite additivity, if you prefer).

Predictive probabilities I

By repeated application of the definition of conditional probability one finds this relation

$$p(x_1, \dots, x_n) = \mathbb{P}(X_1 = x_1, \dots, X_n = x_n) = \mathbb{P}(X_1 = x_1)\mathbb{P}(X_2 = x_2|X_1 = x_1)\dots \mathbb{P}(X_n = x_n|X_1 = x_1, \dots, X_{n-1} = x_{n-1}). \quad (2)$$

Hence, often a stochastic process is characterized in terms of the predictive law $\mathbb{P}(X_m = x_m|X_1 = x_1, \dots, X_{m-1} = x_{m-1})$ for any $m > 0$ and of the “initial” distribution $\mathbb{P}(X_1 = x_1)$. This characterisation automatically obeys the above compatibility conditions.

Predictive probabilities II

The predictive probability has only to fullfill the probability axioms: $\mathbb{P}(X_m = x_m|\mathbf{X}^{(m-1)} = \mathbf{x}^{(m-1)}) \geq 0$, and $\sum_{x_m \in \{1,\ldots,g\}} \mathbb{P}(X_m = x_m|\mathbf{X}^{(m-1)} = \mathbf{x}^{(m-1)}) = 1$. In other words, one has to fix the probability of the "next" observation as a function of all the previous results (the history of the process). Predictive probability laws that are simple functions of x_m and $\mathbf{x}^{(m-1)}$ are particularly interesting.

Exchangeability I

If the finite dimensional probabilities are such that, for any integer $m = 2, \ldots, n$, one has

$$\mathbb{P}(X_1 = x_1, \ldots, X_m = x_m) = \mathbb{P}(X_{i_1} = x_1, \ldots, X_{i_m} = x_m) \qquad (3)$$

for any of the $m!$ possible permutation of the indices, then the stochastic process is called *exchangeable* (or, more precisely, n-exchangeable). Note that this condition differs from the symmetry compatibility condition. The meaning of exchangeability is that all the individual sequences corresponding to the same occupation vector $\mathbf{m}$ have the same probability.

Exchangeability II

Consider an exchangeable process $X_1, \dots, X_n$ whose range is the label set $\{1, \dots, g\}$. We know that there are g^n possible sequences $X_1 = x_1, \dots, X_n = x_n$, whereas the number of possible occupation vectors $\mathbf{n} = (n_1, \dots, n_g)$ is given by

$$\binom{n+g-1}{n}. \tag{4}$$

Moreover, given a particular occupation vector $\mathbf{Y} = \mathbf{n}$, the number of corresponding individual sequences $\mathbf{X}^{(n)}$ is

$$\frac{n!}{n_1! \cdots n_g!}. \tag{5}$$

Exchangeability III

A consequence of definition (3) and of finite additivity is that

$$\mathbb{P}(\mathbf{X}^{(m)} = \mathbf{x}^{(m)}) = \mathbb{P}(\mathbf{Y} = \mathbf{m}) \left(\frac{m!}{m_1! \cdots m_g!} \right)^{-1} \quad (6)$$

where a shortcut notations has been used $\mathbf{Y} = \mathbf{m}$ is the frequency vector of the particular random vector $\mathbf{X}^{(m)} = \mathbf{x}^{(m)}$.
Equation (6) is compatible with any probability distribution on $\mathbf{m}$, $\{\mathbb{P}(\mathbf{m}) : \sum \mathbb{P}(\mathbf{m}) = 1, \sum_{i=1}^{g} m_i = m\}$. It turns out that $(n + g - 1)!/(n!(g-1)!) - 1$ parameters are needed to completely define the probability space. They represent the probability of all possible compositions of the population with respect to the g labels (categories).

