MSc Projects CWI's Machine Learning Group

Best Arm Identification in Survival Models

Supervisor: Wouter M. Koolen
Keywords: Pure Exploration, Bandits, Survival Analysis, Hypothesis Testing

Background: Pure Exploration is an active area of Machine Learning and Statistics. Its central problem of Best Arm Identification has been studied intensively since (Even-Dar, Mannor, and Mansour, 2002), and worst-case optimal methods have been developed for the fixed confidence, fixed budget and simple regret settings (Bubeck, Munos, and Stoltz, 2011). After a long and respectable series of papers establishing worst-case optimality, a revolutionary new approach called Track-and-Stop was pioneered by Garivier and Kaufmann (2016) that delivers instance-optimal methods. Since then, several aspects of Track-and-Stop have been generalised and refined: tighter stopping thresholds were constructed by Kaufmann and Koolen (2021), computational efficiency was improved using saddle-point methods by Degenne, Koolen, and M´enard (2019), and problems with multiple answers were analysed by Degenne and Koolen (2019).

Academic Content: In this project the Best Arm Identification problem will be extended to active testing in survival analysis models. Two complications arise when transitioning from a standard bandit to survival analysis models: data now arrive with delay, and are right-censored. There is extensive recent work on bandits with delayed feedback, both in the regret and pure exploration settings, by Vernade, Capp´e, and Perchet (2017), Pike-Burke et al. (2018), Vernade, Carpentier, Zappella, et al. (2018), Grover et al. (2018), and Vernade, Carpentier, Lattimore, et al. (2020). Bandits with censored feedback are considered by Abernethy, Amin, and Zhu (2016), though the literature seems much less developed. Passive hypothesis testing with survival data (including optional stopping problems) has recently been considered by Grunwald et al. (2020).

The project will focus on developing the model and theory, following these steps:
1. Review literature on Best Arm Identification, in particular the Track-and-Stop methodology.
2. Review literature on nonparametric survivial models, in particular the Cox proportional hazards model.
3. Study the active testing problem (BAI) for survival models. Defining carefully the interaction protocol and metric. Gain experience by addressing the BAI problem for survival models but without delays and censoring.
4. Extend the Track-and-Stop method to active testing in survival models with delays and censoring. The project is envisaged to consist of mostly theoretical work, with only minor computational and empirical components.

References

Abernethy, J. D., K. Amin, and R. Zhu (2016). “Threshold bandits, with and without censored feedback”. In: Advances In Neural Information Processing Systems 29, pp. 4889–4897.

Bubeck, S., R. Munos, and G. Stoltz (2011). “Pure Exploration in Finitely Armed and Continuous Armed Bandits”. In: Theoretical Computer Science 412, 1832-1852 412, pp. 1832–1852.

Degenne, R. and W. M. Koolen (Dec. 2019). “Pure Exploration with Multiple Correct Answers”. In: Advances in Neural Information Processing Systems (NeurIPS) 32, pp. 14591–14600.

Degenne, R., W. M. Koolen, and P. M´enard (Dec. 2019). “Non-Asymptotic Pure Exploration by Solving Games”. In: Advances in Neural Information Processing Systems (NeurIPS) 32, pp. 14492–14501.

Even-Dar, E., S. Mannor, and Y. Mansour (2002). “PAC Bounds for Multi-armed Bandit and Markov Decision Processes”. In: Computational Learning Theory, 15th Annual Conference on Computational Learning Theory, COLT 2002, Sydney, Australia, July 8-10, 2002, Proceedings. Vol. 2375. Lecture Notes in Computer Science, pp. 255–270.

Garivier, A. and E. Kaufmann (2016). “Optimal Best arm Identification with Fixed Confidence”. In: Proceedings of the 29th Conference On Learning Theory (COLT).

Grover, A., T. Markov, P. Attia, N. Jin, N. Perkins, B. Cheong, M. Chen, Z. Yang, S. Harris, W. Chueh, et al. (2018). “Best arm identification in multi-armed bandits with delayed feedback”. In: International Conference on Artificial Intelligence and Statistics. PMLR, pp. 833–842.

