*Letter to the Editor*

# "Combining Probability Distributions from Experts in Risk Analysis"

In their well-researched and almost unbelievably comprehensive paper, Clemen and Winkler[1] give a summary of my "Expert Information" method originally published in 1992.[2] From this summary, and from rereading the original paper, I recognize that in that paper I failed to adequately communicate my idea, probably because the idea at that time was newly formed in my mind. From where I am now, I see that the paper glossed over and left fuzzy a part of the process that should be very sharp and clear.

I would like to remedy that defect in the following.

With regard to the difficult question: "How should we combine probability distributions from different experts?" my suggestion is that we bypass it. We can bypass it by not asking the experts for their probability distributions. Instead, we ask each expert independently what evidence, information, and experience he/she has, relevant to the question at hand. We write these items of evidence down very carefully, and collect them in a combined list. Then, using the experts as a group, we work over these items, clarifying, coalescing, refining, adding new items that come to mind, and all the time being sure to distinguish the actual evidence (i.e., "what happened") from the experts' interpretation of what happened. At the end of this process we should have a single, agreed upon "consensus body of evidence."

Then, together with the group, we process this combined body of evidence, item by item, through Bayes' theorem to arrive finally at a posterior probability curve that expresses our joint state of knowledge about the parameter of interest.

For example, suppose the parameter $\lambda$ denotes the failure fraction, on demand, of a piece of safety equipment. By its definition the numerical value of this parameter must lie in the interval [0, 1]. Let us suppose now that we have made explicit, and compiled into the consensus list, all the evidence items we have relevant to the numerical value of this parameter. If we have truly included in this list *all* the relevant evidence, then our probability distribution "prior to" this evidence is a flat curve on the interval [0, 1]. Let us denote this distribution by $p_0(\lambda)$.

Now take the first item, $E_1$, on the evidence list and process it through Bayes' theorem. The result is the probability curve $p(\lambda \mid E_1)$, which we could also call $p_1(\lambda)$. We next do the same with the second item, $E_2$, using $p_1(\lambda)$ as the prior in Bayes theorem. The resulting posterior is now $p_2(\lambda) = p(\lambda \mid E_1, E_2)$.

Continuing this way we obtain, at the end of the list, $p_N(\lambda)$, where $N$ is the number of evidence items on the list. $p_N(\lambda)$ is now our final probability distribution. It represents our collective "state of knowledge" about the numerical value of $\lambda$ based on all the evidence we have. I like to call it our "credibility distribution." It tells us, for any number, $x$, in the interval [0, 1], how much credibility can be logically assigned to the hypothesis that $x$ is the correct value for $\lambda$.

This process is what I call the "Expert Information" process, as opposed to the "Expert Opinion" process. It attempts to "Let the Evidence Speak" rather than the personalities, positions, politics, reputations, opinions, wishful thinkings, etc.

This expert information process looks really neat, and in principle it is. But now I have to own up to the fact that doing it is not so easy. It is often quite demanding philosophically and tactically. Defining the problem right in the first place can be difficult, getting the consensus body of evidence can require considerable effort, and using Bayes' theorem can be quite challenging.

Within Bayes' theorem is the term $p(E \mid \lambda)$, known as the "likelihood function." This is the crucial term, in fact, in the theorem. Evaluating it is easy or hard depending on the type of evidence item.

For certain types of evidence, such as the result of a pure, random sampling experiment, evaluating the likelihood is easy. I call such types of evidence "direct" evidence. Unfortunately, most of the evidence we have to deal with is "indirect" to some degree. Developing the likelihood function for such types of evidence can require considerable insight and ingenuity, together with a bag of tricks. I discuss this point in an article in a previous issue of this journal[3] and give some examples elsewhere.[2,4]

We need make no apologies for these difficulties, however. Difficult or easy, Bayes' theorem is the fundamental principle for making inferences from evidence,

and therefore cannot be avoided. It must be used, if not explicitly, then implicitly. Better to do it explicitly.

## REFERENCES

1. R. T. Clemen and R. L. Winkler, "Combining Probability Distributions from Experts in Risk Analysis," *Risk Anal.* **19**, 187–203 (1999).
2. S. Kaplan, "'Expert Information' versus 'Expert Opinion.' Another Approach to the Problem of Eliciting/Combining/Using Expert Knowledge in PRA," *Reli. Engin. Syst. Safety* **35**, 61–72 (1992).
3. F. O. Hoffman and S. Kaplan, "Beyond the Domain of Direct Observation: How to Specify a Probability Distribution that Represents the 'State of Knowledge' About Uncertain Inputs," *Risk Anal.* **19**, 131–134 (1999).
4. S. Kaplan, "On the Inclusion of Precursor and Near Miss Events in Quantitative Risk Assessments: A Bayesian Point of View and a Space Shuttle Example," *Reli. Engin. Syst. Safety* **27**, 103–115 (1990).

Stan Kaplan
*Bayesian Systems, Inc.*
*Gaithersburg, MD*