The hypergeometric process I

The physical process of drawing balls without replacement at random and from an urn of given composition $\mathbf{n} = (n_1, \ldots, n_g)$ can be described by the following stochastic process. Note that the state space $\{1, \ldots, g\}$ denotes the g different colours present in the urn. Incidentally, this urn model gives a stylised, but general, representation of *random sampling*. Consider a sequence of individual random variables $X_1, \ldots, X_n$ whose range is the label set $\{1, \ldots, g\}$ and characterized by the following predictive probability (for $m \leq n$, the process stops at $m = n$)

$$\mathbb{P}(X_{m+1} = j | X_1 = x_1, \ldots, X_m = x_m; \mathbf{n}) = \mathbb{P}(X_{m+1} = j | \mathbf{m}; \mathbf{n}) = \frac{n_j - m_j}{n - m}. \quad (7)$$

The hypergeometric process II

Based on the rule (7), one can find the probability of the fundamental sequence $\mathbf{X}_f^{(m)} = (X_1 = 1, \ldots, X_{m_1} = 1, X_{m_1+1} = 2, \ldots X_{m_1+m_2} = 2, \ldots, X_{m-m_g+1} = g, \ldots, X_m = g)$ with $m \leq n$ balls drawn, of which the first m_1 of colour 1, the following m_2 of colour 2, and so on, until the last set of m_g balls of colour g, with $\sum_{i=1}^{g} m_i = m$. By repeated applications of (7), one gets

$$\mathbb{P}(\mathbf{X}_f^{(m)}|\mathbf{m};\mathbf{n}) = \frac{n_1}{n}\frac{n_1-1}{n-1}\cdots\frac{n_1-m_1-1}{n-m_1-1}\frac{n_2}{n-m_1}\frac{n_2-1}{n-m_1-2}\cdots\frac{n_2-m_2-1}{n_2-m_1-m_2-1}\times \frac{n_g}{n-m+m_g}\frac{n_g-1}{n-m+m_g-1}\cdots\frac{n_g-m_g-1}{n-m-1} \tag{8}$$

a result that can be written in a more compact form as follows

$$\mathbb{P}(\mathbf{X}_f^{(m)}|\mathbf{m};\mathbf{n}) = \frac{(n-m)!}{n!}\prod_{i=1}^{g}\frac{n_i!}{(n_i-m_i)!}. \tag{9}$$

The hypergeometric process III

Due to exchangeability, the nice point about the probability given by equation (9) is that it is the same for any individual sequence with the sampling vector **m**, and the number of these sequences is given by the multinomial factor; therefore one has

$$\mathbb{P}(\mathbf{m};\mathbf{n}) = \frac{m!}{\prod_{i=1}^{g} m_i!}\mathbb{P}(\mathbf{X}_f^{(m)}|\mathbf{m};\mathbf{n}) \tag{10}$$

leading to the hypergeometric sampling distribution

$$\mathbb{P}(\mathbf{m};\mathbf{n}) = \frac{m!}{\prod_{i=1}^{g} m_i!}\frac{(n-m)!}{n!}\prod_{i=1}^{g}\frac{n_i!}{(n_i-m_i)!} = \frac{\prod_{i=1}^{g}\binom{n_i}{m_i}}{\binom{n}{m}}. \tag{11}$$

The hypergeometric process is the simplest case of *n-exchangeable process*. The shortcut notation has the following meaning $\mathbb{P}(\mathbf{m};\mathbf{n}) = \mathbb{P}(\{\mathbf{Y}^{(m)} = \mathbf{m}\}|\mathbf{Y}^{(n)} = \mathbf{n})$ and $\mathbf{Y}^{(n)} = \mathbf{n}$ is the exact composition of the urn appearing as a parameter in the distribution.

Finite version(s) of de Finetti's theorem

It turns out that the hypergeometric process has an important role, due to the finite version of de Finetti representation theorem.

Theorem

Any n-exchangeable stochastic process can be written as a mixture of hypergeometric processes.

A good discussion on this result is in P. Diaconis, *Finite Forms of de Finetti's Theorem on Exchangeability*, Synthese **36** 271-281, 1977.

Remark

For processes that do not terminate and can be extended for $n \to \infty$, the usual de Finetti representation theorem can be derived from results on finite processes.

