Sensitivity Analysis via Information Density[1]

By: Gordon B. Hazen (Northwestern University)

The practice of parametric sensitivity analysis is crucial to any responsible decision model implementation. One of its primary purposes is the demonstration to supportive or skeptical audiences of the robustness of an analysis to variations in input parameters, be they probabilities, rates, costs, utilities, resource levels, or other quantities. Important instruments here are, for example, traditional one- and two-way graphical sensitivity analyses, as well as tornado diagrams (e.g., Clemen 1996).

Key in recent literature is the recognition that for a decision model, input parameter estimates are just that – estimates – that reflect underlying parameter uncertainty. When this uncertainty can be probabilistically quantified in terms of parameter distributions, measures of *global* sensitivity have been devised that take into account not just variation in one or two parameters at a time, but all model parameters simultaneously. These include measures based on variance (e.g., Wagner 1995), as well as the probability that the recommended optimal choice might be incorrect (so-called *probabilistic sensitivity analysis*, e.g., Doubilet *et al.* 1985, Critchfield and Willard 1986, Briggs *et al.* 2006), and the expected value of perfect information. All of these measures may be computed for the entire parameter set, or any subset of interest. However, I think information value is the preferred measure of sensitivity (Felli and Hazen 1998), as it takes into account not only the chance that the optimal choice might be incorrect, but also the impact of that potential mistake.

In spite of their limited ability to capture sensitivity to all input parameters simultaneously, traditional one- and two-way graphical sensitivity analyses retain a certain appeal. Of course, they are easily computed and do not require explicit parameter distributions, but beyond these reasons, their appeal also arises from their ability to depict *critical directions* of sensitivity. A one-way sensitivity analysis can tell us, for example, that *policy A* will no longer be optimal should an estimated probability *p* fall below a critical threshold (note: below is a *direction*). In contrast, the expected value of perfect information on *p* might be reported as $5000. This figure provides no indication that variation in *p* in the *downward* direction (in this case) is the crucial concern, a concern that a graphical one-way sensitivity analysis reveals immediately. Of course, in the information value computation itself, the values of *p* that fall below the threshold play an important role – a role that is obscured by the final reported $5000 information value. Although this difficulty is harmless in a single-parameter setting where graphical aids are easily available, their absence in the multi-parameter setting leaves a simple information value report lacking in useful meaning for sensitivity analysis needs.

The purpose of this note is to point out that by augmenting an information value report to include what I term the *information density*, an analyst can provide clear information about directions of concern for sensitivity – the same information so intuitively captured in a graphical sensitivity diagram. The advantage is that this directional information can be made available not just for one- or two-way sensitivity analyses, but for sensitivity/ information value calculations on an arbitrary number of parameters. These ideas are so far only at a nascent stage, and I here content myself to provide only simple numerical and graphical examples for the enjoyment of potential further researchers.

[1] *Decision Analysis Today* Vol. 33 (2014), No.1, pp 24-29.

Information Density

Consider a situation in which we wish to find the information value of a continuous-valued uncertainty ξ with density $f(\xi)$. Quite possibly ξ could be a vector $(\xi_1,\ldots,\xi_q)$. It is well known that the information value $EVPI_\xi$ associated with learning the true value of ξ is the *expected improvement* in payoff obtainable by optimizing the action a taken after learning ξ:

$$EVPI_\xi = \mathrm{E}[\max_a \mathrm{E}[\Delta V_a|\xi]] = \int_\xi f(\xi)d\xi \max_a E[\Delta V_a \mid \xi] \qquad (1)$$

Here $\Delta V_a = V_a - V_{a^*}$ is the improvement in payoff over the optimal payoff V_{a^*}, V_a is the payoff associated with action a, and we are assuming risk neutrality in the computation of information value[2].

Let $d\xi$ denote an infinitesimal region around a particular value of ξ, and let $[d\xi]$ be the indicator variable of $d\xi$ (equal to 1 if $\xi \in d\xi$ and zero otherwise). What is the information value of $[d\xi]$, that is, the information value of learning whether $\xi \in d\xi$? We have

$$EVPI_{[d\xi]} = \mathrm{E}[\max_a \mathrm{E}[\Delta V_a|[d\xi]]]$$

$$= f(\xi)d\xi\cdot\max_a \mathrm{E}[\Delta V_a|\xi] + (1 - f(\xi)d\xi)\cdot\max_a \mathrm{E}[\Delta V_a|\xi\notin d\xi].$$

Because of the infinitesimal nature of the region $d\xi$, the conditioning event $\xi\notin d\xi$ is equivalent to the sure event, so that

$$\max_a \mathrm{E}[\Delta V_a|\xi\notin d\xi] = \max_a \mathrm{E}[\Delta V_a] = 0.$$