Grunwald, P., A. Ly, M. Perez-Ortiz, and J. ter Schure (2020). “The Safe Log Rank Test: Error Control under Optional Stopping, Continuation and Prior Misspecification”. In: arXiv preprint arXiv:2011.06931.

Kaufmann, E. and W. M. Koolen (Nov. 2021). “Mixture Martingales Revisited with Applications to Sequential Tests and Confidence Intervals”. In: Journal of Machine Learning Research 22.246, pp. 1–44.

Pike-Burke, C., S. Agrawal, C. Szepesvari, and S. Grunewalder (2018). “Bandits with delayed, aggregated anonymous feedback”. In: International Conference on Machine Learning. PMLR, pp. 4105–4113.

Vernade, C., O. Capp´e, and V. Perchet (2017). “Stochastic bandit models for delayed conversions”. In: arXiv preprint arXiv:1706.09186.

Vernade, C., A. Carpentier, T.Lattimore, G. Zappella, B. Ermis, and M. Brueckner (2020). “Linear bandits with stochastic delayed feedback”. In: International Conference on Machine Learning. PMLR, pp. 9712–9721.

Vernade, C., A. Carpentier, G. Zappella, B. Ermis, and M. Brueckner (2018). “Contextual Bandits under Delayed Feedback.” In: arXiv preprint arXiv:1807.02089.

MSc Project: ϵ-Best Mixture Identification

Supervisor: Wouter M. Koolen
Keywords: Sequential Decision Making, Bandits, Pure Exploration, Sample Complexity, Mixture Policy Identification

Background: Pure Exploration is an active area of Machine Learning and Statistics. Its central problem of Best Arm Identification has been studied intensively since (Even-Dar, Mannor, and Mansour, 2002), and worst-case optimal methods have been developed for the fixed confidence, fixed budget and simple regret settings (Bubeck, Munos, and Stoltz, 2011). After a long and respectable series of papers establishing worst-case optimality, a revolutionary new approach called Track-and-Stop was pioneered by Garivier and Kaufmann (2016) that delivers instance-optimal methods. Since then, several aspects of Track-and-Stop have been generalised and refined: tighter stopping thresholds were constructed by Kaufmann and Koolen (2021), computational efficiency was improved using saddle-point methods by Degenne, Koolen, and M´enard (2019), and problems with multiple answers were analysed by Degenne and Koolen (2019). A recent extension with subpopulations was proposed by Russac et al. (2021).

Academic Content: The typical pure exploration task is identification of the best arm from samples. Letting μk denote the mean of arm k, the task is to identify arg $\max_k \mu k$. The sample complexity for this problem is well studied. One may reduce the sample complexity — at the cost of incurring some approximation error ϵ — by asking for identification of any ϵ-best arm $k \in \{k \mid \mu k \geq \max k \, \mu k - \epsilon\}$.
The starting point of this project is the realisation that for many applications it is enough to identify an ϵ-optimal mixture policy. That is, we are happy with any probability distribution p

$$\left\{ p \in \triangle_K \middle| \sum_{k=1}^{K} p_k \mu_k \geq \max_k \mu_k - \epsilon \right\}.$$

One may think of p as a randomised policy that is close to optimal. This could find application e.g. when applying bandit techniques to reinforcement learning problems, in particular near-optimal policy identification in MDPs.
This project will investigate these questions: Is the sample complexity of ϵ-best mixture identification strictly lower than that of ϵ-best arm identification? If so, what is the sample complexity of ϵ-best mixture identification? And how can one design efficient algorithms for ϵ-best mixture identification? Possible algorithms include elimination methods based on confidence-intervals, Track-and-Stop and Bayesian-flavoured approaches (Kaufmann, Koolen, and Garivier, 2018).

The project is envisaged to consist of mostly theoretical work, with only minor computational and empirical components.

References

Bubeck, S., R. Munos, and G. Stoltz (2011). “Pure Exploration in Finitely Armed and Continuous Armed Bandits”. In: Theoretical Computer Science 412, 1832-1852 412, pp. 1832–1852.

