Seminar Paper

Quality of information sources and value of information

Management Decision Making | Tandogan (0048035), Lindinger (0508782)
LV-Nr. 040.541/1, Winter Term 2006/07, O.Univ.-Prof. Mag. Dr. Rudolf Vetschera

universität wien

Quality of information sources and value of information

Tamer Tandogan (0048035), Andreas Lindinger (0508782)

5 March 2007

Contents

1 Introduction

Decision making is an almost continuous activity in the everyday business of living our lives. Uncertainty is inherent in the decision making process. In decision making under uncertainity, a decision maker does not know precisely the outcome of actions chosen because of the occurrence of events that are not under his control, events that must be viewed as a random variable. Although the specific realization of the random variable is generally known prior to the choice of action, in decision theory the decision maker identifies several possibilities and classifies them into a set of mutually exclusive state descriptions.

Knowledge is required to control or to take an action, therfore decision making presents a need for information. Decision makers (DMs) who face uncertain prospects have to gather information. The intuitive reason behind gathering information is straightforward: DMs can reduce uncertainty about anticipated outcomes, in other words, they may improve current decisions with a set of information to reach better outcomes.

In general, the lack of information is the nature of the problem. Chavas and Pope (1984) define information as a message which alters probabilistic perceptions of random events. For the measure of information, as a substitute for certainty, probability enters into the process. Information is probabilistic if it expresses a probablity distribution over some exhaustive collection of events. Probability assessment of uncontrollable events quantifies the information gap between what is known and what needs to be known for an optimal decision. In this sense, information has a value, leading to a reduction in the expected expense relative to the situation in which the information is not available.

If there is value of information, there should be determinants of the value of information. What makes information useful is the fitness for use of it. Information quaility becomes a crucial factor for the effectiveness of decision makers. There are numerous disciplines, including management, medicine, cartography, linguistics, rhetorics, history, and athmospheric sciences which have discussed the notion of information quality. The definition of information quality in these disciplines depends on the use of information.

The relationship between information and knowledge has a distinctive role for quality definitions. Information becomes knowledge when it is correctly interpreted and connected with prior knowledge. High-quality information makes it easier to transform information into knowledge, by helping to interpret and evaluate information, by assisting the connection to prior knowledge, and by facilitating the application of information to new contexts.

Lesca and Lesca (1995) analyzed information quality problems. They summarized eight information quality problems: limited usefulness of information, ambiguity, incompleteness, inconsistency, an inadequate presentation format, unreliable, unaccessability. Strong et al. (1997) developed another systematic approach to information quality problems. In their approach, problems considered were mutliple sources, subjective production, production errors, too much information, distributed systems, changing task needs, security and privacy requirements, and lack of computing resources.

If we consider each topic seperately in this paper, value of information and quality of information become broader topics to examine. In fact, being precise, value and quality descriptions are limited for a specific model, a cost-loss ratio situation used in weather forecasts. Therefor, an analytic model by Katz and Murphy (1987) is used. Some theoretical results concerning the nature of the relationship between the scientific quality and economic value of imperfect forecasts were obtained by using this prototype decision making model. The effort shown later in the paper is focused on examining the conditions in which the economic value of information changes as its quality increases.

2 Cost-Loss Ratio

Efforts to quantify the uncertainty in weather forecasts have quite a long history, and many studies and experiments involving objective and subjective probability forecasting have been conducted in meteorology in the intervening period. Interest in probability forecasting in fields other than meteorology has increased considerably in recent years. For example, probability forecasts are used in portfolio analysis by some large banks, formal recognition of uncertainty in terms of probabilities is becoming more common in economic forecasting, and probabilistic diagnosis is receiving increasing attention in the medical field.(Murphy and Winkler, 1984)

Probabilities represent an important input in applications of decision analysis, and the promising area called risk analysis depends heavily on probability forecasts, especially for low-probability events such as an earthquake or a serious accident at a nuclear power plant. Weather information in general and weather forecasts in particular have important roles in decision making models, which describe how the forecast should be used by DMs and what will be the value of the forecasts. A Cost-loss ratio situation is a decision making process often used in the meteorological literature.

