On the generating process and the class typicality measure

Oleg Golubitsky

January 24, 2002

Abstract

In this paper, we consider the stochastic generating process—one of the key concepts of the Evolving Transformation System model [1]—from the formal perspective. First, we give an informal definition of the generating process supported by some intuitive assumptions and consider several examples. Then, we formally define the concept of the generating process as a continuous parameter (c.p.) Markov chain. Some important random variables associated with this c.p. Markov chain are introduced next, followed by the definition of the typicality measure. Two methods for the computation of the typicality measure are proposed. In conclusion, we discuss the problem of compactification of the state space for the c.p. Markov chain. This problem is not only interesting from the points of view of topology and of the c.p. Markov chains theory, but also has important implications for the ETS model, since it is related to the problem of class comparison and to the proper formulation of the learning problem.

1 Introduction

The concepts of generating process and typicality measure [1, Defs. 33–35] are among the key concepts that constitute the formal foundation of the Evolving Transformation System (ETS) model proposed in [1]. The ETS model is a model of "structural", or "symbolic", object representation. The basic assumption of the model is that the object's formative (or evolutionary) history is an integral part of the object's representation. This formative history is viewed as a sequence of states represented as *structs* in the ETS model. Each next struct is obtained from the preceding one by applying a *context-dependent struct transformation*. The choice of a particular transformation to be applied to a state is controlled by a stochastic *generating process*, which is specified by a *transformation system*. An object's representation in the ETS model consists of a struct—the final state in the above sequence—and a transformation system which specifies a process that can generate this struct. Thus, given an object's represnetation, the set of all possible evolutionary (historical and future) paths for this object can be deduced.

The concepts of transformation system and generating process give rise to the notion of a class of objects; elements of the class are structs that can be generated by the class generating process. Thus, the description of the class to which an object belongs is an integral part of the object's represnetation which determines the possible hisories of the object's construction, the possible paths of the object's future evolution, and the set of objects (class elements) that are related to the object by their evolutionary histories.

The stochastic generating process induces a measure on the set of class elements, which we will call the class typicality measure. Essentially, a struct's typicality is the probability that an observer that makes an observation of the process at a random moment will see this struct as the current state of the generating process.

In this paper, we will assume that the reader is familiar with the notions of inductive structure and transformation system (see [1] for the corresponding formal definitions).

2 Informal definition of the generating process

Let an inductive structure $(\Pi, \mathcal{I})$ [1, Def. 20] be fixed. Let $T = \{\bar{\boldsymbol{\tau}}_1, \dots, \bar{\boldsymbol{\tau}}_n\}$ be a transformation set [1, Def. 22], and let $\mathbf{TS} = (T, l, \bar{\boldsymbol{\kappa}})$ be a transformation system [1, Def. 30], where $l : T \to \mathbb{R}_+$ is the weighting mapping, and struct $\bar{\boldsymbol{\kappa}} \in \bar{\Theta}$ is the progenitor. We denote by P_T and EP_T the set of all paths and elementary paths, respectively, generated by the transformation set T [1, Def. 25]. For a path $p \in P_T$, $\text{begin}(p)$ and $\text{end}(p)$ denote the beginning and the end of p, respectively [1, Def. 25].

The following informal definition of the generating process is taken from [1].

Definition 1. For a transformation system $\mathbf{TS} = (T, l, \bar{\boldsymbol{\kappa}})$, the **generating process** $G_{\mathbf{TS}}$ (or simply G) is a c.p. Markov chain (or, equivalently, a stochastic Markov process with a countable state space) defined as follows:

1. The states of G are elements of the set TS of structs generated by transformation system $\mathbf{TS}$.
2. The initial state of G is progenitor struct $\bar{\boldsymbol{\kappa}}$.
3. The amount of time which G spends in state $\bar{\boldsymbol{\gamma}}$ is a random variable distributed exponentially with mean
$$L = \frac{1}{\sum\limits_{[\boldsymbol{\gamma};\boldsymbol{\tau}] \in EP_T} 1/l(\bar{\boldsymbol{\tau}})} \,.^{1} \qquad (*)$$
4. When G leaves state $\bar{\boldsymbol{\gamma}}$, it chooses randomly an elementary path $[\boldsymbol{\gamma}; \boldsymbol{\tau}] \in EP_T$ with probability
$$\frac{L}{l(\bar{\boldsymbol{\tau}})} \qquad (**)$$
and enters state $[\boldsymbol{\gamma} \lhd \boldsymbol{\tau}]$.
5. All random variables in 2 and 3 are mutually independent.

[1] Here $[\boldsymbol{\gamma}; \boldsymbol{\tau}]$ denotes an arbitrary elementary path beginning in struct $\bar{\boldsymbol{\gamma}}$, whose transformation is $\bar{\boldsymbol{\tau}}$. See also [1] for the difference between instance struct $\boldsymbol{\gamma}$ and struct $\bar{\boldsymbol{\gamma}}$, as well as instance transformation $\boldsymbol{\tau}$ and transformation $\bar{\boldsymbol{\tau}}$.

►

The justification of why and under which assumptions conditions 1–5 in the above definition uniquely specify a c.p. Markov chain is postponed until Section 5. Here we would like to state explicitly the assumptions on the generating process which have lead to the above definition.

First, why the generating process is chosen to be a stochastic process. There are two reasons for this. First, that the process has to be able to generate *all* elements of the set TS—the set of structs generated by the transformation system **TS** [1, Def. 31]. This set is typically quite large and has a complex structure, therefore one would need an enormous amount of information (even unlikely to be finite) to specify a *deterministic* process that would still generate the whole set TS. The second reason, related to the first one, is that, in nature, the choice of the next step of the generating process, as well as the time this step takes, depends not only on the transformation system itself but also on a variety of external factors the influence of which cannot be taken into account. These factors are considered as random, thus leading to a *probabilistic* interpretation.

Second, why the random amount of time that the process spends in a state until a transformation is applied is an exponentially distributed random variable. This is because we consider the elementary paths and their transformations to be *indivisible*. I.e., an elementary path is either completed which means that the transformation is applied and the current state of the process has changed, or not completed, in which case no changes are made to the struct being the current state of the process. This implies that if ξ is the amount of time the process spends in a state, then for all $T, t > 0$,

$$\mathcal{P}(\xi < T + t \,|\, \xi > T) = \mathcal{P}(\xi < t).$$

The above property means that the system is memoryless, and implies that ξ has an exponential distribution [2, Chapter I].

Third, the transformations that are applicable to the current state $\bar{\gamma}$ of the process are considered to be independent of each other. Thus, the random variables $\xi_1, \ldots, \xi_k$ which denote the waiting times for the elementary paths beginning in $\bar{\gamma}$ to be passed are mutually independent. The process remains in state $\bar{\gamma}$ until one of the transformations from these elementary paths is applied, thus the time the process spends in $\bar{\gamma}$ is the random variable $\xi = \min(\xi_1, \ldots, \xi_k)$. The minimum of mutually independent exponentially distributed random variables is an exponentially distributed random variable with mean $(*)$ [2, Chapter 1]. Also, the probability that a particular random random variable ξ_i will have the minimal value among $\{\xi_1, \ldots, \xi_k\}$, is proportional to the inverse of expectation of ξ_i, hence formula $(**)$ holds.

Finally, we assume that all random variables are mutually independent, since we postulate that the choice of the transformation that is applied to a particular struct and the time of its application are independent of the particular path that has lead to the struct and of the time it has taken to generate the struct. The latter assumption is due to the fact that a struct represents a set of construction paths which are indistinguishable (or equivalent) from the point of view of applicability of transformations.

3 Examples

In this section, we present three examples of the generating processes: the processes which generate the set of natural numbers, the set of binary sequences representing binary fractions on the segment $[0, 1]$, and the set of strings over a two-letter alphabet. In Section 9, we will compute the typicality measures induced by these processes on the corresponding sets of structs. In Conclusion, we will refer to these examples in connection with the questions about compactification and topology of the set of structs generated by a transformation system.

3.1 Generating process for natural numbers

We begin with the definition of the *inductive structure of natural numbers*. The set of primtypes $\Pi_{\mathbb{N}}$ consists of one element π_1 with one site being simultaneously the input and output site (Fig. 1). The set of semantic identities $\mathcal{I}_{\mathbb{N}}$ is the set of commutativity identities

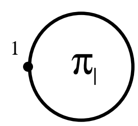

Figure 1: The primitive type π_1 for the inductive structure of natural numbers.

$\mathrm{Comm}(\Pi_{\mathbb{N}})$ [1, Def. 13]. In this case, this set consists of one identity

$$\pi_|\langle 1\rangle \lhd \pi_|\langle 2\rangle \equiv \pi_|\langle 2\rangle \lhd \pi_|\langle 1\rangle.$$

Next, the transformation set $T_{\mathbb{N}}$ consists of one transformation $\bar{\boldsymbol{\tau}}_|$ whose instance transformation is

$$\boldsymbol{\tau}_| = ([\pi_|\langle 1\rangle], [\pi_|\langle 1\rangle]).$$

The weighting mapping $l_{\mathbb{N}}$ is defined as $l_{\mathbb{N}}(\bar{\boldsymbol{\tau}}_|) = 1$. The progenitor $\bar{\boldsymbol{\kappa}}_{\mathbb{N}}$ is a struct whose instance struct is

$$\boldsymbol{\kappa}_{\mathbb{N}} = [\pi_|\langle 1\rangle].$$

Thus, the progenitor corresponds to number $1 \in \mathbb{N}$ and the transformation corresponds to the successor function in Peano axiomatics [4].

