Probability in Control?

Jan C. Willems[1]

1 Introduction

Probability is one of the success stories of applied mathematics. It is universally used, from statistical physics to quantum mechanics, from econometrics to financial mathematics, from information theory to control, from psychology and social sciences to medicine. Unfortunately, in many applications of probability, very little attention is paid to the modeling aspect. That is, the interpretation of the probability used in the model is seldom discussed, and it is rarely explained how one comes to the numerical values of the distributions of the random variables used in the model. The aim of this communication is to put forward some remarks related to the use of probability in Systems and Control.

2 Interpretations of probability

One of the main difficulties both in using probabilistic models and in criticizing their use, is that there are widely diverging interpretations of what probability means. Libraries full of books have been written on the topic on interpretation of probability, starting at the time of Pascal and continuing to the present day. See [2] for a comprehensive treatise, and [5] for some remarks and references.

Two main views have emerged, among an uncountable number of intermediate nuances.

1. Probability as a subjective notion, as *degree of belief*.
2. Probability as an objective notion, as *relative frequency*.

The distinction between these two interpretations can be illustrated by considering coin tossing. To the question *'What is the probability of heads?'*, the subjectivist answers $\frac{1}{2}$ because there is no reason to believe that tails are more likely than heads, or vice-versa. The subjectivist does not claim to predict what will happen when the coin is actually flipped. The answer quantifies the person's individual belief. The objectivist on the other hand argues that the probability of heads is $\frac{1}{2}$ because it is claimed to be a physical law that in a repeated experiment with the number of tosses going to ∞, the average number of heads will be $\frac{1}{2}$.

This example is perhaps atypical because it could be argued that the subjectivist argues $\frac{1}{2}$ because he or she believes that the objectivist's relative frequency of a repeated toss will turn out to be $\frac{1}{2}$. For repeatable experiments there is some agreement between both views. But in other situations, the distinction is more striking. If a sports

[1] ESAT/SCD (SISTA), K.U. Leuven, Kasteelpark Arenberg 10, B-3001 Leuven, Belgium.
Jan.Willems@esat.kuleuven.be, http://www.esat.kuleuven.be/~jwillems.

commentator states that the probability of the Dutch team winning the World Cup in South Africa is 0.1, then it is difficult to interpret this statement as anything but the commentator's subjective belief. On the other hand, if an official of the registry of motor vehicles with knowledge of the prices of the automobiles sold in a country last year states that the probability of the price being below P is 0.5, then this is obviously a relative frequency.

For ease of exposition, we left out a third interpretation, namely

3. Probability as *propensity*.

The propensity interpretation, due to Karl Popper, brought a logical foundation to the single-event probability required in the physical probabilistic interpretation of the wave function in Quantum Mechanics.

3 Probability in Systems and Control

At least three main areas of Systems and Control are dominated by probability: filtering, system identification, and stochastic control. It would take us too far to analyze the use of probability in each of these areas. We therefore limit our remarks to filtering.

In continuous time, and over an infinite interval, the filtering problem may be formulated as follows. Given two (vector-valued) signals, $z : \mathbb{R} \to \mathbb{R}^{\mathtt{z}}$, the to-be-estimated signal, and $y : \mathbb{R} \to \mathbb{R}^{\mathtt{y}}$, the observed signal, construct a filter F that takes y into $\hat{z} : \mathbb{R} \to \mathbb{R}^{\mathtt{z}}$, the estimate of z. F is thus a map that takes (a suitable subset of) $\mathbb{R}^{\mathtt{y}}$-valued functions on $\mathbb{R}$ into $\mathbb{R}^{\mathtt{z}}$-valued functions on $\mathbb{R}$. A basic restriction is that F should be non-anticipating, that is $y_1(t) = y_2(t)$ for $t \leq T$ should imply $F(y_1)(t) = F(y_2)(t)$ for $t \leq T$. Filtering is a very well-motivated problem, and numerous applications of it can immediately be seen. But, in order to set up an algorithm to construct F, we must clarify how z and y are related, so that y contains information allowing to obtain a reasonable estimate $\hat{z}$.

