Imprecise Probabilities in Statistics and Conformal Change Detection: Elementary Results and an Open Question*

Vladimir Vovk V.VOVK@RHUL.AC.UK
Department of Computer Science, Royal Holloway, University of London, UK

IID probability, exchangeability probability, and Cournot's principle Let $\Omega := \{0,1\}^N$ be the set of all sequences of N binary observations. The *upper IID probability* of a set $E \subseteq \Omega$ is $\mathbb{P}^{\text{iid}}(E) := \sup_{p\in[0,1]} B_p^N(E)$, where B_p is the Bernoulli probability measure on $\{0,1\}$ with $B_p(\{1\}) = p$. The *upper exchangeability probability* of $E \subseteq \Omega$ is $\mathbb{P}^{\text{exch}}(E) := \sup_P P(E)$, P ranging over the exchangeable probability measures on Ω. (We will not use the corresponding lower probabilities.)

The function $\mathbb{P}^{\text{iid}}$ can be used when testing the IID hypothesis: if $\mathbb{P}^{\text{iid}}(E)$ is small (say, below 5% or 1%) and the observed sequence ω is in E that is chosen in advance, we are entitled to reject the hypothesis that the observations in ω are IID. Similarly, $\mathbb{P}^{\text{exch}}$ can be used when testing the hypothesis of exchangeability. This is *Cournot's principle*.

Proposition 1 *For any* $E \subseteq \Omega$, $\mathbb{P}^{\text{iid}}(E) \le \mathbb{P}^{\text{exch}}(E) \le 1.5\sqrt{N}\,\mathbb{P}^{\text{iid}}(E)$.

This proposition follows easily from Stirling's formula. (Detailed proofs are given in [1].) The difference between $\mathbb{P}^{\text{iid}}$ and $\mathbb{P}^{\text{exch}}$ (which is $O(\log N)$ on the log scale) has been regarded by some writers on the algorithmic theory of randomness (including Andrei Kolmogorov) as small enough to be disregarded.

Conformal change detection Let us now drop the assumption that the observations are binary, replacing Ω by $\mathbf{Z}^N$ for any measurable space $\mathbf{Z}$. A standard assumption in machine learning and nonparametric statistics is that the data are generated in the IID fashion. How can we test this assumption? *Conformal change detection* (or *anomaly detection* in the terminology of [2]) is a way of turning standard machine-learning algorithms into stochastic processes, called *conformal martingales*, that are martingales under any IID distribution. This method, which is even used in practice [2], is really simple, and all details will be given in the poster. A conformal martingale consists of random variables S_n, $n = 0,1,\dots$, that depend not only on $z_1,\dots,z_n$ but also on internal coin tossing. We only consider nonnegative conformal martingales with $S_0 > 0$. Ville's inequality says that, for any $c > 1$, $\mathbb{P}(\exists n : S_n/S_0 \ge c) \le 1/c$ under any IID distribution. We can interpret S_n/S_0 directly as the amount of evidence detected against the first n observations being IID.

A typical example of change detection is where we observe attacks, which we assume to be IID, on a computer system. When a new kind of attacks appears, the process ceases to be IID, and we would like to raise an alarm soon afterwards. Raising an alarm when S_n/S_0 exceeds a threshold $c > 1$ makes sure that the probability of a false alarm is at most $1/c$.

Conformal probability A natural question is how much we can potentially lose when using conformal martingales as compared with unrestricted testing (with either $\mathbb{P}^{\text{iid}}$ or $\mathbb{P}^{\text{exch}}$). The *upper conformal probability* of $E \subseteq \mathbf{Z}^N$ is $\mathbb{P}^{\text{conf}}(E) := \inf\{S_0 : S_N \ge 1_E \text{ a.s.}\}$, where S ranges over the nonnegative conformal martingales, 1_E is the indicator of E, and "a.s." refers to the internal coin tossing. The smallness of $\mathbb{P}^{\text{conf}}(E)$ means not only that E is unlikely under the IID assumption (by Ville's inequality) but also that this fact can be detected by conformal martingales. Now let us return to the binary case $\mathbf{Z} := \{0,1\}$.

Proposition 2 *For any* $E \subseteq \Omega$, $\mathbb{P}^{\text{iid}}(E) \le \mathbb{P}^{\text{conf}}(E) \le N\,\mathbb{P}^{\text{exch}}(E)$.

Open question: Can the coefficient N be improved?

References

[1] Vladimir Vovk. Power and limitations of conformal martingales, On-line Compression Modelling project (New Series), Working Paper 24. June 2019. URL http://alrw.net.

[2] Xiao Zhang, Peter Lu, Josée Martens, Gary Ericson, and Kent Sharkey. Time Series Anomaly Detection module in Microsoft Azure, May 2019. URL https://docs.microsoft.com/en-gb/azure/machine-learning/studio-module-reference/time-series-anomaly-detection.

* Many thanks to the referees for their valuable comments. This research has been supported by Astra Zeneca and Stena Line.