The diagram for the generating process $G_{\mathbb{N}}$ specified by the transformation system $\mathbf{TS}_{\mathbb{N}} = (T_{\mathbb{N}}, l_{\mathbb{N}}, \bar{\boldsymbol{\kappa}}_{\mathbb{N}})$ is shown in Fig. 2.

3.2 Generating process for binary sequences

Consider the set of finite binary sequences $\mathbb{B} = \{0, 1\}^*$. We will define an inductive structure and a generating process that generates $\mathbb{B}$.

The set of primtypes $\Pi_{\mathbb{B}}$ is shown in Fig. 3. The set of semantic identities $\mathcal{I}_{\mathbb{B}}$ is the set of commutativity identities $\mathrm{Comm}(\Pi_{\mathbb{B}})$. In case when the set of primitives is $\Pi_{\mathbb{B}}$, these

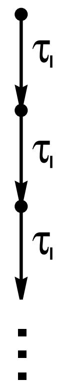

Figure 2: Diagram for generating process $G_{\mathbb{N}}$.

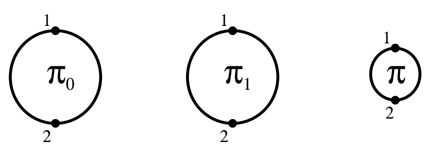

Figure 3: Primitive types for the inductive structure of binary sequences.

identities imply that every "connected" composite is semantically equivalent only to itself. The instance transformations for the transformation set $T_{\mathbb{B}}$ are shown in Fig. 4, and the weighting mapping is defined as $l_{\mathbb{B}}(\bar{\boldsymbol{\tau}}_0) = l_{\mathbb{B}}(\bar{\boldsymbol{\tau}}_1) = 1$. The progenitor $\bar{\boldsymbol{\kappa}}_{\mathbb{B}}$ is a struct whose instance struct is

$$\boldsymbol{\kappa}_{\mathbb{B}} = [\pi\langle 1, 2\rangle].$$

Thus, the progenitor corresponds to the empty binary sequence and transformations $\bar{\boldsymbol{\tau}}_0$, $\bar{\boldsymbol{\tau}}_1$ correspond to the attachments of "0" and "1", respectively, at the end of the sequence.

The diagram for the generating process $G_{\mathbb{B}}$ specified by the transformation system $\mathbf{TS}_{\mathbb{B}} = (T_{\mathbb{B}}, l_{\mathbb{B}}, \bar{\boldsymbol{\kappa}}_{\mathbb{B}})$ is shown in Fig. 5. Notice that for each struct from $TS_{\mathbb{B}}$, there exists a unique path from the progenitor to this struct. In other words, each binary sequence has a unique *constructive history*.

3.3 Generating process for strings

Consider the set $\mathbb{S}$ of strings over the alphabet $\{a, b\}$. Conventionally, this set is considered to be the same set as $\mathbb{B} = \{0, 1\}^*$ and, in fact, the same notation $\{a, b\}^*$ for $\mathbb{S}$ is being used. We will show that binary sequences and strings are, indeed, very different objects by pointing out to the difference in their constructive histories, which is not captured by the conventional notation.

The set of primtypes $\Pi_{\mathbb{S}}$ is shown in Fig. 6. The set of semantic identities $\mathcal{I}_{\mathbb{S}}$ is $\mathrm{Comm}(\Pi_{\mathbb{S}})$ plus the identities shown in Fig. 7. The shown identities have the following meaning: in-

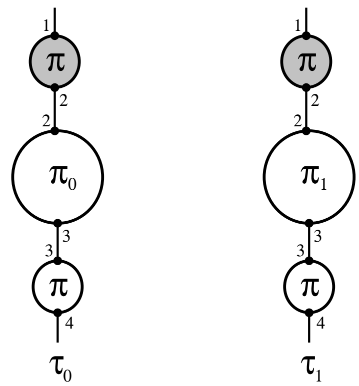

Figure 4: Instance transformations generating binary sequences.

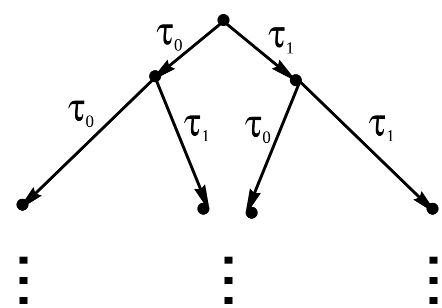

Figure 5: Diagram for generating process $G_{\mathbb{B}}$.

sertion of letter “a”, and then of letter “b” *after* “a” into a string has the same result as insertion of letter “b” first, and then of letter “a” *before* “b”.

The instance transformations for the transformation set $T_{\mathbb{S}}$ are shown in Fig. 8. The weighting mapping is defined as $l_{\mathbb{S}}(\bar{\boldsymbol{\tau}}_a) = l_{\mathbb{S}}(\bar{\boldsymbol{\tau}}_b) = 1$, and the progenitor is equal to the one for binary sequences

$$\boldsymbol{\kappa}_{\mathbb{S}} = [\pi\langle 1, 2\rangle].$$

Thus, the progenitor corresponds to the empty string, and transformations $\bar{\boldsymbol{\tau}}_a$, $\bar{\boldsymbol{\tau}}_b$ correspond to the insertions of “a” and “b” at any position in the string (since the semantic identites from $\mathcal{I}_{\mathbb{S}}$ ensure that a transformation can be applied at any output site of the struct representing

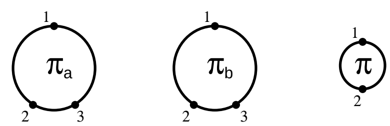

Figure 6: Primitive types for the inductive structure of strings.

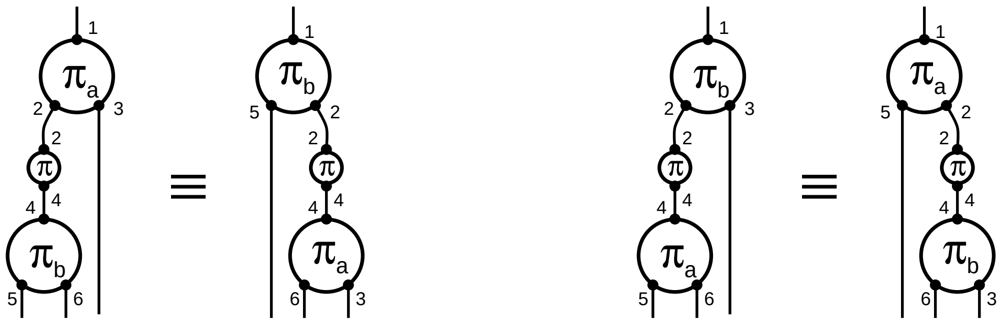

Figure 7: Semantic identities for the inductive structure of strings.

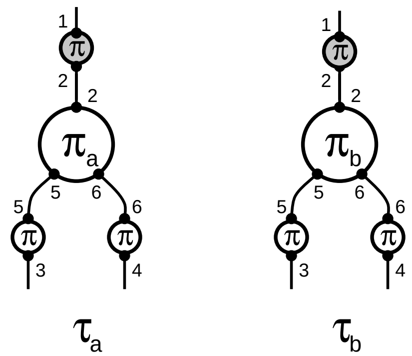

Figure 8: Instance transformations generating strings.

a string).

The diagram of the generating process $G_{\mathbb{S}}$ specified by the transformation system $\mathbf{TS}_{\mathbb{S}} = (T_{\mathbb{S}}, l_{\mathbb{S}}, \bar{\boldsymbol{\kappa}}_{\mathbb{S}})$ is shown in Fig. 9. Notice that, as a consequence of the introduced semantic identities, for each struct from $TS_{\mathbb{S}}$ representing a string of length n, there exist $n!$ construction paths from the progenitor to this struct. Since these paths lead to the same string, they are indistinguishable from the point of view of applicability of further transformations.

Comparing the diagrams in Fig. 5 and Fig. 9, one can see that the constructive histories of binary sequences and strings differ significantly. In case of strings, one may think of other possibilities for the generating process, for example, instead of insertions at any place in the string, only additions of letters at the beginning and at the end of the string may be chosen as transformations. This indicates that the conventional understanding of a string as an object is quite imprecise, leading to arbitrariness in defining classes of strings.