3.1 Wiener and Kalman filtering

Wiener masterly solved this problem by assuming that (z, y) is a realization of a stochastic vector process and taking for $\hat{z}(t) = F(y)(t) = \mathscr{E}[z(t)|y(t')$ for $t' \leq t]$. The problem thus became a precise mathematical one, formulas involving spectral factorization were derived for the filter F in the stationary Gaussian case, and a research field was born that is very succesful up to the present day. The Kalman filter addresses in essence the same problem as Wiener did, but, by taking a very convenient representation of the stochastic process (z, y), a filter algorithm was obtained that is far superior and much more easily generalizable.

The problem with the stochastic formulation of the filtering problem is that it requires to model (z, y) probabilistically. Presumably, in the case that we use a frequentist interpretation, this should be done (in the zero-mean Gaussian stationary case) by obtaining the autocorrelation function of (z, y) using statistical methods. But statistics typically assumes a probabilistic framework to begin with, and therefore it assumes the implied existence and persistence of the limits required to define relative frequencies. But where would, for a physical signal, this regularity of (z,y) come from? Surely, there are some applications where (z,y) can indeed be modeled well as a stochastic process, but these are, in my opinion, few and far between. What physical laws ensure that outcomes of a real physical signal is a realization of a stochastic process in the frequentist sense? If, on the other hand, we use a subjective interpretation of probability, then it should be explained where the detailed numerical values of the degree of belief required for the autocorrelation function of (z,y) come from.

3.2 Least Squares filtering

There is an interpretation of the filtering problem that avoids probability. This is most easily explained in the setting of the Kalman filter. Assume that (z,y) is modeled as

$$\frac{d}{dt}x = Ax + Bw, y = Cx + Dw, z = Hx. \tag{1}$$

In words, the relation between z and y stems from the fact that they are both outputs of a linear system that is driven by an input $w : \mathbb{R} \to \mathbb{R}^{\mathtt{w}}$. Think of w as an underlying latent variable that serves to model (z,y).

We could now solve the filtering problem as follows. In order not to get into limit arguments that are not germane to our purposes, assume that the filtering interval is $[0,\infty)$, instead of $\mathbb{R}$ as in the previous section. The signal $z : [0,\infty) \to \mathbb{R}^{\mathtt{z}}$ needs to be estimated from the observations $y : [0,\infty) \to \mathbb{R}^{\mathtt{y}}$, and the estimate $\hat{z}(t)$ can only depend on $y(t')$ for $0 \leq t' \leq t$. The observed trajectory y and the to-be-estimated trajectory z are completely determined by the corresponding input $w : [0,\infty) \to \mathbb{R}^{\mathtt{w}}$ and initial state $x(0)$. Now, in order to compute $\hat{z}(t)$, use the input w and initial state $x(0)$ that minimize

$$J(w,x(0)) = \int_0^t ||w(t')||^2 \, dt' + ||x(0)||_Q \tag{2}$$

over all $(w,x(0))$ that generate the observations $y(t)$ for $0 \leq t' \leq t$. Here $Q \geq 0$ is a suitable weighting matrix.The estimate $\hat{z}(t)$ defined this way obviously depends on the observations $y(t)$ for $0 \leq t' \leq t$ and on the model parameters A,B,C,D,H. It can be shown that this minimization leads exactly to the Kalman filter formulas. This result is more or less obvious from a maximum likelihood interpretation of the Kalman filter, and is derived in the textbook [3] (see also [4], where more references can be found).