The Pólya process I

The sequence $X_1, X_2, \ldots, X_n$, with $X_i \in \{1, \ldots, g\}$ is a generalized Pólya process if the predictive probability of the process is given by

$$\mathbb{P}(X_{m+1} = j | X_1 = x_1, \ldots, X_m = x_m) = \frac{\alpha_j + m_j}{\alpha + m} \tag{12}$$

where $m_j = \#\{X_i = j, i = 1, \ldots, m\}$, the number of occurrences of the j-th category in the evidence $(X_1 = x_1, \ldots, X_m = x_m)$, and m is the number of observations or trials. As for the parameters α and α_j, in the usual Pólya urn process, α_j is a positive integer, representing the number of balls of type (colour) j in the auxiliary Pólya urn; α is the total number of balls in the urn and it is given by $\alpha = \sum_{j=1}^{g} \alpha_j$.

The Pólya process II

If α_j is positive but real, it can be interpreted as the initial weight of the j-th category. The meaning of α_j becomes clearer if one defines $p_j = \alpha_j/\alpha$, then equation (12) can be written as

$$\mathbb{P}(X_{m+1} = j | X_1 = x_1, \ldots, X_m = x_m) = \frac{\alpha p_j + m_j}{\alpha + m} \qquad (13)$$

and p_j has a natural interpretation as prior (or initial) probability for the j-th category: $p_j = \alpha_j/\alpha = \mathbb{P}(X_1 = j)$.

Sampling with replacement

Remark

The limit $\alpha \to \infty$ has an immediate intepretation as (13) becomes

$$\mathbb{P}(X_{m+1} = j | X_1 = x_1, \ldots, X_m = x_m) = p_j, \tag{14}$$

leading to the i.i.d. multinomial distribution. In other words, the distribution of the generalized Pólya process converges to the distribution of the multinomial process when the parameter α diverges.

Sampling without replacement

Remark

When α_j is a negative integer and $m_j \leq |\alpha_j| = |\alpha| p_j$, equation (12) still represents a probability and it can be written as

$$\mathbb{P}(X_{m+1} = j | X_1 = x_1, \ldots, X_m = x_m) = \frac{|\alpha| p_j - m_j}{|\alpha| - m}, \tag{15}$$

but here, the number of variables (observations or trials) is limited by $|\alpha|$. Equation (15) leads to the hypergeometric distribution. Therefore, in the following, the general case (12) will be studied, with the total weight α belonging to the set

$$\alpha \in (0, \infty) \cup \{-n, -n-1, -n-2, \ldots, -\infty\}$$

which is the union of the generalized Pólya and the hypergeometric domain, with the multinomial case appearing both as $\alpha \to -\infty$ and $\alpha \to +\infty$.

Finite dimensional distributions for the Pólya process I

The Pólya process is exchangeable. The fundamental sequence $\mathbf{X}_f^{(m)} = (1, \ldots, 1, \ldots, g, \ldots, g)$, consisting of m_1 labels 1 followed by m_2 labels 2, and so on, ending with m_g occurrences of label g, has probability

$$\frac{\alpha_1}{\alpha}\frac{\alpha_1+1}{\alpha+1}\cdots\frac{\alpha_1+m_1-1}{\alpha+m_1-1}\frac{\alpha_2}{\alpha+m_1}\frac{\alpha_2+1}{\alpha+m_1+1}\cdots\frac{\alpha_2+m_2-1}{\alpha+m_1+m_2-1}\cdots$$

$$\cdots\frac{\alpha_g}{\alpha+m_1+m_2+\ldots+m_{g-1}}\frac{\alpha_g+1}{\alpha+m_1+m_2+\ldots+m_{g-1}+1}\cdots$$

$$\cdots\frac{\alpha_g+m_g-1}{\alpha+m_1+m_2+\ldots+m_g-1} = \frac{\alpha_1^{[m_1]}\cdots\alpha_g^{[m_g]}}{\alpha^{[m]}}, \tag{16}$$

as a direct consequence of equations (2) and (12) and using the Pochhammer symbol $x^{[n]}$ to denote the rising factorial $x(x+1)\cdots(x+n-1)$ (recall that $\sum_{i=1}^{g} m_i = m$).