4 Continuous parameter Markov chains

In this section, we present some basic facts from the theory of c.p. Markov chains, which are necessary to define rigorously the generating process and some random variables related

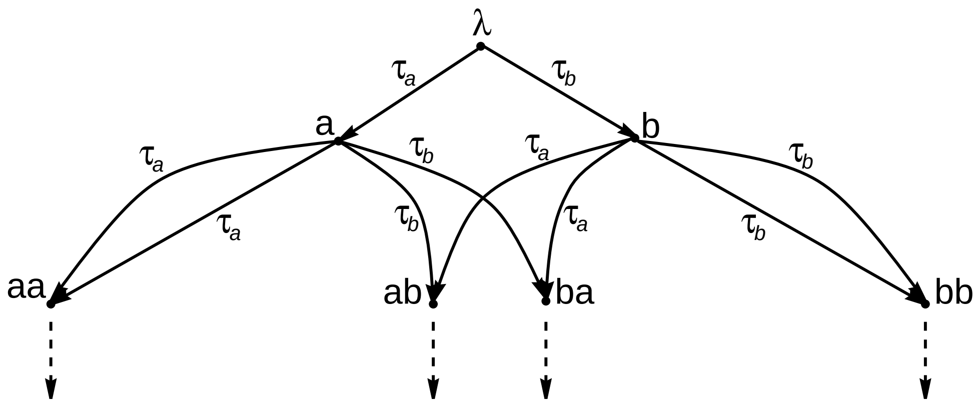

Figure 9: Diagram for the generating process $G_{\mathbb{S}}$.

to it.

Note that the states of the generating process are structs, and not real numbers as usually assumed in the theory to stochastic processes. Many of the results, including those needed to define the typicality measure, will still hold, since the set of structs is countable. In fact, one can always embed the set of structs into the set of real numbers, and many of the results will be independent of the embedding. However, some of them, particularly those related to the questions about compactification of the state space and limit points of the process, do depend on the embedding. These issues will be briefly discussed in Conclusion. Here we will give a general definition of the c.p. Markov chain with an arbitrary countable state space and present all necessary concepts which are independent of the particular embedding.

Let X be a countable set. Let $(\Omega, \mathcal{F}, \mathcal{P})$ be a probability triple, i.e., Ω is a set called the probability space, $\mathcal{F}$ is a Borel field of subsets in Ω, and $\mathcal{P}$ is a probability measure on Ω.

An *X-valued random variable* is a measurable mapping $\xi : \Omega \to X$ defined for almost all $\omega \in \Omega$, i.e., for all $i \in X$, set $\{\omega \,|\, \xi(\omega) = i\}$ belongs to field $\mathcal{F}$.

A *continuous parameter (c.p.) Markov chain with state space* X [3, §II.4] is defined as a family of X-valued random variables $\{x_t(\omega)\}$, $t \in [0, \infty)$, satisfying the Markov property: $\forall\, t_1 < t_2 < \ldots < t_n \in [0, \infty),\ i_1, \ldots, i_n \in X$,

$$\mathcal{P}\left\{x_{t_n}(\omega) = i_n \,|\, x_{t_1}(\omega) = i_1, \ldots, x_{t_{n-1}}(\omega) = i_{n-1}\right\} = \mathcal{P}\left\{x_{t_n}(\omega) = i_n \,|\, x_{t_{n-1}}(\omega) = i_{n-1}\right\}. \quad (1)$$

We will consider c.p. Markov chains with *stationary transition probabilities* [3, Section II.4], i.e., conditional probabilities

$$\mathcal{P}\{x_{s+t}(\omega) = j \,|\, x_s(\omega) = i\}$$

are independent of s.

For a c.p. Markov chain $\{x_t(\omega)\}$, the matrix $P(t) = (p_{ij}(t))$, $i, j \in X$ is defined by

$$p_{ij}(t) \stackrel{\text{def}}{=} \mathcal{P}\{x_{s+t}(\omega) = j \,|\, x_s(\omega) = i\}, \quad t > 0.$$

The *initial distribution* is defined to be the set $\{p_i,\ i \in X\}$, where

$$p_i \stackrel{\text{def}}{=} \mathcal{P}\{x_0(\omega) = i\}, \qquad \sum_i p_i = 1.$$

For any Markov chain with stationary transitional probabilities, the matrix $P(t)$ satisfies the following three properties [3, §§II.1,II.4]: for all $i, j \in X$ and $s, t > 0$,

(A) $p_{ij}(t) \geq 0$

(B) $\sum_j p_{ij}(t) = 1$

(C) $\sum_k p_{ik}(s) p_{kj}(t) = p_{ij}(s+t)$.

Matrices satisfying (A–C) are called *transition matrices*. The converse is true: for any transition matrix $P(t)$ and any initial distribution $\{p_i,\ i \in X\}$, there exists a corresponding Markov chain $\{x_t(\omega),\ t \geq 0\}$ with state space X [3, p. 137].

If $P(t)$ is *standard*, i.e.,

$$\lim_{t \to 0} p_{ij}(t) = \delta_{ij},$$

then, according to [3, Theorem II.2.5], there exist derivatives $p'_{ij}(0)$ (which may be infinite). The matrix $Q = (q_{ij}) = P'(0)$ will be called the *transition rate matrix* corresponding to the standard transition matrix $P(t)$.

We will study the converse question, i.e. under what conditions a given finite or infinite matrix Q is a transition rate matrix corresponding to some transition matrix $P(t)$. We restrict ourselves to the case when matrix Q satisfies the following conditions:

$$\begin{array}{ll} q_{ij} = 0, & i \not\leq j \\ q_{ij} \geq 0, & i < j \\ \sum_j q_{ij} = 0, & \forall j \\ q_{ii} > -\infty, & \forall i \end{array} \tag{2}$$

Matrices satisfying (2) will be called *transformation matrices*. It follows from [3, Theorem II.18.1] that for any transformation matrix Q, the following system of *Kolmogorov equations*:

$$P'(t) = QP(t), \qquad P'(t) = P(t)Q \tag{3}$$

has a solution $\bar{P}(t) = (\bar{p}_{ij}(t))$ called the *minimal solution corresponding to the transformation matrix* Q, which is constructed as follows:

$$\begin{aligned} p_{ij}^{\langle 0 \rangle}(t) &= \delta_{ij} e^{q_{ii} t} = \delta_{ij} e^{q_{jj} t} \\ p_{ij}^{\langle n+1 \rangle}(t) &= \sum_{k<j} \int_0^t p_{ik}^{\langle n \rangle}(s) q_{kj} e^{q_{jj}(t-s)} ds \\ \bar{p}_{ij}(t) &= \sum_{n=0}^{\infty} p_{ij}^{\langle n \rangle}(t). \end{aligned} \tag{4}$$

Matrix $\bar{P}(t)$ is standard and satisfies (A) and (C) but, instead of (B), only the following condition is guaranteed to hold:

$$\sum_j \bar{p}_{ij}(t) \leq 1. \tag{B'}$$

If (B) holds, matrix $\bar{P}(t)$ is the unique solution of system (3) [3, Theorem II.17.2], otherwise there are infinitely many of them [3, Corollary to Theorem II.19.4] and thus infinitely many essentially different c.p. Markov chains corresponding to matrix Q.

Matrices satisfying (A,B',C) are called *substochastic transition matrices*. Given an initial distribution $\{p_i,\, i \in X\}$, a substochastic transition matrix specifies a *c.p. Markov chain stopped at random time* $\tau(\omega)$, $\{x_t(\omega), t < \tau(\omega)\}$. Random variable $\tau(\omega)$ is called the *stopping random variable*, and its distribution is given by

$$\mathcal{P}\{\tau(\omega) \leq t \,|\, x_0(\omega) = i\} = 1 - \sum_{j \in X} \bar{p}_{ij}(t).$$

If the stopping random variable is infinite with probability one, i.e.

$$\mathcal{P}\{\tau(\omega) = \infty\} = 1,$$

then (B) holds [3, §II.19]. Otherwise, $\tau(\omega)$ is finite with a non-zero probability. To describe this situation, we will study the behaviour of the process further. First, the c.p. Markov chain corresponding to $\bar{p}_{ij}$ can be chosen so that for all $t < \tau(\omega)$,

$$\lim_{s \downarrow t} x_s(\omega) = x_t(\omega)$$

(in [3, §II.7] such $x_t(\tau)$ is called the x_+-version of the process). Under the assumptions on matrix Q made above, for almost all ω, we have that function $y(t) = x_t(\omega)$ is a piecewise constant function on $[0, \tau(\omega))$ with finitely or infinitely many points of discontinuity. If the number of these points is finite, then $\tau(\omega) = \infty$ and the last point i is called an *absorbing state* of the process, necessarily having $q_{ii} = 0$. Otherwise, let $\{\tau_n(\omega),\, n \geq 1\}$ be the increasing sequence of these points. Then,

$$\begin{aligned} &x_t(\omega) = x_{\tau_n(\omega)}, \qquad \tau_n(\omega) \leq t < \tau_{n+1}(\omega), \qquad \text{and} \\ &\lim_{n \to \infty} \tau_n(\omega) = \tau(\omega). \end{aligned}$$

Define the *jump chain* associated with $x_t(\omega)$ as a discrete parameter Markov chain $\chi_n(\omega) = x(\tau_n(\omega), \omega)$ [3, p. 236]. Then, according to [3, Theorem II.19.1],

$$\tau(\omega) = \infty \iff -\sum_n q^{-1}_{\chi_n(\omega)\chi_n(\omega)} = \infty. \tag{5}$$

Thus, the process reaches infinity in a finite time with non-zero probability if and only if

$$\mathcal{P}\{-\sum_n q^{-1}_{\chi_n(\omega)\chi_n(\omega)} < \infty\} > 0. \tag{6}$$