This least squares interpretation of the Kalman filter uses the model (1) in an essential way. The use of such representations was new when the Kalman filter was derived. The least squares interpretation of the filtering algorithm readily extends to the Wiener filtering formulas, using a similar representation as (1), or by analogously representing the process (z,y) as an integral of white noise, and then replacing the white noise by an unknown input. But such models were not available when the Wiener filter was first derived, whereas models with (z,y) a stochastic process were very much part of mathematics and probability. The adoption of a stochastic process framework allowed Wiener to pose the filtering problem in precise mathematical terms. This fact, and not the underlying physics, was undoubtedly the main motivation for introducing stochastics in the filtering problem.

Using the least squares interpretation of the Wiener and Kalman filter shifts the burden of justifying the methodology from the *descriptive* to the *prescriptive*. Using the stochastic process interpretation makes claims regarding the model of reality, in other words, stochastic assumptions are part of the physics, of the descriptive part of the problem. On the other hand, using the minimization of 2 specifies the performance. It is up to the designer to choose this as the prescriptive part of the problem. The choice of 2 as the functional to be minimized does not impose anything on the system model.

In a sense, it is possible to interpret the least squares algorithm in terms of a subjective 'degree of belief'. But this interpretation only requires one to state that the $(w, x(0))$ which minimizes 2 is the most believable explanation of the observations. This interpretation is much more parsimonious than the subjective probabilistic interpretation of $(w, x(0))$, which requires giving numerical values of the degree of belief of many more events.

Similar least squares interpretations of many of the algorithms used in system identification and in stochastic control are readily given.There is also the $\mathscr{H}_2$ interpretation of the Wiener and Kalman filter, and its extension to $\mathscr{H}_\infty$ filtering. Also, when the optimal LQG controller is interpreted as minimizing the $\mathscr{H}_2$ norm of the closed loop system, we shift again the burden of the descriptive to the prescriptive, while opening up the generalization to $\mathscr{H}_\infty$ control. These deterministic methods of designing filters and controllers are, in my opinion, very much to be preferred above the stochastic formulations, precisely because they shift the problem justification from the descriptive part of the model to the prescriptive part of the design. Often the argument is put forward that since the stochastic interpretation and the least squares interpretation of filtering and system identification algorithms lead to the same formulas, it is pointless to argue that one interpretation is to be preferred to another one. Of course, this is a two-edged sword, and cannot be used as a defense of the stochastic interpretation. But the fact that the same formulas are obtained does not mean that it is wise or valid to suggest an unverified structure on a model.

In engineering (and prescriptive aspects of economics) one can, it seems to me, also take the following intermediate position as a justification of the use of stochastics. An algorithm-based engineering device, say in signal processing, communication, or control, comes with a set of 'certificates', that guarantee that the device or the algorithm will work well under certain specific circumstances. These circumstances need not be the ones under which the device will be used in practice. They may not even be circumstances which can happen in the real world. These certicates are merely guarantees of good performance under benchmark conditions. Examples of such performance guarantees may be that an error correcting code corrects an encoded message that is received with on the average not more than a certain percentage of errors, or that a filter generates the conditional expectation of an unobserved signal from an observed one under certain prescribed stochastic assumptions, or that a controller ensures robust stability if the plant is in a certain neighborhood of a nominal one.

4 Why resorting to fate can be wise

In [1] it is argued that in many problems in Systems and Control, advantage can be taken of randomization. It is difficult to argue with this. Randomized algorithms are used for secure communication in cryptography, they lead to game theoretic equilibria, they can be effective for evaluating integrals, and so forth. As such many of the points made in [1] are eminently valid.

But it becomes more difficult sometimes to follow the thesis in [1] when it comes to control. For instance, in the discussion of Example 1, it is implied that a design that does not lead to robust stability but that leads to stability with probability 0.1 is to be preferred to a design that does not lead to robust stability, but without being able to attach a value to the probability of stability. But is this so? If we want robust stability then, by the definition of robustness, we should have stability for all values of the uncertain parameters, and randomization has little to offer.

