**Reading Notes: Alexandra M. Jurgens and James P. Crutchfield (2021), Divergent predictive states: The statistical complexity dimension of stationary, ergodic hidden Markov processes, Chaos 31, 083114**

**Summary**

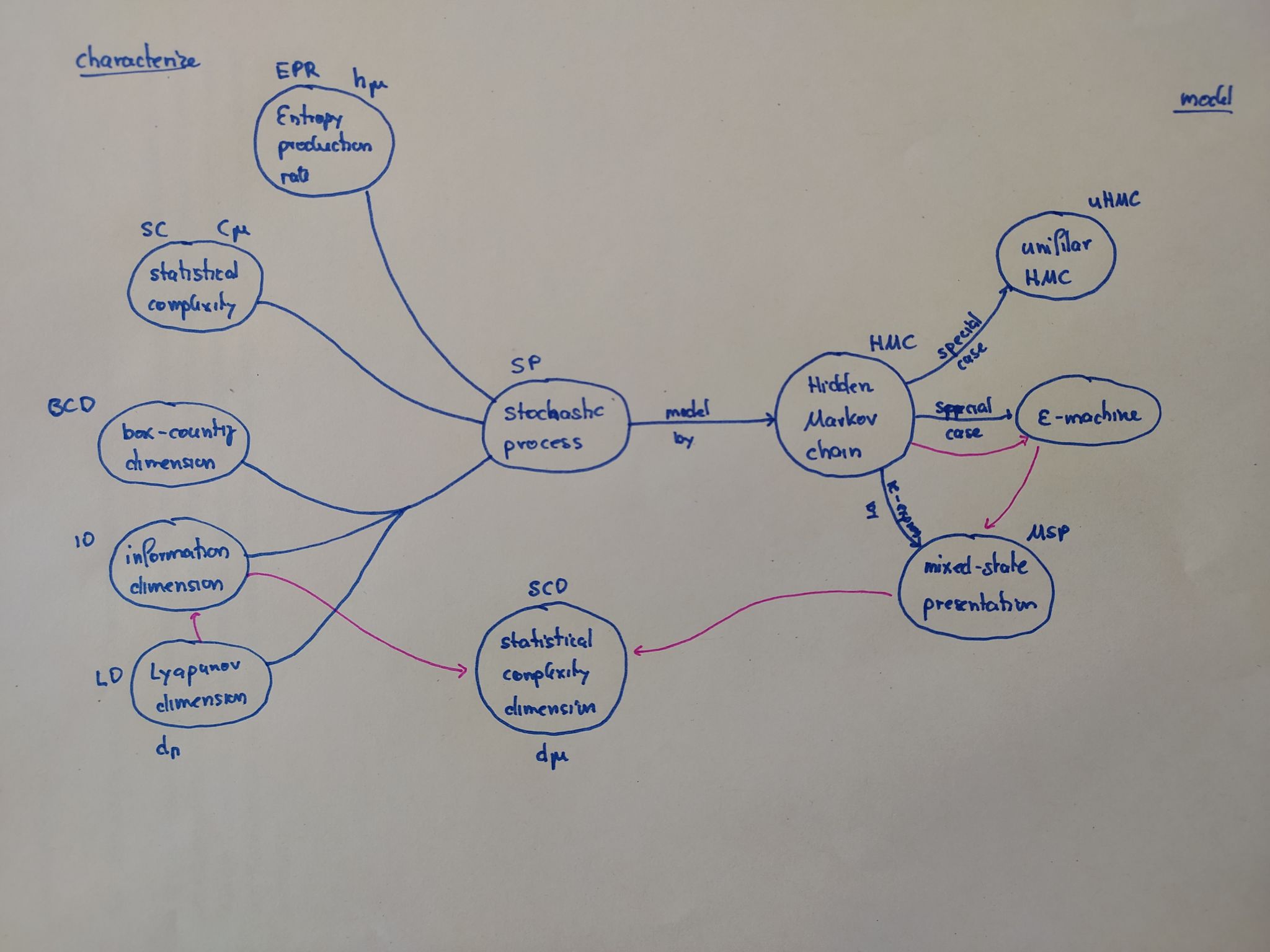
This paper addresses the question of how to characterize infinite-state hidden Markov chains, and provides a pathway for a solution.

**Problem statement and goal**

Stochastic processes (SP) can be characterized in terms of intrinsic randomness (typically measured by entropy production rate EPR), and intrinsic structuredness (also referred to as structural complexity). For the latter, the authors suggested 'statistical complexity' (SC) as a suitable measure in previous work.

However, even simply defined, finite state generators can produce stochastic processes which, if modelled by a hidden Markov chain (HMC), require **infinite** states for optimal prediction, which raises difficulties to characterize intrinsic randomness and structuredness. For the first, the authors provided a solution in previous work, and it is the goal of this paper to provide a solution for the latter, based on Shannons' concept of 'dimension rate'. The resulting measure they call 'statistical complexity dimension' (SCD).

**Solution pathway**: All key elements used in the solution, and their connections are shown in the sketch below. I first give a very colloquial explanation of each, and how they are connected, before then explaining how the authors use these elements to achieve their goal.



Stochastic process (SP)

* A probability measure over a bi-infinte chain of random variables. I.e. for each point in time, the variables are probability distributions, and their evolution from one step to the next involves randomness

Markov chain (MC)

* A Markov chain or Markov process is a stochastic model describing a sequence of possible events in which the probability of each event depends only on the state attained in the previous event.
* Markov property = "memorylessness". In other words, conditional on the present state of the system, its future and past states are independent.