5 Formal definition of the generating process

Let $\mathbf{TS} = (T, l, \bar{\boldsymbol{\kappa}})$ be a transformation system, and let TS be the set of structs generated by it. There exists a partial ordering relation $\leq$ on TS: we have $\bar{\boldsymbol{\alpha}} \leq \bar{\boldsymbol{\beta}}$ if and only if there exists a path $p \in P_T$ from $\bar{\boldsymbol{\alpha}}$ to $\bar{\boldsymbol{\beta}}$. Let the set of states be $X = TS$. Define matrix $Q = (q_{\bar{\boldsymbol{\alpha}}\bar{\boldsymbol{\beta}}})$, $\bar{\boldsymbol{\alpha}}, \bar{\boldsymbol{\beta}} \in X$ as follows:

$$\begin{aligned} q_{\bar{\boldsymbol{\alpha}}\bar{\boldsymbol{\alpha}}} &= -\sum_{[\boldsymbol{\alpha};\boldsymbol{\tau}]\in EP_T} \frac{1}{l(\bar{\boldsymbol{\tau}})} \\ q_{\bar{\boldsymbol{\alpha}}\bar{\boldsymbol{\beta}}} &= \sum_{[\boldsymbol{\alpha};\boldsymbol{\tau}]\in EP_T, [\boldsymbol{\alpha}\lhd\boldsymbol{\tau}]=\bar{\boldsymbol{\beta}}} \frac{1}{l(\bar{\boldsymbol{\tau}})}, \quad \bar{\boldsymbol{\beta}} \neq \bar{\boldsymbol{\alpha}}. \end{aligned} \tag{7}$$

Conditions (2) hold, hence Q is a transformation matrix.[2] Therefore, there exists the minimal solution $\bar{P}(t)$ to the system of Kolmogorov backward equations (3), which is a standard substochastic transition matrix. Matrix $\bar{P}(t)$ and initial distribution such that $p_{\bar{\boldsymbol{\kappa}}} = 1$ specify a stopped c.p. Markov chain $\{x_t(\omega), t < \tau(\omega)\}$, which is called the **generating process for transformation system TS**.

In what follows, we assume that the elements of matrices Q and $P(t)$ that are not specified are equal to zero.

For the transformation system $\mathbf{TS}_{\mathbb{N}}$ of natural numbers from Section 3.1, the corresponding transformation matrix $Q_{\mathbb{N}}$ is defined by

$$\begin{aligned} q_{\bar{\boldsymbol{\alpha}}\bar{\boldsymbol{\alpha}}} &= -1 \\ q_{\bar{\boldsymbol{\alpha}},[\boldsymbol{\alpha}\lhd\boldsymbol{\tau}_|]} &= 1. \end{aligned}$$

Denote $\boldsymbol{\alpha} \lhd n\boldsymbol{\tau} \stackrel{\text{def}}{=} \boldsymbol{\alpha} \lhd \boldsymbol{\tau} \lhd \ldots \lhd \boldsymbol{\tau}$, where $\boldsymbol{\tau}$ is attached to $\boldsymbol{\alpha}$ n times, $n \geq 0$. The minimal solution $\bar{P}(t)$ can be computed using formulas (4):

$$\begin{aligned} \bar{p}_{\bar{\boldsymbol{\alpha}}\bar{\boldsymbol{\alpha}}}(t) &= e^{-t} \\ \bar{p}_{\bar{\boldsymbol{\alpha}},[\boldsymbol{\alpha}\lhd\boldsymbol{\tau}]}(t) &= \int_0^t e^{-s}e^{-(t-s)}ds = te^{-t} \\ \bar{p}_{\bar{\boldsymbol{\alpha}},[\boldsymbol{\alpha}\lhd 2\boldsymbol{\tau}]}(t) &= \int_0^t se^{-s}e^{-(t-s)}ds = \tfrac{t^2}{2!}e^{-t} \\ &\ldots \\ \bar{p}_{\bar{\boldsymbol{\alpha}},[\boldsymbol{\alpha}\lhd n\boldsymbol{\tau}]}(t) &= \tfrac{t^n}{n!}e^{-t}. \end{aligned}$$

Thus, $x_t^{\mathbb{N}}(\omega)$ is the Poisson process. Since for every $\bar{\boldsymbol{\alpha}}$ we have

$$\sum_{\boldsymbol{\beta}} \bar{p}_{\bar{\boldsymbol{\alpha}}\bar{\boldsymbol{\beta}}}(t) = \sum_{n=0}^{\infty} \frac{t^n}{n!}e^{-t} = 1,$$

matrix $\bar{P}(t)$ is a standard transition matrix, thus the stopping time of the process is infinite with probability 1.

For the transformation system $\mathbf{TS}_{\mathbb{B}}$ of binary sequences from Section 3.2, the corresponding transformation matrix $Q_{\mathbb{B}}$ is defined by

$$\begin{aligned} q_{\bar{\boldsymbol{\alpha}}\bar{\boldsymbol{\alpha}}} &= -2 \\ q_{\bar{\boldsymbol{\alpha}},[\boldsymbol{\alpha}\lhd\boldsymbol{\tau}_0]} &= 1 \\ q_{\bar{\boldsymbol{\alpha}},[\boldsymbol{\alpha}\lhd\boldsymbol{\tau}_1]} &= 1. \end{aligned}$$

[2] Matrix Q is not a matrix in a conventional sense, since its rows and columns are not indexed by natural numbers. Still, for a matrix $A = (a_{\bar{\boldsymbol{\alpha}}\bar{\boldsymbol{\beta}}})$ with non-negative components, the products AQ and QA are correctly defined in a usual way. The corresponding infinite summations are independent of the ordering of summands, since each row and column of Q has at most one negative element.

The minimal solution $\bar{P}(t)$ is

$$\begin{array}{rcl}\bar{p}_{\bar{\boldsymbol{\alpha}}\bar{\boldsymbol{\alpha}}}(t) & = & e^{-2t} \\ \bar{p}_{\bar{\boldsymbol{\alpha}},[\boldsymbol{\alpha}\lhd\boldsymbol{\tau}_{i_1}]}(t) & = & te^{-2t} \\ \bar{p}_{\bar{\boldsymbol{\alpha}},[\boldsymbol{\alpha}\lhd\boldsymbol{\tau}_{i_1}\lhd\boldsymbol{\tau}_{i_2}]}(t) & = & \frac{t^2}{2!}\cdot e^{-2t} \\ & \cdots & \\ \bar{p}_{\bar{\boldsymbol{\alpha}},[\boldsymbol{\alpha}\lhd\boldsymbol{\tau}_{i_1}\lhd\ldots\lhd\boldsymbol{\tau}_{i_n}]}(t) & = & \frac{t^n}{n!}\cdot e^{-2t}, \qquad n>0,\end{array}$$

where $i_1, i_2, \ldots, i_n \in \{0, 1\}$. Similarly to the case of natural numbers, one can show that this is a standard transition matrix and hence the stopping time of the process is infinite with probability 1.

For the transformation system $\mathbf{TS}_{\mathbb{S}}$ of strings from Section 3.3, the corresponding transformation matrix $Q_{\mathbb{S}}$ is defined by

$$\begin{array}{rcl}q_{\bar{\boldsymbol{\alpha}}\bar{\boldsymbol{\alpha}}} & = & -\big|\{[\boldsymbol{\alpha};\boldsymbol{\tau}]\in EP_{T_{\mathbb{S}}}\}\big| = -(\mathrm{len}(\bar{\boldsymbol{\alpha}})+1) \\ q_{\bar{\boldsymbol{\alpha}},[\boldsymbol{\alpha}\lhd\boldsymbol{\tau}_i\langle f\rangle]} & = & \big|\{[\boldsymbol{\alpha};\boldsymbol{\tau}]\in EP_{T_{\mathbb{S}}}\,|\,[\boldsymbol{\alpha}\lhd\boldsymbol{\tau}]=[\boldsymbol{\alpha}\lhd\boldsymbol{\tau}_i\langle f\rangle]\}\big|,\end{array}$$

where $i \in \{a, b\}$, f is any site replacement such that $\boldsymbol{\alpha} \lhd \boldsymbol{\tau}_i\langle f\rangle$ exists, and $\mathrm{len}(\bar{\boldsymbol{\alpha}})$ denotes the length of the string corresponding to $\bar{\boldsymbol{\alpha}}$. For almost all ω, the jump chain $\chi_n(\omega)$, $n \geq 0$ corresponds to the sequence of strings whose lengths are $0, 1, \ldots$. Thus,

$$-\sum_n q^{-1}_{\chi_n(\omega)\chi_n(\omega)} = \sum_{n=0}^{\infty}(n+1)^{-1} = \infty,$$

and the stopping time of the process is infinite with probability one, according to criterion (6).

In the next section, we will consider a transformation system whose generating process has an finite running time with probability 1.

6 Graph generating process

Consider the set of unlabeled directed multigraphs $\mathbb{G}$. We define a corresponding inductive structure and a generating process that generates a subset of $\mathbb{G}$.