In effect, learning $\xi \notin d\xi$ tells us nothing, so has zero incremental value. So we conclude

$$EVPI_{[d\xi]} = f(\xi)d\xi\cdot\max_a \mathrm{E}[\Delta V_a|\xi]. \qquad (2)$$

On comparing (2) with (1), we see that

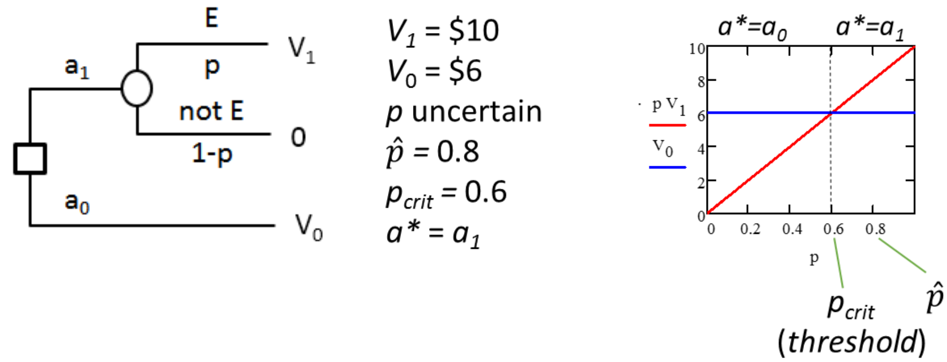
$$EVPI_\xi = \int_\xi EVPI_{[d\xi]} .$$

Based on this equality, we are justified in calling the function $i(\xi) = f(\xi)\cdot\max_a \mathrm{E}[\Delta V_a|\xi]$ in (2) the *information density* for ξ. It qualifies as a density because it is nonnegative ($\max_a \mathrm{E}[\Delta V_a|\xi] \geq \mathrm{E}[\Delta V_{a^*}|\xi] = 0$), and its integral is the overall information value. It is an *information* density because $EVPI_{[d\xi]} = i(\xi)d\xi$. It is of interest to note that the information density is zero in the region of ξ-values that would not change the optimal decision.

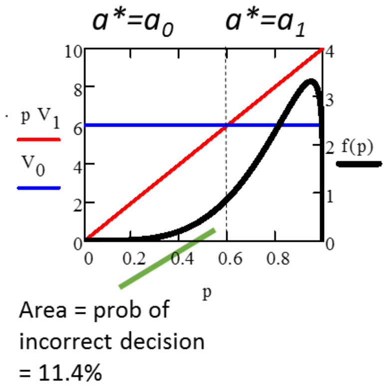
Information Density – Examples

To recapitulate the notions in my introductory remarks in a concrete setting, consider the following simple toy problem

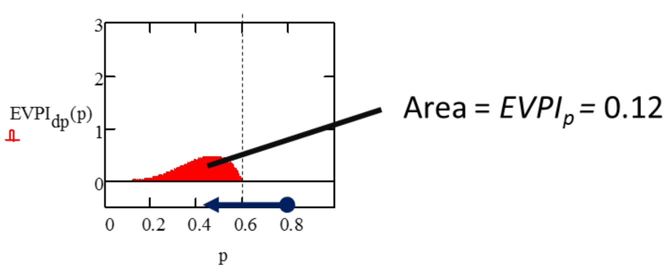
[2] However, information value may be formulated in the same "simple" way under expected utility. See Hazen and Sounderpandian (1999).



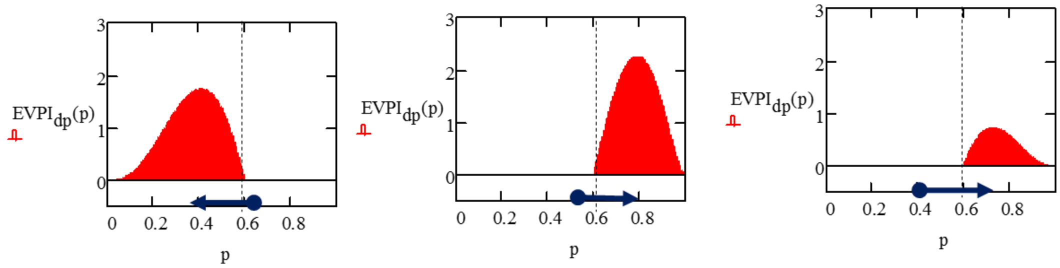
in which the only uncertain parameter ξ is the probability p of the event E, whose baseline estimate is $\hat{p}$ = 0.8. A one-way sensitivity analysis on p reveals a critical value p_{crit} = 0.6 below which the optimal action changes from the baseline a_1 to a_0. Suppose that p has the beta(a=4.8, b=1.2) distribution below with mean $\hat{p}$:



