How Can One Test if a Binary Sequence Is Exchangeable? Fork-convex Hulls, Supermartingales, and Snell Envelopes

Aaditya Ramdas ARAMDAS@CMU.EDU
Statistics and Machine Learning, Carnegie Mellon University
Johannes Ruf J.RUF@LSE.AC.UK
Mathematics, London School of Economics
Martin Larsson MARTINL@ANDREW.CMU.EDU
Mathematics, Carnegie Mellon University
Wouter M. Koolen WMKOOLEN@CWI.NL
Machine Learning Group, CWI Amsterdam

Suppose we observe a sequence of binary coin flips $X_1, X_2, \ldots$ one at a time. We consider the fundamental problem of sequentially testing if our data $(X_t)_{t\geq 1}$ is an exchangeable sequence:

$$H_0 : X_1, X_2, \ldots \text{ are exchangeable.}$$

Nonnegative supermartingales (NSMs) are a workhorse of sequential inference [1], but we prove that they are powerless for this problem. First, utilizing a geometric concept called fork-convexity — which can be thought of as a sequential analog of convexity — we show that any process which is an NSM under two distributions is also necessarily an NSM under their "fork-convex hull". Second, we demonstrate that the fork-convex hull of the exchangeable null consists of all possible laws over binary sequences. This implies that any NSM under exchangeability is necessarily nonincreasing, hence always yields a powerless test for any alternative.

Since testing arbitrary deviations from exchangeability is information theoretically impossible, we focus on Markovian alternatives. We combine ideas from universal inference [5] and the method of mixtures to derive a "safe e-value". For the unfamiliar, $(R_t)_{t\geq 1}$ is called a safe e-value if it is a nonnegative process with expectation at most one under the null at any stopping time, meaning that

$$\sup_{\mathsf{Q}\in\overline{\mathscr{Q}}} \sup_{\tau} \mathbb{E}_{\mathsf{Q}}[R_\tau] \leq 1,$$

where the second supremum is over all stopping times τ, possibly infinite and $\overline{\mathscr{Q}}$ is the set of all distributions resulting in exchangeable sequences. We develop such an (R_t) that is upper bounded by a martingale, but is not itself an NSM.

Rejecting the null whenever R_t crosses $1/\alpha$ in turn yields a level α sequential test that is consistent, meaning that its power tends to one with sample size. Further, regret bounds from universal coding also demonstrate rate-optimal power. We present ways to extend these results to any finite alphabet and to Markovian alternatives of any order using a "double mixture" approach. We provide a wide array of simulations, and give general approaches based on betting for unstructured or ill-specified alternatives.

Finally, inspired by the book of Shafer and Vovk [2], and the original results of Ville [3], we provide game-theoretic interpretations of our e-values and pathwise results. Our results complement the recent advances made by Vovk [4] on this topic using a very different approach.

References

[1] Aaditya Ramdas, Johannes Ruf, Martin Larsson, and Wouter Koolen. Admissible anytime-valid sequential inference must rely on nonnegative martingales. *arXiv:2009.03167*, 2020.

[2] Glenn Shafer and Vladimir Vovk. *Game-Theoretic Foundations for Probability and Finance*, volume 455. John Wiley & Sons, 2019.

[3] J Ville. *Étude Critique de la Notion de Collectif.* Gauthier-Villars, Paris, 1939.

[4] Vladimir Vovk. Testing randomness online. *Statistical Science*, 2021.

[5] Larry Wasserman, Aaditya Ramdas, and Sivaraman Balakrishnan. Universal inference. *Proceedings of the National Academy of Sciences*, 2020. ISSN 0027-8424.