The set of primtypes $\Pi_{\mathbb{G}}$ is shown in Fig. 10. The set of semantic identities $\mathcal{I}_{\mathbb{G}}$ is the

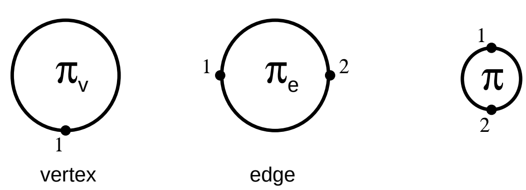

Figure 10: Primitive types for the inductive structure of graphs.

set of commutativity identities $\mathrm{Comm}(\Pi_{\mathbb{B}})$. Out of many possibilities for the choice of

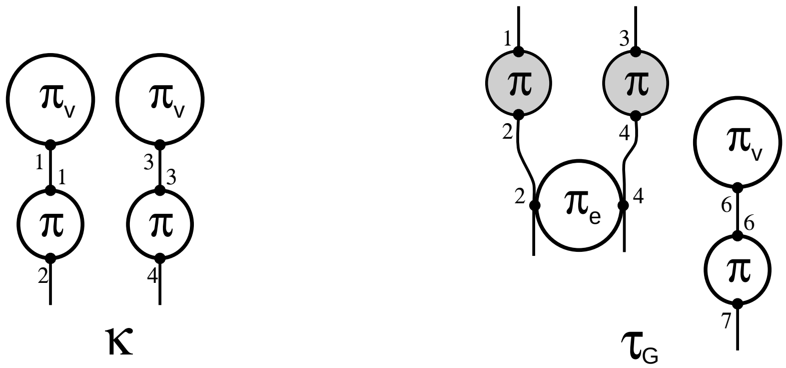

Figure 11: Progenitor and transformation of the graph transformation system.

transformation system to generate a subset of graphs, we select a particular one, $T_{\mathbb{G}} = \{\bar{\boldsymbol{\tau}}_{\mathbb{G}}\}$, shown in Fig. 11. The weighting mapping is defined as $l_{\mathbb{G}}(\bar{\boldsymbol{\tau}}_{\mathbb{G}}) = 1$. The progenitor $\bar{\boldsymbol{\kappa}}_{\mathbb{G}}$ is a struct whose instance struct is

$$\boldsymbol{\kappa}_{\mathbb{G}} = [\pi_v\langle 1\rangle \lhd \pi\langle 1, 2\rangle \lhd \pi_v\langle 3\rangle \lhd \pi\langle 3, 4\rangle].$$

Thus, the progenitor corresponds to the graph with two vertices and no edges, and transformation $\bar{\boldsymbol{\tau}}_{\mathbb{G}}$ corresponds to creation of an edge and a vertex simultaneously. Note that this transformation system generates only graphs whose number of edges is equal to the number of vertices minus 2. Of course, the set of all graphs could have been generated as well by another transformation system; this one is chosen as an example of transformation system whose generating process has finite running time with probability one.

If $\bar{\boldsymbol{\gamma}}$ is a struct corresponding to a graph with n vertices, then

$$q_{\bar{\boldsymbol{\gamma}}\bar{\boldsymbol{\gamma}}} = -|\{[\boldsymbol{\gamma}; \boldsymbol{\tau}] \in EP_{\mathbb{G}}\}| = -n(n-1).$$

For almost all ω, the jump chain $\chi_n(\omega)$, $n \geq 0$ corresponds to the sequence of graphs with $2, 3, \ldots$ vertices. Thus,

$$-\sum_n q^{-1}_{\chi_n(\omega)\chi_n(\omega)} = \sum_{n=0}^{\infty} \frac{1}{(n+2)(n+1)} < \infty,$$

and the stopping time of the process is finite with probability 1, according to criterion (6). In other words, the process constructs an "infinite graph" (the notion clarified later in Conclusion) in a finite time.

7 The time that the process spends in a state and the running time

Let **TS** be a transformation system, Q be the corresponding transformation matrix, $\bar{P}(t)$ be its minimal solution, and $x_t(\omega)$ be the corresponding generating process. Assume also that for each $\bar{\boldsymbol{\gamma}} \in TS$, the number of ancestors of $\bar{\boldsymbol{\gamma}}$ is finite:[3]

$$\mathrm{anc}(\bar{\boldsymbol{\gamma}}) \stackrel{\text{def}}{=} |\{\bar{\boldsymbol{\alpha}} \mid \bar{\boldsymbol{\alpha}} \leq \bar{\boldsymbol{\gamma}}\}| < \infty. \tag{8}$$

Define

$$c_{\bar{\boldsymbol{\alpha}}}(\omega, t) \stackrel{\text{def}}{=} \begin{cases} 1, & x_t(\omega) = \bar{\boldsymbol{\alpha}} \\ 0, & \text{otherwise.} \end{cases}$$

Then,

$$m_{\bar{\boldsymbol{\alpha}}}(\omega) \stackrel{\text{def}}{=} \int_0^\infty c_{\bar{\boldsymbol{\alpha}}}(\omega, t) dt$$

is the random variable corresponding to the *total time the process spends in* $\bar{\boldsymbol{\alpha}}$. Its expectation is

$$M_{\bar{\boldsymbol{\alpha}}} \stackrel{\text{def}}{=} \int_0^\infty E\{c_{\bar{\boldsymbol{\alpha}}}(t, \omega)\} dt = \int_0^\infty \bar{p}_{\bar{\boldsymbol{\kappa}}\bar{\boldsymbol{\alpha}}}(t) dt.$$

Let

$$M_{\bar{\boldsymbol{\alpha}}\bar{\boldsymbol{\beta}}} \stackrel{\text{def}}{=} \int_0^\infty \bar{p}_{\bar{\boldsymbol{\alpha}}\bar{\boldsymbol{\beta}}}(t) dt.$$

For the matrix $N = MQ = (n_{\bar{\boldsymbol{\alpha}}\bar{\boldsymbol{\beta}}})$, we have

$$n_{\bar{\boldsymbol{\alpha}}\bar{\boldsymbol{\beta}}} = \sum_{\bar{\boldsymbol{\gamma}}} \int_0^\infty \bar{p}_{\bar{\boldsymbol{\alpha}}\bar{\boldsymbol{\gamma}}}(t) dt \cdot q_{\bar{\boldsymbol{\gamma}}\bar{\boldsymbol{\beta}}}. \tag{9}$$

Since Q is a transformation matrix, it is upper triangular (see (2)), and it follows from the assumption (8) that the summation in (9) is finite. Thus, summation and integration can be exchanged. Since $\bar{P}(t)$ satisfies the second system in (3), we have:

$$n_{\bar{\boldsymbol{\alpha}}\bar{\boldsymbol{\beta}}} = \int_0^\infty \sum_{\bar{\boldsymbol{\gamma}}} \bar{p}_{\bar{\boldsymbol{\alpha}}\bar{\boldsymbol{\gamma}}}(t) q_{\bar{\boldsymbol{\gamma}}\bar{\boldsymbol{\beta}}} dt = \int_0^\infty \bar{p}'_{\bar{\boldsymbol{\alpha}}\bar{\boldsymbol{\beta}}}(t) dt = \lim_{t \to \infty} \bar{p}_{\bar{\boldsymbol{\alpha}}\bar{\boldsymbol{\beta}}}(t) - \bar{p}_{\bar{\boldsymbol{\alpha}}\bar{\boldsymbol{\beta}}}(0).$$

By definition, we have that $\bar{p}_{\bar{\boldsymbol{\alpha}}\bar{\boldsymbol{\beta}}}(0) = \delta_{\bar{\boldsymbol{\alpha}}\bar{\boldsymbol{\beta}}}$. Next, since (8) holds, only a finite number of $\bar{p}^{\langle n \rangle}_{\bar{\boldsymbol{\alpha}}\bar{\boldsymbol{\beta}}}(t)$ are non-zero functions. Therefore, since for each n,

$$\lim_{t \to \infty} \bar{p}^{\langle n \rangle}_{\bar{\boldsymbol{\alpha}}\bar{\boldsymbol{\beta}}}(t) = 0$$

(see [3, p. 232]), we have that

$$\lim_{t \to \infty} \bar{p}_{\bar{\boldsymbol{\alpha}}\bar{\boldsymbol{\beta}}}(t) = 0$$

and that $n_{\bar{\boldsymbol{\alpha}}\bar{\boldsymbol{\beta}}} = -\delta_{\bar{\boldsymbol{\alpha}}\bar{\boldsymbol{\beta}}}$. Thus we have $MQ = -I$, where I is the identity matrix. Since Q is upper triangular and (8) holds, this system has a unique solution. The elements of the first row of M, $M_{\bar{\boldsymbol{\kappa}}\bar{\boldsymbol{\alpha}}} = M_{\bar{\boldsymbol{\alpha}}}$, can be computed using the following algorithm:

$$\begin{aligned} M_{\bar{\boldsymbol{\kappa}}} &= -1/q_{\bar{\boldsymbol{\kappa}}\bar{\boldsymbol{\kappa}}} \\ M_{\bar{\boldsymbol{\alpha}}} &= -(\textstyle\sum_{\bar{\boldsymbol{\gamma}} < \bar{\boldsymbol{\alpha}}} M_{\bar{\boldsymbol{\gamma}}} q_{\bar{\boldsymbol{\gamma}}\bar{\boldsymbol{\alpha}}})/q_{\bar{\boldsymbol{\alpha}}\bar{\boldsymbol{\alpha}}}, \end{aligned} \tag{10}$$

[3] There exist transformation systems violating this assumption.

whose time complexity is $O(\mathrm{anc}(\bar{\boldsymbol{\alpha}})^2)$. Formulas (10) still hold if $q_{\bar{\boldsymbol{\alpha}}\bar{\boldsymbol{\alpha}}} = 0$, in which case $M_{\bar{\boldsymbol{\alpha}}} = \infty$.