Finite dimensional distributions for the Pólya process II

In summary, the finite dimensional distributions only depend on the frequency vector **m**. They are

$$\mathbb{P}(\mathbf{x}^{(m)}) = \frac{1}{\alpha^{[m]}} \prod_{i=1}^{g} \alpha_i^{[m_i]} \quad (17)$$

So the predictive probability (12) yields the law (17); it turns out that the converse is also true, as from the definition of conditional probability

$$\mathbb{P}(X_{m+1} = j | \mathbf{x}^{(m)}) := \frac{\mathbb{P}(\mathbf{x}^{(m)}, X_{m+1} = j)}{\mathbb{P}(\mathbf{x}^{(m)})} = \frac{\alpha_j + m_j}{\alpha + m} \quad (18)$$

where, $\mathbf{x}^{(m)}$ is an evidence with frequency vector $\mathbf{m} = (m_1, \ldots, m_g)$, both the numerator and the denominator follow (17).

Finite dimensional distributions for frequency descriptions

Remark

The sampling distribution is the probability of the frequency vector $\mathbf{m}$*. In order to get it as a function of the size* m*, it is sufficient to multiply (17) by the number of distinct equiprobable sequences leading to*

$$\mathbb{P}(\mathbf{m}) = \mathrm{Polya}(\mathbf{m}; \boldsymbol{\alpha}) = \frac{m!}{\prod_{i=1}^{g} m_i!} \frac{\prod_{i=1}^{g} \alpha_i^{[m_i]}}{\alpha^{[m]}} = \frac{m!}{\alpha^{[m]}} \prod_{i=1}^{g} \frac{\alpha_i^{[m_i]}}{m_i!}; \qquad (19)$$

this equation defines the generalised Pólya distribution of parameters m *and* $\boldsymbol{\alpha}$*.*

Example: The dichotomous Pólya distribution

The dichotomous (or bivariate) case, $g = 2$, is sufficient to illustrate all the essential properties of the Pólya process. Let $X_i = 0, 1$, and define $(m_1 = h, m_0 = m - h)$ as the frequency vector of $X_1, \ldots, X_m$. Then $h = \#\{X_i = 1\}$ is the number of observed successes and $S_m = h$ is a random variable whose values h vary according to $h \in \{0, 1, \ldots, m\}$. Moreover, one can write $S_m = \sum_{i=1}^{m} X_i$. In this simple case the multivariate equation (19) simplifies to

$$\mathbb{P}(S_m = h|\alpha_0, \alpha_1) = \frac{m!}{m_0!m_1!} \frac{\alpha_0^{[m_0]}\alpha_1^{[m_1]}}{\alpha^{[m]}} = \binom{m}{h} \frac{\alpha_1^{[h]}\alpha_0^{[m-h]}}{\alpha^{[m]}}. \quad (20)$$

Uniform distribution

One has $1^{[n]} = n!$ and $2^{[n]} = (n+1)!$; if $\alpha_0 = \alpha_1 = 1$, then $\alpha = \alpha_0 + \alpha_1 = 2$, and it turns out that

$$\mathbb{P}(S_m = h|1,1) = \frac{m!}{h!(m-h)!}\frac{h!(m-h)!}{(m+1)!} = \frac{1}{m+1}, \qquad (21)$$

this is the uniform distribution on all frequency vectors of m elements and $g = 2$ categories.

Binomial distribution

For $x \gg n$, one has $x^{[n]} = x(x+1)\dots(x+n-1) \simeq x^n$; moreover, one can set $\alpha_0 = \alpha p_0$, and $\alpha_1 = \alpha p_1$. Therefore, one obtains that

$$\lim_{\alpha\to\infty} \mathbb{P}(S_m = h|\alpha p_0, \alpha p_1) =$$

$$\lim_{\alpha\to\infty} \binom{m}{h} \frac{\alpha_1^h \alpha_0^{m-h}}{\alpha^m} = \binom{m}{h} p_1^h (1-p_1)^{m-h}, \quad (22)$$

this is the binomial distribution. In the urn interpretation, if the initial urn has a very large number of balls, the Pólya prize is inessential to calculate the predictive probability, they are all equal for each category and the Pólya process cannot be distinguished from repeated Bernoulli trials, as already discussed above.