Degenne, R. and W. M. Koolen (Dec. 2019). “Pure Exploration with Multiple Correct Answers”. In: Advances in Neural Information Processing Systems (NeurIPS) 32, pp. 14591–14600.

Degenne, R., W. M. Koolen, and P. M´enard (Dec. 2019). “Non-Asymptotic Pure Exploration by Solving Games”. In: Advances in Neural Information Processing Systems (NeurIPS) 32, pp. 14492–14501.

Degenne, R., H. Shao, and W. M. Koolen (July 2020). “Structure Adaptive Algorithms for Stochastic Bandits”. In Proceedings of the 37th International Conference on Machine Learning (ICML).

Even-Dar, E., S. Mannor, and Y. Mansour (2002). “PAC Bounds for Multi-armed Bandit and Markov Decision Processes”. In: Computational Learning Theory, 15th Annual Conference on Computational Learning Theory, COLT 2002, Sydney, Australia, July 8 10, 2002, Proceedings. Vol. 2375. Lecture Notes in Computer Science, pp. 255–270.

Garivier, A. and E. Kaufmann (2016). “Optimal Best arm Identification with Fixed Confidence”. In: Proceedings of the 29th Conference On Learning Theory (COLT).

Katariya, S., B. Kveton, C. Szepesv´ari, C. Vernade, and Z. Wen (2017). “Stochastic Rank-1 Bandits”. In: Proceedings of the 20th International Conference on Artificial Intelligence and Statistics, AISTATS 2017, 20-22 April 2017, Fort Lauderdale, FL, USA. Vol. 54. Proceedings of Machine Learning Research, pp. 392–401.

Kaufmann, E. and W. M. Koolen (Nov. 2021). “Mixture Martingales Revisited with Applications to Sequential Tests and Confidence Intervals”. In: Journal of Machine Learning Research 22.246, pp. 1–44.

Kaufmann, E., W. M. Koolen, and A. Garivier (Dec. 2018). “Sequential Test for the Lowest Mean: From Thompson to Murphy Sampling”. In: Advances in Neural Information Processing Systems (NeurIPS) 31, pp. 6333–6343.

Russac, Y., C. Katsimerou, D. Bohle, O. Capp´e, A. Garivier, and W. M. Koolen (Dec. 2021). “A/B/n Testing with Control in the Presence of Subpopulations”. In: Advances in Neural Information Processing Systems (NeurIPS) 34.

MSc Project: Life-long Online Learning

Supervisor: Wouter M. Koolen
Keywords: Sequential Decision Making, Online Learning, Expert Setting

Background: A fundamental task in sequential decision making is Prediction with Expert Advice, also known as Decision Theoretic Online Learning or the Hedge Setting (Freund and Schapire, 1997). Here a decision maker plays a game of T rounds, picking one of K actions every round, and incurring a bounded loss. The objective is the regret compared the best fixed action in hindsight. The minimax regret is known to scale as $\sqrt{T \ln K}$. Many variations and extensions of the classic Hedge algorithm have been developed, featuring refined notions of time (de Rooij et al., 2014), and adaptivity to the complexity of the action set (Koolen and van Erven, 2015), etc.

Academic Content: This project explores what can be achieved in a long sequence of M Hedge interactions. It could be considered an investigation into "life-long-learning". Of course, one may play the M games with overall worst-case summed regret at most

$$M\sqrt{T \ln K}$$

The question raised here is whether one can adapt to the frequency distribution of the games' best actions.

1. We first ask what one can achieve if the distribution of winning actions were known. Say action k is the best in fraction p_k of the games. In this sense $\vec{p}$ could be considered a *prior distribution* on actions. Can one play the M games to achieve regret bounded by something of the form

$$M\sqrt{T \operatorname{cost}(\vec{p})}$$

and what would be the right cost function? A first, intuitive suggestion could be the Shannon entropy, but this is shown admissible but suboptimal by Koolen (2013). There are likely deep connections to the concept of mixability introduced by Vovk (1998).

2. We then ask what one can achieve if the fraction p is not known a priori. How can one still adapt to it, and how can one quantify the cost for learning p?

