COMMENTS ON *A NATURAL GENERALIZATION OF JEFFREY CONDITIONING*

Suppose that you have a probability measure p_Θ defined on the measurable space $(\Theta, \wp(\Theta))$. Jeffrey (1965)'s update rule, *Jeffrey Conditioning*, provides a mapping from probability functions on a measurable space and inputs to other probability functions on the same measurable space. The inputs to Jeffrey conditionalization are partitions $\mathcal{E}_\Theta$ of Θ and a collection of positive real numbers $\{\mu_E\}_{E \in \mathcal{E}_\Theta}$ which sum to 1.

Wagner (1992) was interested in where these inputs come from. He suggested that, in some cases, they come from *another* probability measure, p_Ω, on *another* measurable space $(\Omega, \wp(\Omega))$, together with a mapping, Γ, from the set Ω to the set $\wp(\Theta)$—the mapping which tells us, intuitively, the strongest proposition in $\wp(\Theta)$ which follows from the truth of the proposition $\{\omega\}$, for every $\omega \in \Omega$.

Thus, on Wagner's approach, the inputs to a probabilistic transition on the possibility space $(\Theta, \wp(\Theta), p_\Theta)$ are 1) an additional probability space $(\Omega, \wp(\Omega), p_\Omega)$; and 2) a mapping Γ from worlds in Ω to sets of worlds in Θ.

$$\begin{tikzcd}
(\Omega, \wp(\Omega), p_\Omega) \arrow[dr] & \Gamma : \Omega \to \wp(\Theta) \arrow[d] & (\Theta, \wp(\Theta), p_\Theta) \arrow[dl] \\
& (\Theta, \wp(\Theta), p'_\Theta) &
\end{tikzcd}$$

While, in some cases, these inputs determine the inputs the Jeffrey's rule, in other cases, they do not. Thus, Wagner provides a rule which 'generalizes' Jeffrey conditioning. Wagner's rule is that, for each $A \in \wp(\Theta)$, your posterior probability that A, $p'_\Theta(A)$, ought to be as specified in (W).

(W) $$p'_\Theta(A) = \sum_{B \in \wp(\Theta)} p_\Omega(\Gamma^{-1}(B)) \cdot p_\Theta(A \mid B)$$

(Where $\Gamma^{-1}(B)$ is the set of $\omega \in \Omega$ which Γ maps to B.) Wagner argues that (W) reduces to Jeffrey's update rule in a special case. He additionally claims, in a short paragraph in the conclusion of the article, that (W) does not agree with Jaynes' 'Principle of Maximum Entropy'; in particular, it does not agree with the recommendation of minimizing the cross-entropy between p_Θ and p'_Θ, subject to the constraint that $p'_\Theta(A)$ be bounded from below by $\sum_{B\subseteq A} p_\Omega(\Gamma^{-1}(B))$ [which Wagner takes to be the only constraint which a defender of the Principle of Maximum Entropy should endorse].

The author contends:

a) (W) does agree with the Principle of Maximum Entropy; the two are consistent given the assumption that probabilities must be precise.

b) The Principle of Maximum Entropy generalizes (W).

It's unclear to me why the precision of probabilities is relevant here (see point (4) below). However, the author accepts (b) for the following reason: in the course of deriving his update rule (W), Wagner introduced a hypothetical probability distribution $p_{\Omega\times\Theta}$ over the product space $\Omega\times\Theta$. He then showed that, while there are infinitely many joint distributions $p_{\Omega\times\Theta}$ which have various nice properties which Wagner felt it appropriate to impose, all of these share a marginal distribution $\sum_{\omega_j\in\Omega} p_{\Omega\times\Theta}(\,\cdot\,,\omega_j)$, and this marginal is $p'_\Theta(\,\cdot\,)$ as defined in (W) above.

The author contends, however, that rather than calculating a posterior by choosing the one which minimizes cross-entropy subject to some constraint, the defender of the Principle of Maximum Entropy should first choose a prior, $p_{\Omega\times\Theta}$, over the product space $\Omega\times\Theta$ which maximizes entropy, and then update that prior by choosing the one which meets certain constraints (constraints corresponding to the ones imposed by Wagner (1992)). The author contends that such an update will always agree with Wagner's prescription (W) (though this is not shown here, but rather claimed to follow in the technical companion paper).