Hypergeometric distribution

Given that $x^{[n]} = x(x+1)\ldots(x+n-1)$, if x is negative, and $n \leq |x|$, one gets $x^{[n]} = (-1)^n|x|(|x|-1)\ldots(|x|-n+1)$, meaning that $x^{[n]} = (-1)^n\,|x|_{[n]}$, where $x_{[n]} = x(x-1)\ldots(x-n+1)$ is the so-called *falling (or lower) factorial*, and setting $\alpha_1 = -N_1 = -Np_1$, $\alpha_0 = -N_0 = -Np_0$, with $p_1 + p_0 = 1$, one gets:

$$\mathbb{P}(S_m = h| - N_0, -N_1) = \frac{m!}{h!(m-h)!}\times$$
$$\times\frac{(-1)^h N_1\cdots(N_1-h+1)(-1)^{m-h}N_0\cdots(N_1-n+h+1)}{(-1)^m N\cdots(N-m+1)}, \quad (23)$$

leading to the hypergeometric distribution

$$\mathbb{P}(S_m = h| - N_0, -N_1) = \frac{\binom{N_1}{h}\binom{N_0}{m-h}}{\binom{N}{m}} = \binom{m}{h}\frac{\binom{N-m}{N_1-h}}{\binom{N}{N_1}} \quad (24)$$

which is the usual sampling distribution from a hypergeometric urn with N_0, N_1 balls of two types and $N = N_1 + N_0$.

Limit of the hypergeometric distribution

Also for the falling factorial $x \gg n$ yields $x_{[n]} \simeq x^n$. Therefore, one gets

$$\lim_{\alpha \to -\infty} \mathbb{P}(S_m = h | \alpha p_0, \alpha p_1) = \binom{m}{h} p_1^h (1 - p_1)^{m-h}; \qquad (25)$$

in other words, sampling without replacement from a very large urn cannot be distinguished from sampling with replacement.

The continuous limit of the Pólya distribution I

Consider the multivariate generalized Pólya sampling distribution given by equation (19). Using the fact that

$$\alpha^{[m]} = \frac{\Gamma(m+\alpha)}{\Gamma(\alpha)} \tag{26}$$

equation (19) can be re-written as follows

$$\mathrm{Polya}(\mathbf{m};\boldsymbol{\alpha}) = \frac{\Gamma(\alpha)}{\prod_{i=1}^{g}\Gamma(\alpha_i)}\frac{m!}{\Gamma(m+\alpha)}\prod_{i=1}^{g}\frac{\Gamma(m_i+\alpha_i)}{m_i!}. \tag{27}$$

The variables $u_i = m_i/m$ are such that one has

$$\sum_{i=1}^{g} u_i = \sum_{i=1}^{g}\frac{m_i}{m} = 1; \tag{28}$$

moreover, $\forall i \in \{1,\ldots,g\}$, one further has that $0 \leq u_i \leq 1$. If one considers the *continuous* limit in which $m \to \infty$, $m_i \to \infty$ with constant $u_i = m_i/m$ for all the categories i, one gets that

$$\frac{\Gamma(m_i+\alpha_i)}{m_i!} = \frac{\Gamma(m_i+\alpha_i)}{\Gamma(m_i+1)} \simeq m_i^{\alpha_i-1}. \tag{29}$$

The continuous limit of the Pólya distribution II

Replacing (29) for any m_i and for m in (27) leads to

$$\mathrm{Polya}(\mathbf{m};\boldsymbol{\alpha}) \simeq \frac{\Gamma(\alpha)}{\prod_{i=1}^{g}\Gamma(\alpha_i)}\frac{\prod_{i=1}^{g} m_i^{\alpha_i-1}}{m^{\alpha-1}} = \frac{\Gamma\left(\sum_{i=1}^{g}\alpha_i\right)}{\prod_{i=1}^{g}\Gamma(\alpha_i)}\prod_{i=1}^{g} u_i^{\alpha_i-1}\cdot\frac{1}{m^{g-1}}. \quad (30)$$