Since $\sum_{\bar{\boldsymbol{\alpha}}} c_{\bar{\boldsymbol{\alpha}}}(t,\omega) = 1$ if and only if $t < \tau(\omega)$, we have for the *expected running time* of the generating process that

$$E\{\tau(\omega)\} = \int_0^\infty E\{\sum_{\bar{\boldsymbol{\alpha}}} c_{\bar{\boldsymbol{\alpha}}}(t,\omega)\}dt = \sum_{\bar{\boldsymbol{\alpha}}} M_{\bar{\boldsymbol{\alpha}}}.$$

In particular, $E\{\tau(\omega)\}$ is finite if and only if $\sum_{\bar{\boldsymbol{\alpha}}} M_{\bar{\boldsymbol{\alpha}}}$ is finite.

8 Typicality measure

In this section, we define the notion of typicality measure. For a struct $\bar{\boldsymbol{\gamma}}$, its typicality is informally defined as the probability that this struct is randomly encountered by an observer. We assume that the expected waiting time for the observer to come is the same at each moment, from what it follows that the observation moment is an exponentially distributed random variable, call it ξ_u:

$$\mathcal{P}\{\xi_u(\omega) < t\} = 1 - e^{-ut}, \qquad t \geq 0,$$

where u is the parameter of the exponential distribution. We also assume that this random variable is independent of the process $\{x_t(\omega)\}$, i.e., independent of each random variable $x_t(\omega)$ for all $t \geq 0$.

The probability that process $x_t(\omega)$ will be in struct $\bar{\boldsymbol{\gamma}}$ at random moment $t = \xi_u(\omega)$ is

$$\mathcal{P}\{x_{\xi_u(\omega)}(\omega) = \bar{\boldsymbol{\alpha}}\} = \int_0^\infty \bar{p}_{\bar{\boldsymbol{\kappa}}\bar{\boldsymbol{\alpha}}}(t) \cdot ue^{-ut}dt.$$

If the running time of the process is infinite with probability 1, then (B) holds, and we have

$$\sum_{\bar{\boldsymbol{\alpha}}} \mathcal{P}\{x_{\xi_u(\omega)}(\omega) = \bar{\boldsymbol{\alpha}}\} = \int_0^\infty \sum_{\bar{\boldsymbol{\alpha}}} \bar{p}_{\bar{\boldsymbol{\kappa}}\bar{\boldsymbol{\alpha}}}(t) \cdot ue^{-ut}dt = \int_0^\infty ue^{-ut}dt = 1,$$

therefore

$$\mathbf{g}_u(\bar{\boldsymbol{\alpha}}) \stackrel{\mathrm{def}}{=} \mathcal{P}\{x_{\xi_u(\omega)}(\omega) = \bar{\boldsymbol{\alpha}}\}$$

is a probability measure on the set of states of the process, called the **typicality measure**. If the running time $\tau(\omega)$ is finite with positive probability, the typicality measure is not a probablity measure, since the typicality of the set of all states is less than one.

To compute the typicality measure for a given generating process induced by a transformation system **TS**, we introduce the notion of an *observed transformation system* $\tilde{\mathbf{TS}}$ corresponding to **TS**. Then, the typicality measure for a state of the original process is expressed via the expected time the terminated process spends in the corresponding state, which has been computed in the previous section.

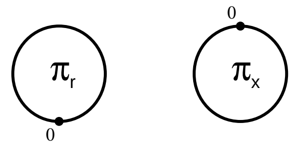

Figure 12: Additional primitives for the observed inductive structure.

Let $(\Pi, \mathcal{I})$ be an inductive structure. The **observed inductive structure** $(\tilde{\Pi}, \tilde{\mathcal{I}})$ corresponding to $(\Pi, \mathcal{I})$ is defined as follows (assume that $\pi_r, \pi_x \notin \Pi$, see Fig. 12): add a new site 0 to the set of sites and let

$$\begin{aligned} \tilde{\Pi} &\stackrel{\text{def}}{=} \Pi \cup \{\pi_r, \pi_x\} \\ \tilde{\mathcal{I}} &\stackrel{\text{def}}{=} \mathcal{I} \cup (\cup_{\pi\in\Pi} \operatorname{Comm}(\{\pi, \pi_r, \pi_x\}))\,. \end{aligned}$$

In other words, we add two new primitives π_r and π_x to Π and make them commutative with all other primitives. For every instance struct $\boldsymbol{\alpha} \in \Theta$, let

$$\boldsymbol{\alpha}_r \stackrel{\text{def}}{=} \boldsymbol{\alpha} \lhd [\pi_r] \in \tilde{\Theta}.$$

For a transformation system $\mathbf{TS} = (T, l, \bar{\boldsymbol{\kappa}})$ in $(\Pi, \mathcal{I})$ and $u \in \mathbb{R}_+$, define the corresponding **observed transformation system** $\tilde{\mathbf{TS}}(u) = (\tilde{T}, \tilde{l}, \bar{\boldsymbol{\kappa}}_r)$ as follows:

$$\begin{aligned} \tilde{T} &\stackrel{\text{def}}{=} T \cup \{\bar{\boldsymbol{\tau}}_x\} \\ \boldsymbol{\tau}_x &\stackrel{\text{def}}{=} ([\pi_r], [\pi_x]) \\ \tilde{l}(\bar{\boldsymbol{\tau}}) &\stackrel{\text{def}}{=} l(\bar{\boldsymbol{\tau}}) \\ \tilde{l}(\bar{\boldsymbol{\tau}}_x) &\stackrel{\text{def}}{=} u^{-1}. \end{aligned}$$

The new set of states is

$$\tilde{TS} = \{\bar{\boldsymbol{\alpha}}_r \,|\, \bar{\boldsymbol{\alpha}} \in TS\} \cup \{[\boldsymbol{\alpha}_r \lhd \boldsymbol{\tau}_x] \,|\, \bar{\boldsymbol{\alpha}} \in TS\},$$

and the new observed transformation matrix (dependent on u) $\tilde{Q}(u) = (\tilde{q}_{\bar{\boldsymbol{\alpha}}_x\bar{\boldsymbol{\beta}}_x})_{\bar{\boldsymbol{\alpha}}_x,\bar{\boldsymbol{\beta}}_x\in\tilde{TS}}$ is defined by

$$\begin{aligned} \tilde{q}_{\bar{\boldsymbol{\alpha}}_r\bar{\boldsymbol{\alpha}}_r} &= q_{\bar{\boldsymbol{\alpha}}\bar{\boldsymbol{\alpha}}} - u, && \bar{\boldsymbol{\alpha}} \in TS \\ \tilde{q}_{\bar{\boldsymbol{\alpha}}_r\bar{\boldsymbol{\beta}}_r} &= q_{\bar{\boldsymbol{\alpha}}\bar{\boldsymbol{\beta}}}, && \bar{\boldsymbol{\alpha}}, \bar{\boldsymbol{\beta}} \in TS,\ \bar{\boldsymbol{\alpha}} \neq \bar{\boldsymbol{\beta}} \\ \tilde{q}_{\bar{\boldsymbol{\alpha}}_r[\boldsymbol{\alpha}_r\lhd\boldsymbol{\tau}_x]} &= u, && \bar{\boldsymbol{\alpha}} \in TS \\ \tilde{q}_{[\boldsymbol{\alpha}_r\lhd\boldsymbol{\tau}_x]\bar{\boldsymbol{\beta}}_x} &= q_{\bar{\boldsymbol{\alpha}}\bar{\boldsymbol{\beta}}}, && \bar{\boldsymbol{\alpha}} \in TS,\ \bar{\boldsymbol{\beta}}_x \in \tilde{TS}. \end{aligned}$$

