5. Willems JC. Thoughts on system identification. *Control of Uncertain Systems: Modelling, Approximation, and Design: Springer Lecture Notes in Control and Information Sciences* 2006; 329: 389–416
6. Willems JC. The behavioral approach to open and interconnected systems. *Control Sys Mag* 2007; 27: 46–99

Final Comments by the Author M.C. Campi

It has been a pleasure receiving and reading the discussion articles three outstanding scientists, namely, S. K. Mitter, A. Nemirovski and J. C. Willems have generously provided. Their comments increase the value of the discussion.

In my article [1], I presented simple examples to get facts and concepts underlying randomized algorithms easily through. Arkadi Nemirovski presents nontrivial examples in optimization which suitably complement those in my article and add concreteness to the role of randomized algorithms in difficult problems.

Sanjoy Mitter's comments open up new directions where randomization is used and broaden the scope of the discussion in a significant manner. I would like here to only add some remarks about the PAC learning methodology Mitter mentions in his comment.

1. The Lesson of PAC Learning

In PAC learning, [2], one is given a set of data and is asked to select a hypothesis from a given class of functions. The acronym PAC, probably approximately correct, refers to the fact that, with *high probability*, the selected function must have low generalization error that is it is *approximately correct*. PAC learning offers in a sense a broader set-up than it is usually done in system identification, while it is narrower in another way. It is broader because it assumes little or nothing about the underlying distribution that generates the data, and it is narrower because data are assumed to be independent one of the other. I believe that importing this framework into system identification is most desirable, the main challenge in this process being the treatment of dynamics.

From a more general point of view, in my appreciation the PAC paradigm teaches us an important lesson: meaningful results can be achieved *in the presence of little prior information* provided that we relax our requirements for precision: we should be content with answers which are approximately correct most of the time. Both aspects, *approximate correctness* and *most of the time* as opposed to always are intimately tied to the generality of the approach. Randomized methods are a means to pursue this philosophy, and this is one reason why randomized methods provide powerful tools of synthesis, see e.g. the examples in the discussion paper of A. Nemirovski.

Turning to Jan Willems' discussion article, I would like to say that I have had the opportunity to talk more than once with him on topics related to probability and its interpretation, and I share his view that justifying probabilistic models is most important and yet this issue has at times been underestimated by the control and systems community. This is particularly evident in the field of system identification, where much of my interest lies. Educating ourselves to the use of probability is a priority and, in my opinion, this is even more important than replacing probability with alternative models.

2. The Need for Probability

A central issue in filtering and identification is to provide guarantees in the form of quantitative statements capable to credit an estimation result with reliability. This is relevant to the practice of these methods and it is necessary for their scientific use. A single estimate (point estimation) is unsuitable to the purpose of providing guarantees since a single estimate can hardly be announced to be the true value. Intervals and regions have to be used instead[1]. However, claims like "the true parameter value *certainly* lies in this interval" can only be made under very stringent assumptions on the noise, assumptions that are difficult to justify from a modeling point of view. We therefore see that we have more realistically to seek guarantees valid for most of the noise sequences, not for all of them, and our goal is to look for regions having this property. However, if we really want to be quantitative, we need to provide ourselves with a mathematical tool to measure the "extension" of these regions, and this is measure theory; and when

[1] Interval estimation is a very well-established field in statistics, pioneered by masters such as J. Neyman, [3], A. Wald, [4], and J. W. Tukey, [5].

a measure is interpreted as “chance of happening”, this measure is called probability. So, I believe, renouncing to use probability leaves us unarmed to tackle the challenge to be quantitative.

3. A Cautious Use of Probability

While probability is an essential element in the formulation of many problems, I again agree with Jan Willems that it is also true that probability should be used with care and it is preferable to design methods whose probabilistic justification can be made at different levels, that is with different degrees of probabilistic knowledge. Kalman filtering is one such examples: in a highly structured Gaussian framework, a Kalman filter computes the conditional mean, the best *nonlinear* estimation in norm 2. However, only assuming knowledge of the second-order moments still permits one to justify the Kalman filter equations as a recursive method to derive the best *linear* norm 2 estimation.

There are fields where so-called distribution-free results are common. In Probably Approximately Correct (PAC) learning theory, for example, beautiful results have been derived that are valid independently of the underlying probability distribution. Thus, a probabilistic model is needed to justify the approach, but the results hold independently of the probability distribution that generates the data. To a similar result one can arrive in prediction using Interval Predictor Models (IPMs), as done in [6]. Justifying a result, algorithm or method without a full description of the underlying probabilistic model is important simply because this result, algorithm or method becomes more widely applicable and it gains credibility.

Randomized methods have a very special role in this discourse. As I pointed out in my article [1] the probability used in a randomized method is not introduced for the purpose of modeling, it is instead part of the algorithm, we create it for use in the algorithm and therefore we have no reason to doubt its validity. This is a significant positive mark in favor of the use of randomization.

4. Randomization and the Probability of Success

Jan Willems says that it has been difficult for him to follow the thesis in [1] when it comes to control. I take this opportunity to better clarify my view.

When we board a plane we would in principle like the idea that the plane crashes did not exist. However, planes do crash, and it is a relief to know that statistically a plane only crashes once every some 10^6 to 10^7 flights. If a full guarantee is not possible, we seek for a probabilistic guarantee, and this leads to a shift of the notion of robustness.

Further commenting on this point, I would like here to express doubts as to whether a full guarantee has always to be preferred to a probabilistic guarantee. Full guarantee is with respect to an assumed level of uncertainty, but how sure can we be about the level of uncertainty assumed? A robust design is tailored to the prescribed uncertainty level and falls apart if uncertainty is not in the prescribed uncertainty set. At times, it may be convenient to introduce a larger uncertainty set to cover more situations for which a robust design is not possible and allow for a small probability of failure. This approach may possibly lead to a design that safeguards against more situations simply because the algorithm accounts for all the situations in the larger uncertainty set when doing the design. Voltaire said “doubt is not a pleasant condition, but certainty is absurd”; I personally tend to look suspiciously at any theory expressing certainty.

5. About a Bayesian Perspective

In Section 4, Jan Willems comments about the Bayesian perspective that I have in Section 2.3 of my article [1] and says that a situation where “the designer knows exactly the relative frequency of the various plants, but seems to be unable to actually measure the unknown parameters in the actual plant” is unnatural. His point is my point. In Section 2.3, I argue that the big difference between the randomized approach and the Bayesian perspective is that in the latter poor modeling is a constant risk, a thing that cannot happen with the artificial probability introduced for use in a randomized algorithm.

References

1. Campi MC. Why is resorting to fate wise? A critical look at randomized algorithms. *Eur. J. Control*, 2010; this issue.
2. Valiant L. A theory of the learnable. *Commun. ACM*, 1984; 27: 1134–1142.
3. Neyman J. Outline of a theory of statistical estimation based on the classical theory of probability. *Philos. Trans. R. Soc. Lond. A, Math. Phys. Eng. Sci.*, 1937; 236: 333–380.
4. Wald A. An extension of Wilks method for setting tolerance limits. *Ann. Math. Stat.*, 1943; 14: 45–55.
5. Tukey JW. Nonparametric estimation II. Statistically equivalent blocks and tolerance regions—The continuous case. *Ann. Math. Stat.*, 1947; 18: 529–539.
6. Campi MC, Calafiore G, Garatti S. Interval predictor models: Identification and reliability. *Automatica*, 2009; 45: 382–392.