The continuous limit of the Pólya distribution III

Equation (30) can be intepreted as follows; based on the exchangeability of the variables $Y_i = m_i$, the probability of the variables $U_i = Y_i/m$ of assuming values $U_1 = u_1, \ldots U_n = u_n$ with $u_i = m_i/m$ is

$$\mathbb{P}(U_1 = u_1, \ldots, U_n = u_n) \simeq \frac{\Gamma\left(\sum_{i=1}^{g} \alpha_i\right)}{\prod_{i=1}^{g} \Gamma(\alpha_i)} \prod_{i=1}^{g} u_i^{\alpha_i - 1} \cdot \frac{1}{m^{g-1}} \simeq \frac{\Gamma\left(\sum_{i=1}^{g} \alpha_i\right)}{\prod_{i=1}^{g} \Gamma(\alpha_i)} \prod_{i=1}^{g} u_i^{\alpha_i - 1} du_1 \cdots du_{g-1}, \quad (31)$$

where the relationship becomes exact in the continuous limit. The function

$$f_{\mathbf{U}}(u_1, \ldots, u_g; \alpha_1, \ldots \alpha_g) = f_{\mathbf{U}}(\mathbf{u}; \boldsymbol{\alpha}) = \frac{\Gamma\left(\sum_{i=1}^{g} \alpha_i\right)}{\prod_{i=1}^{g} \Gamma(\alpha_i)} \prod_{i=1}^{g} u_i^{\alpha_i - 1} \quad (32)$$

defined on the simplex $\sum_{i=1}^{g} u_i = 1$ and $0 \leq u_i \leq 1$ for all the $i \in \{1, \ldots g\}$ is the probability density function of the *Dirichlet distribution*. Let $\mathbf{U} \sim \mathrm{Dir}(\mathbf{u}; \boldsymbol{\alpha})$ denote the fact that the random vector $\mathbf{U}$ is distributed according to the Dirichlet distribution.

Aggregation property of the Pólya distribution I

The g-variate generalized Pólya distribution has a nice aggregation property which is useful for the so-called *label mixing* by R. von Mises. Starting from (12), how can $\mathbb{P}(X_{m+1} \in A | X_1 = x_1, \ldots, X_m = x_m)$ be obtained, where the set A is a set of categories $A = \{j_1, \ldots, j_r\}$? The answer to this question is not difficult: it turns out that

$$\mathbb{P}(X_{m+1} \in A | \mathbf{x}^{(m)}) = \sum_{i=1}^{r} \mathbb{P}(X_{m+1} = j_i | \mathbf{x}^{(m)}), \tag{33}$$

where, as usual, $\mathbf{x}^{(m)} = (X_1 = x_1, \ldots, X_m = x_m)$ summarizes the evidence. In the Pólya case, $\mathbb{P}(X_{m+1} = j | \mathbf{x}^{(m)})$ is a linear function of both the weights and the occupation numbers; therefore one gets

$$\mathbb{P}(X_{m+1} \in A | \mathbf{x}^{(m)}) = \frac{\sum_{j \in A} \alpha_j + \sum_{j \in A} m_j}{\alpha + m} = \frac{\alpha_A + m_A}{\alpha + m}, \tag{34}$$

where $\alpha_A = \sum_{j \in A} \alpha_j$ and $m_A = \sum_{j \in A} m_j$.

Aggregation property of the Pólya distribution II

As a direct consequence of (34), the marginal distributions of the g-variate generalized Pólya distribution are given by the dichotomous Pólya distribution of weights α_i and $\alpha - \alpha_i$, where i is the category with respect to which the marginalization is performed. In other words, one gets that

$$\sum_{m_j,\, j \neq i} \mathrm{Polya}(\mathbf{m}; \boldsymbol{\alpha}) = \mathrm{Polya}(m_i, m - m_i; \alpha_i, \alpha - \alpha_i). \qquad (35)$$

Aggregation property of the Dirichlet distribution

The aggregation property is inherited by the Dirichlet distribution. In particular, the probability density function for the one-point marginal is nothing else than the Beta distribution. In particular, if $U_1, \ldots, U_g \sim \mathrm{Dir}(u_1, \ldots, u_g; \alpha_1, \ldots, \alpha_g)$ then

$$U_i \sim \mathrm{Beta}(u_i; \alpha_i, \alpha - \alpha_i) \tag{36}$$

corresponding to the probability density function

$$f_{U_i}(x_i) = \frac{\Gamma(\alpha)}{\Gamma(\alpha_i)\Gamma(\alpha - \alpha_i)} u_i^{\alpha_i - 1}(1 - u_i)^{\alpha - \alpha_i - 1}. \tag{37}$$

Johnson's representation theorem I

Consider an n-step stochastic process defined by the sequence of n random variables $X_1, \ldots, X_n$ with values in g categories.

