Statistical Monitoring of Stochastic Processes

Statistical Monitoring

Given:
- *an alphabet* $\mathscr{X}$,
- *a stochastic process* $\overrightarrow{X} := (X_t)_{t\in\mathbb{N}}$ *over* $\mathscr{X}$,
- *a property* $\varphi : \mathscr{X}^\infty \to \mathbb{R}$,
- *and a confidence level* $\delta \in (0,1)$.

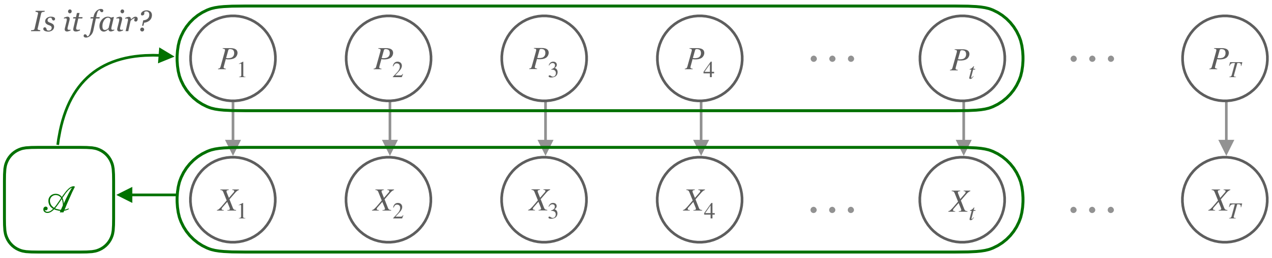
Develop a monitor $\mathscr{A}_\varphi : \mathscr{X}^* \to I_\mathbb{R}$ *such that*

$$\forall t \in \mathbb{N} : \mathbb{P}\left(\varphi(\overrightarrow{X}_t) \in \mathscr{A}_\varphi(\overrightarrow{X}_t)\right) \geq 1 - \delta$$

Or

$$\mathbb{P}\left(\forall t \in \mathbb{N} : \varphi(\overrightarrow{X}_t) \in \mathscr{A}_\varphi(\overrightarrow{X}_t)\right) \geq 1 - \delta$$

Example: Fair Coins



Given a sequence of coins $\overrightarrow{P} = (P_t)_{t\in\mathbb{N}}$ *and their corresponding tosses* $\overrightarrow{X} := (X_t)_{t\in\mathbb{N}}$. *At time* $t \in \mathbb{N}$ *the monitor observes the sequence of tosses up until t, i.e.,* $\overrightarrow{X}_t$.

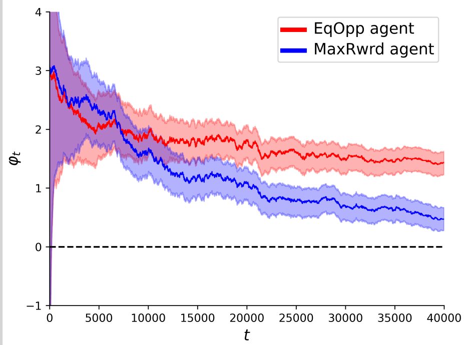
The questions we are interested in are among others:
(i) Was the sequence of chosen coins fair on average? (ii) Is the current coin fair? (iii) Will the sequence of coins be fair on average in the near future? (iv) What about the the limit? (v) How certain are we in our answer? (vi) What are the necessary assumptions?

Answering question (i) reduces to estimating the average bias of the past coins, i.e.,

$$\mathbb{P}\left(\forall t \in \mathbb{N} : \frac{1}{t}\sum_{i=1}^{t} P_i \in \mathscr{A}(\overrightarrow{X}_t)\right) \geq 1 - \delta$$

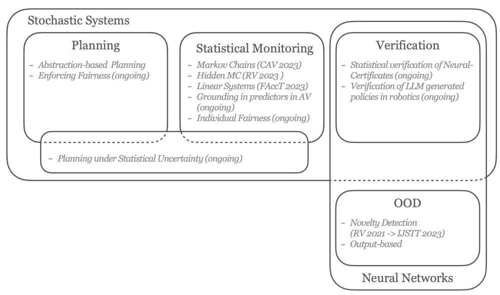
Applications

Statistical Monitoring can be used to detect
- *unfair behaviour of black-box systems, e.g., is a neural network biased;*
- *uncertainty in the model, e.g., is the (machine learning) model correct;*
- *undesirable feedback loops in dynamic systems, e.g., is a government policy leading to a degradation in well-being of a population over time (see image).*



Our Work

Using statistical methods, e.g., concentration inequalities, we developed monitors for monitoring algorithmic fairness beyond the i.i.d. setting extending it to Markov chains, Hidden Markov Models, and stochastic linear dynamic systems.



Thomas A. Henzinger | Mahyar Karrabi | Konstantin Kueffner | Kaushik Mallik

Institute of Science and Technology Austria