Probabilistic sensitivity analysis reveals an 11.4% chance that the optimal choice might be incorrect. However, the information value of p is \$0.12, relatively small compared to the optimal expected payoff of \$8, indicating, as is common (see Felli and Hazen 1999) that sensitivity should not be a concern in spite of this 11.4% figure. Unfortunately, the value $EVPI_p$ = \$0.12 by itself reveals nothing about the *direction of concern* for p, which is that p might fall below the critical value p_{crit} = 0.6, a fact that is obvious from the one-way sensitivity graph. So *EVPI* by itself seems inadequate for sensitivity concerns. If, however, one graphs the information density for p, one obtains:



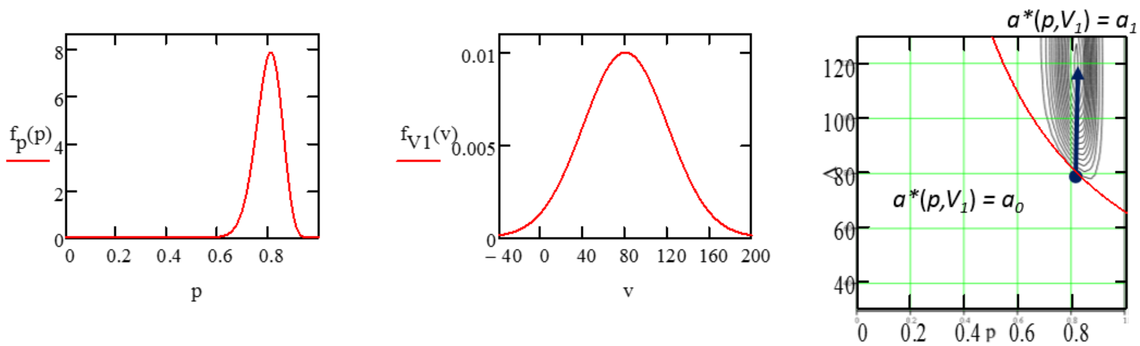
clearly indicating that the direction of concern for p (depicted here by an arrow from the baseline estimate of p to the mode of the information density) is *down*. Other possible information densities for p, as the baseline estimate $\hat{p}$ varies (keeping the standard deviation of p the same), are as follows:



$\hat{p}$ = 0.65 $EVPI_p$ = \$0.53 | $\hat{p}$ = 0.55 $EVPI_p$ = \$0.55 | $\hat{p}$ = 0.40 $EVPI_p$ = \$0.15

Note that based on the information density, the direction of concern for p switches from down to up when the baseline estimate $\hat{p}$ moves below the critical threshold p_{crit} = 0.6 where the optimal action switches from a_1 to a_0. Of course, this would be obvious anyway from a corresponding one-way sensitivity analysis graph. However, I ask the reader to imagine the multidimensional parameter setting, where graphical sensitivity analysis is infeasible, but an information density may still be calculated.

Moving towards that setting, suppose the uncertain parameter set in our example is expanded to ξ = (p,V_1), with baseline estimates $(\hat{p},\hat{V}_1)=(0.80,\$80)$ (and V_0 = \$65 with certainty). The situation becomes as follows:

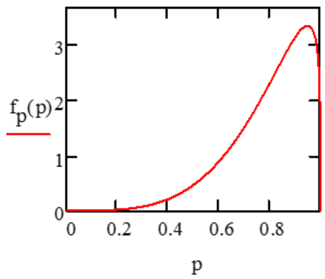


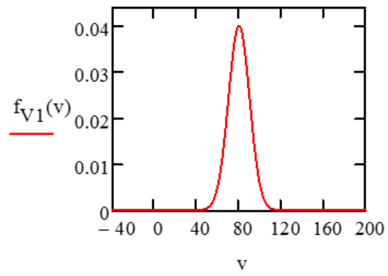
Here we take p,V_1 to be independent with densities having means $(\hat{p},\hat{V}_1)=(0.80,\$80)$ in the leftmost two panels above (p is beta with $a+b$ = 60, and V_1 is normal with standard deviation \$40). In the right panel, the threshold curve between $a^* = a_1$ and $a^* = a_0$ is given in red, and the baseline optimal action is $a^* = a_0$ with payoff \$65. The joint information value for (p,V_1) is $EVPI_{p,V_1}$= \$12.4, which is relatively large compared to the problem payoffs – in my judgment there is sensitivity here. The information density is revealed by the contour lines.