We prove by induction that for all $n \geq 0$, $\bar{\boldsymbol{\alpha}}, \bar{\boldsymbol{\beta}} \in TS$ and $t \geq 0$,

$$\tilde{\bar{p}}^{\langle n\rangle}_{\bar{\boldsymbol{\alpha}}_r\bar{\boldsymbol{\beta}}_r}(t) = \bar{p}^{\langle n\rangle}_{\bar{\boldsymbol{\alpha}}\bar{\boldsymbol{\beta}}}(t)e^{-ut}\ :$$

$$\begin{aligned} \tilde{\bar{p}}^{\langle 0\rangle}_{\bar{\boldsymbol{\alpha}}_r\bar{\boldsymbol{\beta}}_r}(t) &= \delta_{\bar{\boldsymbol{\alpha}}_r\bar{\boldsymbol{\beta}}_r} e^{q_{\bar{\boldsymbol{\alpha}}_r\bar{\boldsymbol{\alpha}}_r}t} = \delta_{\bar{\boldsymbol{\alpha}}\bar{\boldsymbol{\beta}}} e^{(q_{\bar{\boldsymbol{\alpha}}\bar{\boldsymbol{\alpha}}}-u)t} = \bar{p}^{\langle 0\rangle}_{\bar{\boldsymbol{\alpha}}\bar{\boldsymbol{\beta}}}(t)e^{-ut} \\ \tilde{\bar{p}}^{\langle n+1\rangle}_{\bar{\boldsymbol{\alpha}}_r\bar{\boldsymbol{\beta}}_r}(t) &= \textstyle\sum_{\bar{\boldsymbol{\gamma}}_r<\bar{\boldsymbol{\beta}}_r} \int_0^t \tilde{\bar{p}}^{\langle n\rangle}_{\bar{\boldsymbol{\alpha}}_r\bar{\boldsymbol{\gamma}}_r}(s)\tilde{q}_{\bar{\boldsymbol{\gamma}}_r\bar{\boldsymbol{\beta}}_r} e^{\tilde{q}_{\bar{\boldsymbol{\beta}}_r\bar{\boldsymbol{\beta}}_r}(t-s)}ds = \\ &\quad \textstyle\sum_{\bar{\boldsymbol{\gamma}}<\bar{\boldsymbol{\beta}}} \int_0^t \bar{p}^{\langle n\rangle}_{\bar{\boldsymbol{\alpha}}\bar{\boldsymbol{\gamma}}}(s)e^{-us} q_{\bar{\boldsymbol{\gamma}}\bar{\boldsymbol{\beta}}} e^{q_{\bar{\boldsymbol{\beta}}\bar{\boldsymbol{\beta}}}(t-s)} e^{-u(t-s)}ds = \bar{p}^{\langle n+1\rangle}_{\bar{\boldsymbol{\alpha}}\bar{\boldsymbol{\beta}}}(t)e^{-ut}. \end{aligned}$$

Hence,

$$\tilde{\bar{p}}_{\bar{\boldsymbol{\alpha}}_r\bar{\boldsymbol{\beta}}_r}(t) = \sum_{n=0}^{\infty} \bar{p}^{\langle n\rangle}_{\bar{\boldsymbol{\alpha}}\bar{\boldsymbol{\beta}}}(t)e^{-ut} = \bar{p}_{\bar{\boldsymbol{\alpha}}\bar{\boldsymbol{\beta}}}(t)e^{-ut}.$$

Thus, if $\tilde{M}_{\bar{\boldsymbol{\alpha}}_r}$ denotes the expected time the observed process spends in $\bar{\boldsymbol{\alpha}}_r$ (which depends on u), we have the following formula for the typicality:

$$\mathbf{g}_u(\bar{\boldsymbol{\alpha}}) = \int_0^\infty \bar{p}_{\bar{\boldsymbol{\kappa}}\bar{\boldsymbol{\alpha}}}(t)\cdot ue^{-ut}dt = u\int_0^\infty \tilde{\bar{p}}_{\bar{\boldsymbol{\kappa}}_r\bar{\boldsymbol{\alpha}}_r}(t)dt = u\tilde{M}_{\bar{\boldsymbol{\alpha}}_r}.$$

9 Computation of typicality using probabilities of paths

In this section, we give another method to compute the typicality measure $\mathbf{g}_u(\bar{\boldsymbol{\alpha}})$. For each path connecting the progenitor and $\bar{\boldsymbol{\alpha}}$, we will first compute the probability that the process passes this path. Then, the expected time the process spends in $\bar{\boldsymbol{\alpha}}$ is the sum of these probabilities times the expected time the process spends in $\bar{\boldsymbol{\alpha}}$ given that it has entered $\bar{\boldsymbol{\alpha}}$.

As previously, let $\mathbf{TS} = (T, l, \bar{\boldsymbol{\kappa}})$ be a transformation system. For each $\bar{\boldsymbol{\alpha}} \in TS$, $\bar{\boldsymbol{\beta}} = [\boldsymbol{\alpha} \lhd \boldsymbol{\tau}]$, where $\bar{\boldsymbol{\tau}} \in T$, let $y_{\bar{\boldsymbol{\alpha}}\bar{\boldsymbol{\beta}}}(\omega)$ be a random variable whose values are elementary paths beginning in $\bar{\boldsymbol{\alpha}}$ and ending in $\bar{\boldsymbol{\beta}}$ with the following distribution:

$$\mathcal{P}\{y_{\bar{\boldsymbol{\alpha}}\bar{\boldsymbol{\beta}}}(\omega) = [\boldsymbol{\alpha};\boldsymbol{\tau}']\} \stackrel{\text{def}}{=} \left(l(\bar{\boldsymbol{\tau}}') \sum_{[\boldsymbol{\alpha};\boldsymbol{\tau}'']:[\boldsymbol{\alpha}\lhd\boldsymbol{\tau}'']=\bar{\boldsymbol{\beta}}} \frac{1}{l(\bar{\boldsymbol{\tau}}'')}\right)^{-1} = (l(\bar{\boldsymbol{\tau}}')q_{\bar{\boldsymbol{\alpha}}\bar{\boldsymbol{\beta}}})^{-1}.$$

In other words, the probability that elementary path $[\boldsymbol{\alpha};\boldsymbol{\tau}']$ is the value of $y_{\bar{\boldsymbol{\alpha}}\bar{\boldsymbol{\beta}}}(\omega)$ is proportional to the inverse of $l(\bar{\boldsymbol{\tau}}')$. Assume that all $y_{\bar{\boldsymbol{\alpha}}\bar{\boldsymbol{\beta}}}(\omega)$, $\bar{\boldsymbol{\alpha}}, \bar{\boldsymbol{\beta}}$ as above, are independent of the process and of each other.

Let $\bar{p} = [\boldsymbol{\gamma};\boldsymbol{\tau}_1,\ldots,\boldsymbol{\tau}_n]$ be a path from P_T, and let $\boldsymbol{\gamma}_i = \boldsymbol{\gamma} \lhd \boldsymbol{\tau}_1 \lhd \ldots \lhd \boldsymbol{\tau}_i$ $(0 \le i \le n)$. Define a random variable

$$\eta_{\bar{p}}(\omega) \stackrel{\text{def}}{=} \begin{cases} 1, & \exists\, k \;\; \chi_{k+i}(\omega) = \bar{\boldsymbol{\gamma}}_i,\ 0 \le i \le n \;\; \text{and} \;\; y_{\bar{\boldsymbol{\gamma}}_i\bar{\boldsymbol{\gamma}}_{i+1}}(\omega) = [\boldsymbol{\gamma}_i;\boldsymbol{\tau}_{i+1}],\ 0 \le i < n \\ 0, & \text{otherwise.} \end{cases}$$

We will say that the process $\{x_t(\omega)\}$ *passes path* $\bar{p}$ if and only if $\eta_{\bar{p}}(\omega) = 1$. The probability that the process passes $\bar{p}$ is:

$$\mathcal{P}\{\eta_{\bar{p}}(\omega) = 1\} = \sum_{k=0}^{\infty} \mathcal{P}\{\chi_{k+i}(\omega) = \bar{\boldsymbol{\gamma}}_i,\ 0 \le i \le n\} \cdot \mathcal{P}\{y_{\bar{\boldsymbol{\gamma}}_i\bar{\boldsymbol{\gamma}}_{i+1}}(\omega) = [\boldsymbol{\gamma}_i;\boldsymbol{\tau}_{i+1}],\ 0 \le i < n\}.$$

By Markov property (1) and [3, p. 236],

$$\begin{aligned} &\mathcal{P}\{\chi_{k+i}(\omega) = \bar{\boldsymbol{\gamma}}_i,\ 0 \le i \le n\} = \\ &\quad \mathcal{P}\{\chi_k(\omega) = \bar{\boldsymbol{\gamma}}\} \cdot \mathcal{P}\{\chi_{k+1}(\omega) = \bar{\boldsymbol{\gamma}}_1 \,|\, \chi_k(\omega) = \bar{\boldsymbol{\gamma}}\} \cdots \mathcal{P}\{\chi_{k+n}(\omega) = \bar{\boldsymbol{\gamma}}_n \,|\, \chi_{k+n-1}(\omega)\} = \\ &\quad \mathcal{P}\{\chi_k(\omega) = \bar{\boldsymbol{\gamma}}\} \cdot \textstyle\prod_{i=0}^{n-1} \frac{q_{\bar{\boldsymbol{\gamma}}_i\bar{\boldsymbol{\gamma}}_{i+1}}}{-q_{\bar{\boldsymbol{\gamma}}_i\bar{\boldsymbol{\gamma}}_i}}. \end{aligned}$$

Next, from mutual independence of $\{y_{\bar{\boldsymbol{\alpha}}\bar{\boldsymbol{\beta}}}(\omega)\}$, we have

$$\mathcal{P}\{y_{\bar{\boldsymbol{\gamma}}_i\bar{\boldsymbol{\gamma}}_{i+1}}(\omega) = [\boldsymbol{\gamma}_i;\boldsymbol{\tau}_{i+1}],\, 0 \le i < n\} = \prod_{i=0}^{n-1} \mathcal{P}\{y_{\bar{\boldsymbol{\gamma}}_i\bar{\boldsymbol{\gamma}}_{i+1}}(\omega) = [\boldsymbol{\gamma}_i;\boldsymbol{\tau}_{i+1}]\} = \prod_{i=0}^{n-1}(-l(\bar{\boldsymbol{\tau}}_{i+1})q_{\bar{\boldsymbol{\gamma}}_i\bar{\boldsymbol{\gamma}}_{i+1}})^{-1}.$$

In the particular case of $\bar{\boldsymbol{\gamma}} = \bar{\boldsymbol{\kappa}}$,

$$\mathcal{P}\{\chi_k(\omega) = \bar{\boldsymbol{\kappa}}\} = \delta_{k0}.$$

Combining the last three formulas, we obtain for a path starting at $\bar{\boldsymbol{\kappa}}$:

$$\mathcal{P}\{\eta_{\bar{p}} = 1\} = \prod_{i=0}^{n-1}(-l(\bar{\boldsymbol{\tau}}_{i+1})q_{\bar{\boldsymbol{\gamma}}_i\bar{\boldsymbol{\gamma}}_i})^{-1}. \tag{11}$$

Now, the expected time the process spends in $\bar{\boldsymbol{\alpha}}$ can be computed using the following lemma:

Lemma 1.

$$M_{\bar{\boldsymbol{\alpha}}} = (-q_{\bar{\boldsymbol{\alpha}}\bar{\boldsymbol{\alpha}}})^{-1} \sum_{\bar{p}\in P_T:\mathrm{begin}(\bar{p})=\bar{\boldsymbol{\kappa}},\mathrm{end}(\bar{p})=\bar{\boldsymbol{\alpha}}} \mathcal{P}\{\eta_{\bar{p}}(\omega) = 1\}.$$

Proof. The proof is by induction on $\bar{\boldsymbol{\alpha}}$.