3. Finally, we ask the same question in the multi-scale extension of the setting, where recent adaptivity breakthroughs were made by P´erez-Ortiz and Koolen (2022).

The project is primarily expected to develop new theory: results include algorithms with performance guarantees as well as lower bounds. The project may benefit significantly from exploratory programming skills.

References

De Rooij, S., T. van Erven, P. Gr¨unwald, and W. M. Koolen (Apr. 2014). “Follow the Leader If You Can, Hedge If You Must”. In: Journal of Machine Learning Research 15, pp. 1281 1316.

Freund, Y. and R. E. Schapire (1997). “A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting”. In: Journal of Computer and System Sciences 55, pp. 119–139.

Koolen, W. M. (Dec. 2013). “The Pareto Regret Frontier”. In: Advances in Neural Information Processing Systems (NeurIPS) 26, pp. 863–871.

Koolen, W. M. and T. van Erven (June 2015). “Second-order Quantile Methods for Experts and Combinatorial Games”. In: Proceedings of the 28th Annual Conference on Learning Theory (COLT), pp. 1155–1175.

P´erez-Ortiz, M. and W. M. Koolen (Dec. 2022). “Luckiness in Multiscale Online Learning”. In: Advances in Neural Information Processing Systems (NeurIPS) 35. Accepted.

Vovk, V. (1998). “A Game of Prediction with Expert Advice”. In: Journal of Computer and System Sciences 56.2, pp. 153–173.

Various Projects on E-Values, Always-Valid Confidence Sequences and "Safe Testing"

Supervisor: Peter Grünwald
Keywords: Testing; uncertainty quantifications; foundations of statistics and machine learning

How much evidence do the data give us about one hypothesis versus another? The standard way to measure evidence is still the p-value, despite a myriad of problems surrounding it – problems which are one of the reasons for the ongoing *replicability crisis* in the applied sciences such as medicine and psychology: the fraction of published results that are irreproduceable (which often just means 'wrong') is much higher than one would hope.

The e-value is a recently popularized notion of evidence which overcomes some of the issues with p-values. While e-values have lain dormant until 2019, interest in them has recently exploded with papers in the world's top machine learning conferences and statistics journals. In June 2022 we held a first international workshop on e-values with attendees from the areas of clinical trial design and meta-analysis but also from some of the large tech companies interested in A/B testing.

Unlike p-values, E-values allow for tests with strict ' classical' Type-I error control under *optional continuation and combination of data from different sources*. They are also easier to interpret than p-values, having a straightforward interpretation in terms of sequential betting.

E-values are also the basic building blocks of *anytime-valid confidence intervals* that remain valid under optional stopping and that are crucial for gaining trustworthy uncertainty quantification in e.g. A/B testing and bandit settings. In simple cases, inference based on evalues coincides with a particular Bayesian method, the *Bayes factor*. But if the null is composite or nonparametric, or an alternative cannot be explicitly formulated, e-values and Bayes factors become distinct and e-processes can be seen as a generalization of nonnegative supermartingales, a central topic in stochastic process theory.

The theory of E-Values is still very young, so many types of projects are possible. Here are a few examples:
- Design and implementation (in R or Python) of E-Values for Cox Regression, standard Regression, Mixed Models, Confidence Sequences for Effect Size in stratified contingency tables
- Comparison of different existing E-Values for the 'Model-X' Conditional Independence Tests. Investigating the claim that the Model-X assumption is unavoidable
- Comparing the GRAPA and the REGROW design principles for e-variables
- (more theoretical) Investigating the relation between E-variables and the Likelihood Principle

Learning to Attend to Classify

Supervisor: Sander Bohté
Keywords: biologically plausible deep learning; attention in humans and machines

In humans, attention focuses neural resources on a limited part of the sensory experience. Psychophysics also tells us that we only learn about that to which we attend. In deep learning, attention models are typically applied to sequence learning, where attention dynamically masks part of the stream [1]. Can we model the biological kind of attention to learn more efficiently?

[1] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). Attention is all you need. In *Advances in Neural Information Processing Systems (*pp. 5998-6008).