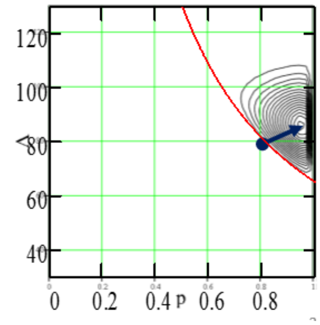
The mode of the information density is (0.82, \$120). Knowing whether (p,V_1) is close to this point would give the greatest gain in information value. The direction $(\Delta p,\Delta V_1)$ of concern for (p,V_1) is therefore indicated by the arrow from the baseline $(\hat{p},\hat{V}_1)=(0.80,\$80)$ to the mode of the information density, yielding $(\Delta p,\Delta V_1)$ = (0.02, \$40). This value reveals that for both p and V_1, the direction of concern is *up*, but it also seems to reveal that V_1 plays a stronger role than p in information value.

In the absence of a picture (as in the multidimensional case), one might reason that Δp = 0.02 is only 13% of the 3σ range for p, whereas ΔV = \$40 is 33% of the 3σ range for V, so that variation in V plays a stronger role in the joint information value.

However, below is a situation in which the reverse is the case:







The baseline estimates / means remain as $(\hat{p},\hat{V}_1)=(0.80,\$80)$, but here p has a large standard deviation ($a+b$ = 6) and V_1 a small one (σ = \$10). Information value is $EVPI_{p,V_1}$= \$5.3, not so large as before. Knowing whether (p,V_1) is close to the modal value (0.98, \$85) of the information density would give the greatest gain in information value.

The direction of concern (Δp,ΔV_1) = (0.18, \$5) is *up* for both p and V_1, and reveals a larger value-of-information role for p and a smaller one for V_1 than before. In the absence of a diagram, the reasoning would be that Δp = 0.18 is 39% of the 3σ range for p, whereas ΔV_1 = \$5 is only 18% of the 3σ range for V_1, so p would seem to play a more important role in the joint information value. These numbers merely indicate *relative* contribution to sensitivity – whether there is sensitivity or not depends on the information value \$5.3, revealing that one could increase the optimal expected payoff from \$65 to \$70.3 by learning p and V_1. Perhaps this indicates moderate sensitivity – the degree of sensitivity is a subjective judgment.

The Multidimensional Setting – Open Issues

Assuming the notion of information density is an attractive one for sensitivity analysis, how should analysts proceed in the multidimensional parameter setting ξ = (ξ_1,…,ξ_q) where no graphical aids are available? One may still compute the mode of the information density, and from there, if desired, the direction of concern. One should remember, however, that the magnitude of this direction will not be related to the magnitude of the information value. Presumably the relative magnitudes of its components relate to their degree of contribution to information value, but how exactly?

The overarching question, however, is whether these quantities are the “right” way to think about the information density. Perhaps some other measure of central tendency of the information density, such as a center of gravity, would be more appropriate than modal value? However, what if the information density is bimodal (corresponding perhaps to multiple threshold surfaces for multiple alternatives a)? In this case, center of gravity would be misleading –there would really be multiple directions of concern, so multiple local modes should be computed. But are simple direction(s) of concern the right way to summarize the information density?

Returning to the definition of information density itself, how this should be extended to discrete-valued parameters (where the second term in $EVPI_{[d\xi]}$ does not zero out) is less than obvious.

Questions such as these need to be addressed before this procedure can become operational. I leave these issues to interested readers.

For additional information contact: Gordon B. Hazen (gbh305@northwestern.edu)

References

A. Briggs, M. Sculpher, K. Claxton, 2006, *Decision Modeling for Health Economic Evaluation*, Oxford University Press.

R. T. Clemen, 1996, *Making Hard Decisions: An Introduction to Decision Analysis*, 2nd ed, Duxbury Publishing.

G. G. Critchfield, K. E. Willard, 1986, Probabilistic analysis of decision trees using Monte Carlo simulation, *Medical Decision Making*, **6**, 85-92.

P. Doubilet, C. B. Begg, M. C. Weinstein, P. Braun, B. J. McNeil, 1985, Probabilistic sensitivity analysis using Monte Carlo simulation, *Medical Decision Making*, **5**, 157-177.

J. C. Felli, G. B. Hazen, 1998, Sensitivity analysis and the expected value of perfect information, *Medical Decision Making*, **18,** 95-109. (Errata in volume **21** (2001) 3, p. 254, and volume **23** (2003) 1, p. 97.)

J. C. Felli, G. B. Hazen, 1999, Do Sensitivity analyses really capture problem sensitivity? An Empirical Analysis Based on Information Value, *Risk, Decision and Policy*, **4**, 79-98.

G. B. Hazen, J. Sounderpandian, 1999, Expected-utility preference reversals in information acquisition, *Journal of Risk and Uncertainty*, **18**, 125-136.

H. M. Wagner, 1995, Global sensitivity analysis, *Operations Research*, **43** (6), 948-969.