Hidden Markov chain (HMC)

* A Markov chain with unobservable (hidden) states. Some (unobserved) variable x is a state variable of a Markov process, and this variable x influences some (observed) variable y of interest (output) in a probabilistic manner

Unifilar Hidden Markov chain (uHMC)

* Special case of a HMC, where for the transition matrices of all pairs of state variables, for any (discrete, finite) value the ingoing state variables assumes, there is only one (discrete, finite) value the outgoing variable can take.
* This establishes a deterministic relationship between the past and the future states of the HMC. In contrast to this, non-unifilar (generative) HMCs may generate varying future distributions, even if conditioned on the same past

ε-machine

* The smallest possible, maximally predictive uHMC representing a SP
* As the number of state variables in an ε-machine is the smallest possible for the given SP, their number serves as a measure of intrinsic structure/memory of the SP.

Mixed-state presentation (MSP)

* Observing the output of an HMC, one can from the observations make guesses about the internal state of the HMC at every time step. These guesses will in most cases be probabilistic and are called belief distributions.
* The belief distributions, together with the transition matrices that capture their evolution over time, define the HMCs mixed-state presentation.
* An MSP is unifilar by construction, and can therefore be used to turn an non-uifilar HMC to a unifilar one, potentially at the price of an infinite number of mixed states

Entropy (production) rate (EPR)

* The increase of entropy of a dynamical system per unit time (or timestep), conditioned on knowledge of its past. Measures the systems' intrinsic randomness.
* If the system can be modelled by a Markov chain, the EPR can be directly derived from its transition matrix.
* EPR provides a link between the underlying physics and timeseries statistics of a dynamical system: EPR values > 0 indicate chaotic/generative systems, = 0 indicate conservative systems, < 0 indicate dissipative systems

Statistical complexity (SC)

* If a SP is represented by its ε-machine, the intrinsic memory of the SP can be measured either by the number states of the ε-machine, or by the amount of historical Shannon entropy they store. The latter is called statistical complexity, and a unique measure for the structural complexity of a stochastic process. SC describes how much of a system's past is transmitted to the future via the system states at the present.
* For infinite-state HMCs, SC diverges and therefore becomes meaningless. This is the motivation to develop statistical complexity dimensions (SCD) as a substitute in such cases.

Box-counting dimension (BCD)

* A measure for the 'dimension' of a process, calculated by dividing the state space volume into bins of side length 'delta', and then counting the number of bins 'F' visited by the process trajectory at least once. This is done for increasingly fine-grained bins. BCD then is the limit of log(F)/log(delta) as delta goes to zero.

Information dimension (ID)

* A measure for the 'dimension' of a process, calculated by dividing the state space volume into bins of side length 'delta', fitting a (finite-state) Markov chain to the coarse-grained process, and then calculating the entropy of the transition matrix T of that chain. This is done for increasingly fine-grained bins. ID then is the limit of H(T)/log(delta) as delta goes to zero.
* For infinite-state HMCs, as delta goes to zero, ID is a divergence rate, which also describes the rate of divergence of statistical complexity SC. Thus, it can be used to expand the SC concept to infinite-state HMCs

Lyapunov exponent (LE), Lyapunov spectrum (LS), Lyapunov dimension (LD)

* A measure of a dynamical systems' dissipative nature (like EPR). Measures the rate of separation of initially very close trajectories (if the stochastic nature of the system is represented by an ensemble of deterministic trajectories) per time step. Specifically, LE is the exponent of an exponential function describing the rate of separation.
* The rate of separation can be different for different orientations of initial separation vector (orientation in state space). Thus, there is a spectrum of Lyapunov exponents (LS)—equal in number to the dimensionality of the phase space. Usually, only the largest is reported and used to characterize the system, as it will govern predictability, which is termed Lyapunov characteristic exponent (LCE)
* LE values > 1 indicate chaotic/generative systems, = 1 indicate conservative systems, < 1 indicate dissipative systems
* If calculated locally in time, a spectrum of LCE's will be obtained over time. The sum of the positive LCE's is the dynamical system's entropy production rate (EPR).
* From the LCE spectrum, a Lyapunov dimension (LD) can be calculated, and this is linked to the other measures of a processes' dimension (BCD, ID): For 'typical dynamical systems', LD equals ID.

Statistical complexity dimension (SCD)

* A unique measure of the statistical complexity of an infinite-state HMC, defined as the information dimension ID of the ε-machine of the processes' mixed-states presentation (MSP).
* It describes the divergence of memory/structural resources when attempting to optimally predict an infinite-state process
* For processes that can be optimally predicted with a finite number of states, SCD goes to zero and instead, statistical complexity SC can be applied.

**The solution**

* Establish the mixed-state presentation (MSP) ε-machine of the HMC
* Calculate the Lyapunov spectrum (LS) of the MSP
* From the LS, calculate the Lyapunov spectrum (LS) as an estimate of the information dimension (ID)
* The so-obtained ID is the divergence rate of the HMCs statistical complexity (and hence structural complexity), and can therefore serve as a unique measure for the HMCs structural complexity, even for infinite-state HMCs

**How this might be useful**

* Guide model development for cases where both dissipative and generative processes occur at the same time (typical for Earth Science systems)

**Open questions**

* Distinction of structure and state: in HMCs this is clear: states change with time, structure does not. In real-world ES systems, the distinction is less clear, rather we have an interplay of fast- and slow-evolving system properties. How can we apply the concept of statistical complexity dimension to such nonstationary, and non-ergodic systems (which was developed for stationary , ergodic systems)?
* How can we calculate the entropy rate and the SCD for forced systems with deterministic input (systems with dynamic input)?