I find it difficult to sympathize with the discussion surrounding what is called 'Bayesian approach'. To begin with, there is difficulty with the semantics here. The term 'Bayesian' always refers to subjective probability, that is, to degree of belief, whereas in [1] Bayesian probability is explained to mean 'how often a system occurs as compared to the other systems'. In other words, the term is used in the sense of relative frequency. Forgetting about this matter of nomenclature, the control problem discussed assumes that we have a situation in which a designer needs to design one single control algorithm for a whole family of plants. The designer knows exactly (from measurements?) the relative frequency of the various plants, but seems to be unable to actually measure the unknown parameters in the actual plant. It would have been nice to see a description of such a situation. I struggled to imagine a convincing engineering example where such a problem would come up. But could it be that once again probability is used here as a panacea for uncertainty, as a way to make the problem into a mathematical one, without going back to the physics?

5 Let us get the physics right

It is my belief that modeling is the most neglected aspect of theoretical engineering in general and, more specifically, of Systems and Control. This is very evident in areas which use probabilistic models. To begin with, the interpretation of probability is seldom explained. This would pose no problem if the interpretation of a particular concept is evident, but in the case of probability with its highly divergent interpretations, this neglect to explain the interpretation is objectionable. Often, it is vaguely implied that a frequentist interpretation is used. But then, why can measurement inaccuracy be modeled as an additive stochastic process? This is perhaps more or less acceptable to model the effect of quantization, but what about all the other sources of measurement uncertainty? Why should an unmeasured nuisance signal in system identification be a stochastic process? Why should a communication channel change a 0 to a 1 and a 1 to a 0 with a fixed relative frequency? Where would this regularity of error generation come from? I do not claim that in many circumstances, these probabilistic methods cannot be rationalized. What I claim is that by and large, we are doing research and teaching without bothering to explain the physics that leads to probabilistic models.

The neglect of the physics in Systems and Control, is much more widespread than for probability alone. As I have argued extensively before [6], it applies to the input/output thinking that is universally used for modeling open systems. Physical systems are not signal processors. The methods based on inputs and outputs are especially awkward for system interconnection. Interconnection of physical systems leads to variable sharing, not to output-to-input assignment. This is evident in simple systems as electrical circuits and mechanical devices. There is no reason why this situation should suddenly be different for complex systems, as those found in biology. Signal flow graphs have their place, but as a description of the functioning of an interconnected physical system, they miss the crucial point of expressing what interconnection entails.

How can such a situation have occurred? Why is the physics of models not more prominently present in areas as Systems and Control? Why are probability, inputs, outputs, and signal flow graphs used without analyzing the physical situations to which they claim to pertain? The explanation, in my opinion, lies in the sociology of science. Normal science uses an established paradigm in which to operate. When a problem is cast in an input/output setting with disturbances modeled as stochastic processes, we are operating in a clear and often sophisticated mathematical framework, with results that may be difficult to obtain and to prove, and that are verifiable mathematically. The results are judged by their mathematical depth and difficulty. In other words, the explanation lies in the *Lure of Mathematics*. There is no other explanation.

References

[1] M.C. Campi, Why resorting to fate is wise? A critical look at randomized algorithms in systems and control, European Journal on Control, volume , pages – , 2010.

[2] E.T. Jayes, *Probability Theory, The Logic of Science*, Cambridge University Press, 2003.

[3] E.D. Sontag, *Mathematical Control Theory, Deterministic Finite Dimensional Systems*, Springer Verlag, 1990.

[4] J.C. Willems, Deterministic least squares filtering, *Journal of Econometrics*, volume 118, pages 341–370, 2004.

[5] J.C. Willems, Thoughts on system identification, *Control of Uncertain Systems: Modelling, Approximation, and Design*, Springer Lecture Notes in Control and Information Sciences, volume 329, pages 389–416, 2006.

[6] J.C. Willems, The behavioral approach to open and interconnected systems, *Control Systems Magazine*, volume 27, pages 46–99, 2007.