Most Important Points.

(1) This is rather technical for a paper which is meant for a general philosophical audience (and which has its own technical companion paper). The eponymous generalization is presented very briefly on page 12 of the article, in terms that require consulting reference texts which I suspect are beyond the reach of much of the journal's readership.

(2) Relatedly, while the title of the paper makes it appear as though the focus will be on providing a generalization of Jeffrey conditioning, much of the focus of the paper is taken up with criticizing Wagner (1992) for a stray comment in his concluding paragraphs. My understanding is that neither Wagner's 1992 article nor its criticism of the Principle of Maximum Entropy have been so influential

that the journal's readership will be interested in whether Wagner's comments were accurate.

(3) Much of the paper's criticism of Wagner seemed to focus on the fact that he endorsed imprecise probabilities, simply because he failed to specify the joint distribution $p_{\Omega\times\Theta}$. In the first place, failure to specify a distribution is not the same as specifying an imprecise probability distribution (see point 4). In the second place, Wagner was clear that he did not intend for the hypothetical joint distribution to be taken seriously. He writes:

> It is to be emphasized that probability measures on $\Omega \times \Theta$ enter the preceding discussion as formal, conceptual tools, there being no presupposition... that a fully defined probability measure on $\Omega \times \Theta$ be attainable. Indeed, formula [(W)] involves only uncertainty measures on Θ, the prior [p_Θ], and the measure...induced by [p_Ω] and Γ...[1]

If the paper is going to be focused on criticizing Wagner, then I believe that this paragraph should be discussed at some point (though I'd probably tollens on this one).

(4) I don't understand what role the precision of probabilities had to play in the article. It seems as though the right thing to say is this: Wagner thought that a defender of the Principle of Maximum Entropy would want to update by minimizing the cross-entropy between p_Θ and p'_Θ, subject to the constraint that $p'_\Theta(A)$ be bounded from below by $\sum_{B\subseteq A} p_\Omega(\Gamma^{-1}(B))$. The author thinks otherwise. The author thinks that the right way to respond to these kinds of learning scenarios is by populating the product space $\Omega \times \Theta$ with MAXENT, and updating with *Infomin* on Wagner's constraints. I don't see what role precision or imprecision has to do with this disagreement.

(5) I was confused about why a companion technical paper was required in order to prove that updating the joint distribution $p_{\Omega\times\Theta}$ with *Infomin* on Wagner's constraints yields a marginal distribution over Θ which agreed with (W). Doesn't this follow trivially from Wagner's result that *any* joint distribution satisfying his constraints has a marginal which agrees with (W)? What more is there to prove?

Less Important Points.

(6) I believe that the quotation from Caticha and Giffin (2006) is misleading. The placement of the quotation makes it appear that Caticha and Giffin are claiming that $P(\omega, \theta)$ must be defined *in the kinds of cases Wagner (1992) is theorizing about*. This is not correct. Rather, Caticha and Giffin are simply claiming that, if you update by Bayesian conditionaliation, then, in order to make an inference about the values of the parameters $\theta_i \in \Theta$ on the basis of the values of some variables

[1] Wagner 1992, p. 251, with notational changes

$x_i \in \mathcal{X}$, your prior probability function must be defined over $\Theta \times \mathcal{X}$. Otherwise, the probability $P(\theta_i \mid x_i)$ will not be defined, so you cannot update *via* conditionaliation. This point does not appear to have any application to the kinds of learning episodes being discussed in the article.

Least Important Points.

(7) I found the diagram on page 2 incredibly confusing. It is unclear what the black dots represent as well as what the '×' and '✓' symbols are supposed to represent, and what the remarks in the final column are directed toward (neither of the final two columns are labeled).

References

Caticha, Ariel & Adom Giffin. 2006. "Updating Probabilities." In *MaxEnt 2006, the 26th International Workshop on Bayesian Inference and Maximum Entropy Methods*. Paris, France.

Jeffrey, Richard. 1965. *The Logic of Decision*. New York: McGraw-Hill.

Wagner, Carl. 1992. "Generalized Probability Kinematics." *Erkenntnis* 36 : 245–257.