If $\bar{\boldsymbol{\alpha}} = \bar{\boldsymbol{\kappa}}$, then there exists only one path $\bar{p}$ such that $\mathrm{begin}(\bar{p}) = \mathrm{end}(\bar{p}) = \bar{\boldsymbol{\kappa}}$, namely, $[\boldsymbol{\kappa}]$. For this path, $\mathcal{P}\{\eta_{[\boldsymbol{\kappa}]}(\omega) = 1\} = 1$. Hence, according to the first formula in (10), the statement of the lemma holds.

Assume, the statement holds for all $\bar{\boldsymbol{\gamma}} < \bar{\boldsymbol{\alpha}}$. If $\bar{\boldsymbol{\alpha}} \ne \bar{\boldsymbol{\kappa}}$, then, by the second formula in (10),

$$\begin{aligned}
M_{\bar{\boldsymbol{\alpha}}} = (-q_{\bar{\boldsymbol{\alpha}}\bar{\boldsymbol{\alpha}}})^{-1} &\textstyle\sum_{\bar{\boldsymbol{\gamma}}<\bar{\boldsymbol{\alpha}}} M_{\bar{\boldsymbol{\gamma}}} q_{\bar{\boldsymbol{\gamma}}\bar{\boldsymbol{\alpha}}} = \\
(-q_{\bar{\boldsymbol{\alpha}}\bar{\boldsymbol{\alpha}}})^{-1} &\textstyle\sum_{\bar{\boldsymbol{\gamma}}<\bar{\boldsymbol{\alpha}}} \sum_{\bar{p}':\mathrm{begin}(\bar{p}')=\bar{\boldsymbol{\kappa}},\mathrm{end}(\bar{p}')=\bar{\boldsymbol{\gamma}}} \mathcal{P}\{\eta_{\bar{p}'}(\omega) = 1\} q_{\bar{\boldsymbol{\gamma}}\bar{\boldsymbol{\alpha}}} = \\
(-q_{\bar{\boldsymbol{\alpha}}\bar{\boldsymbol{\alpha}}})^{-1} &\textstyle\sum_{\bar{\boldsymbol{\gamma}}<\bar{\boldsymbol{\alpha}}} \sum_{\bar{p}':\mathrm{begin}(\bar{p}')=\bar{\boldsymbol{\kappa}},\mathrm{end}(\bar{p}')=\bar{\boldsymbol{\gamma}},\,[\boldsymbol{\gamma};\boldsymbol{\tau}]:[\boldsymbol{\gamma}\lhd\boldsymbol{\tau}]=\bar{\boldsymbol{\alpha}}} \mathcal{P}\{\eta_{\bar{p}'\blacktriangleleft[\boldsymbol{\gamma};\boldsymbol{\tau}]}(\omega) = 1\} = \\
(-q_{\bar{\boldsymbol{\alpha}}\bar{\boldsymbol{\alpha}}})^{-1} &\textstyle\sum_{\bar{p}\in P_T:\mathrm{begin}(\bar{p})=\bar{\boldsymbol{\kappa}},\mathrm{end}(\bar{p})=\bar{\boldsymbol{\alpha}}} \mathcal{P}\{\eta_{\bar{p}}(\omega) = 1\}.
\end{aligned}$$

(In the above formulas operation $\blacktriangleleft$ means composition of paths). ∎

Method presented in this section can be applied to compute the typicality of and element, if the latter has many ancestors but the paths leading to it are "symmetric" so that they make equal contributions to the typicality. We will compute the typicality measure for the above examples, using formula (11) and Lemma 1.

1. Transformation system of natural numbers ($\mathbf{TS}_{\mathbb{N}}$):

$$\mathbf{g}_u([\boldsymbol{\kappa} \lhd n\boldsymbol{\tau}_|]) = u \cdot (1+u)^{-1} \cdot \prod_{i=0}^{n-1}(1+u)^{-1} = \frac{u}{(1+u)^{n+1}}.$$

2. Transformation system of binary sequences ($\mathbf{TS}_{\mathbb{B}}$):

$$\mathbf{g}_u([\boldsymbol{\kappa} \lhd \boldsymbol{\tau}_{i_1} \lhd \ldots \lhd \boldsymbol{\tau}_{i_n}]) = u \cdot (2+u)^{-1} \prod_{i=0}^{n-1}(2+u)^{-1} = \frac{u}{(2+u)^{n+1}}.$$

3. Transformation system of strings ($\mathbf{TS}_{\mathbb{S}}$): for a struct $\bar{\boldsymbol{\gamma}}$ corresponding to a string of length n,

$$\mathbf{g}_u(\bar{\boldsymbol{\gamma}}) = u \cdot (2(n+1)+u)^{-1} \cdot n! \prod_{i=0}^{n-1} (2(i+1)+u)^{-1} = \frac{un!}{2^{n+1} \prod_{i=0}^{n} (i+1+u/2)}.$$

   In particular, if $u = 2$, we have

$$\mathbf{g}_u(\bar{\boldsymbol{\gamma}}) = \frac{1}{2^n(n+1)(n+2)} = \frac{1}{2^n} \left(\frac{1}{n+1} - \frac{1}{n+2} \right).$$

It is easy to verify that in each of the above three cases the typicality measure is a probability measure on the set of structs generating by the corresponding transformation system, i.e.,

$$\sum_{\bar{\boldsymbol{\gamma}} \in TS} \mathbf{g}_u(\bar{\boldsymbol{\gamma}}) = 1.$$

In general, this is true if and only if the original non-observed process has infinite running time with probability one.

10 Conclusion

We have formally defined the concept of generating process induced by a transformation system. We also introduced the notions of the time the process spends in a struct generated by the transformation system and of the (total) running time of the process. We have modelled the observation of the process by a random observer using the concept of observed process and, based on it, have defined the typicality measure.

This measure is intended to serve as a measure of the quality of training sets in the learning problem, which now can be understood informally as the problem of inference of the transformation system, given a finite training set generated by it. The next question is, of course, how to compare and relate to each other different transformation systems. Eventually this comparison will result in an optimization criterion, which will allow to choose the optimal transformation system corresponding to the given training set.

We compare two transformation systems by comparing their generative processes. In order to do that, it is not sufficient to consider only the finite part of the process—i.e., the set of structs generated before the stopping time $\tau(\omega)$, for the following intuitive reason: a process that "converges" to a single limiting state is significantly different from that converging to infinitely many states, even if the transformation matrices of these processes are identical. This means that the transformation matrix extracted from a transformation system contains only partial information about the corresponding process; the other part of information, perhaps even more significant one, is contained in the *topology* on the set of states, which is naturally induced by the transformation system, since the states are structs generated by it. Moreover, since the topology on the set of states can be defined through transformations, so are the limit states of the process, and therefore the process can be extended to the limit states. For example, transformation that inserts a letter into a string can be extended to “infinite strings”, i.e. the limit states of the process induced by $\mathbf{TS}_{\mathbb{S}}$. Preliminary considerations suggest that there are only countably many such limit states for strings, while binary sequences have a continuum of limit states corresponding to the points on the segment $[0, 1]$. This fact makes us believe that strings and binary sequences are generated by very different transformation systems, and neither of them induces a process which could be considered as a sub-process of the other.

The formal specification of the above topology, limit states, extension of transformations onto them, and comparison of transformation systems via morphisms will be given in a forthcoming paper.

References

[1] L. Goldfarb, O. Golubitsky, D. Korkin, What is a structural representation? Technical Report TR00-137, Faculty of Computer Science, U.N.B., October 2001.

[2] W. Feller, *An Introduction to Probability Theory and Its Applications*, Volume II, Second Edition, John Wiley & Sons, New York, 1966.

[3] K. Chung, *Markov Chains with Stationary Transition Probabilities*, Springer-Verlag, Berlin, 1960.

[4] P. Suppes, *Axiomatic Set Theory*, D. Van Nostrand, Princeton, 1960.