Definition

The *coefficient of heterorelevance* is defined as

$$Q_i^j(X_1 = x_1, \ldots, X_m = x_m) = \frac{\mathbb{P}(X_{m+2} = i | X_1 = x_1, \ldots, X_m = x_m, X_{m+1} = j)}{\mathbb{P}(X_{m+1} = i | X_1 = x_1, \ldots, X_m = x_m)}. \quad (38)$$

The meaning of this coefficient is the ratio between the probability of observing category i at step $m+2$ given the history known up to m observations and knowing that at the $m+1$-th step category $j \neq i$ was observed and the probability of observing category i at step $m+1$ given the history known up to m observations.

Johnson's representation theorem II

Theorem

Assume that $Q_i^j(X_1 = x_1, \ldots, X_m = x_m) = Q_m$, meaning that the coefficient of heterorelevance only depends on the number of observations m. Then

$$\mathbb{P}(X_{m+1} = j | X_1 = x_1, \ldots, X_m = x_m) = p(j|m_j, m) = \frac{\lambda p_j + m_j}{\lambda + m}, \tag{39}$$

where

$$\lambda = \frac{Q_0}{1 - Q_0} \tag{40}$$

and

$$p_j = \mathbb{P}(X_1 = j). \tag{41}$$

Proofs of the theorem

A proof of Johnson's theorem can be found in U. Garibaldi, E. Scalas, *Finitary Probabilistic Methods in Econophysics*, Cambridge University Press, 2010. A discussion of its meaning and history together is in S. Zabell, *The Continuum of Inductive Methods Revisited* in *Symmetry and its Discontents*, Cambridge University Press, 2012. An alternative proof is in S. Zabell, *Confirming Universal Generalizations*, Erkentniss, **45** (2/3), 1996.

Johnson's representation theorem II

Theorem

Assume that $Q^j_i(X_1 = x_1, \ldots, X_m = x_m) = Q_m$, meaning that the coefficient of heterorelevance only depends on the number of observations m. Then

$$\mathbb{P}(X_{m+1} = j | X_1 = x_1, \ldots, X_m = x_m) = p(j|m_j, m) = \frac{\lambda p_j + m_j}{\lambda + m}, \tag{42}$$

where

$$\lambda = \frac{Q_0}{1 - Q_0} \tag{43}$$

and

$$p_j = \mathbb{P}(X_1 = j). \tag{44}$$

Sampling and quantum statistics

Let me conclude with a remark on quantum statistics

Remark

There is a correspondence between sampling procedures and quantum statistics.

1. *Pólya sampling "corresponds" to Bose-Einstein statistic.*
2. *Sampling without replacement "corresponds" to Fermi-Dirac statistic.*
3. *Sampling with replacement "corresponds" to Maxwell-Boltzmann statistic.*

Do you understand why?

Exercise 1: Pólya random walk

Consider a dichotomous polya process with parameters α_0 and α_1 and define the random variable $Y(m) = +1$ if $X(m) = 1$ and $Y(m) = -1$ if $X(m) = 0$. In other words, one has

$$Y = 2X - 1. \tag{45}$$

Then define the random variable $S(m)$ (Pólya random walk) as

$$S(m) = \sum_{i=1}^{m} Y(i). \tag{46}$$

Write a program that simulates the Pólya random walk and discuss some interesting cases.

Exercise 2: Binomial limit

We have seen the the dichotomous Pólya distribution converges to the binomial distribution for $\alpha \to \infty$. But how close is the symmetric dichotomous Pólya to the binomial distribution when $\alpha_0 = \alpha_1 = 100$? Try to answer this question by means of a Monte Carlo simulation.

Exercise 3: Birthday problem

A classical problem in probability theory is often stated as follows. Imagine there are n persons in a room, with $n < 365$. What is the probability that at least two of them have a common birthday? This problem and indeed other problems related to the birthday distribution can be solved using the concepts we have introduced so far. How? Usually, it is assumed that people born on February 29th in leap years are not taken into account. For the sake of simplicity, you can make this assumption in your solution.