Thompson (1952) formulated the original cost-loss ratio situation. It is a very simple normative model that describes the situations faced by many forecast-sensitive decision makers. As a result, the model has been widely used in both real and hypothetical decision making situations. This single stage model involves a decision maker who must decide whether or not to take protective action, with respect to some activity or operation, in the face of uncertainty as to whether or not weather adverse to the activity will occur.

In the next section, the model will be presented extensively.

3 Model

The random variable Θ represents two possible states $\Theta = 1$ and $\Theta = 0$. In the face of uncertainty about whether the adverse weather event S_1 will occur, the decision maker must decide whether to take protective action. If protective action is taken, then a cost C is incurred. If protective action is not taken but the event does occur, then a loss L is incurred ($0 < C < L$ for protective action to be admissible). The decision maker adopts the criterion of selecting the action (i.e. protect or do not protect) that minimizes the expected expense (in general, equivalent to maximizing expected return or profit).(Katz and Murphy, 1987)

An example will be helpful to simplify the model. The example chosen for the cost-loss ratio model is the umbrella problem. Cost-loss ratio models are common in atmospheric studies, therefore the most famous setup is taking protection as an action for the state of weather conditions. The 2x2 contingency table in Figure 1 shows the cost analysis for the possible actions for two different states. If protection is taken, cost C occurs no matter what the state is. On the other hand, if decision makers do not choose protection then in case of adverse weather a loss L occurs.

		States	
		S_1 Θ=1 (Rain)	S_2 Θ=0 (No rain)
Actions	a_1 (take umbrella)	C	C
	a_2 (do not take umbrella)	L	0

Figure 1: 2x2 Contingency table: Actions, States

The information about state (weather condition) is a single prior probability of adverse weather condition (rain), based on historical observations, from imperfect information (weather forecasts) specifiying a conditional probalility of rain that varies from occasion to occasion, to perfect information specifying (with probability one) that adverse weather will or will not occur on a given occasion.

Climatological (or prior) information consists of a single probability of rain, $\pi = Pr\Theta = 1$, that is available to the decision maker each day. Imperfect information about Θ is assumed to consist of the random variable Z, indicating a forecast of rain ($Z = 1$) or of no rain ($Z = 0$). Figure 2 shows the possible types of information.

Prior information (a single probability of adverse weather)	π = Pr[Θ=1],	0<Θ<1
Imperfect information Forcast of wheather condition, Z	Adverse weather	1
	No adverse weather	0

Figure 2: Information types

The conditional (or posterior) probabilities of rain are denoted by

$$p_1 = Pr\left[\Theta = 1|Z = 1\right], p_0 = Pr\left[\Theta = 1|Z = 0\right] \quad (1)$$

It is assumed that the probabilities satisfy the ordering $0 \leq p_0 \leq \pi \leq p_1 \leq 1$.

3.1 Value of information

The most comprehensive normative structure for information evaluation is the theory of information economics. This theory is based on economic and statistical decision theory and models the selection of optimal information in a cost/benefit sense, by explicitly recognizing the decision maker's beliefs and preferences. A major concept in information economics is the value of information. This value is equivalent to the value of the information provided by the system and is the precise analogue for the more casual term "information usefulness".(Lawrence, 1999)

In decision theory the demand price of information is defined as the maximum non-stochastic cost, payable from initial wealth prior to the receipt of any message, that makes DMs indifferent between purchasing the information or not. For the cost-loss decision-making model, one way to measure the economic value of forecast probability estimators is in terms of the reduction in expected expense incurred by the decision maker. This measure involves a comparison of the expected expense using an estimator with that if climatological information alone were available to the decision maker.

Information has a value sofar if it changes the decision maker's action, leading to a reduction in expected expense relative to the situation in which the information is not available. Both the value of perfect and imperfect information can be measured if the decision maker has the information. Let E_C, E_F and E_P denote the expenses associated with climatological information, imperfect information and perfect information, respectively. The economic values of perfect V_P and imperfect information V_F, are given by

$$V_F = E_C - E_F,\ V_P = E_C - E_P \tag{2}$$

The value of imperfect information satisfies $0 \leq V_F \leq V_P$. In the further section, we show how V_F behaves as a function of quality in the cost-loss ratio model.

From the contingency table, for the prior information alone, a comparison of expected expenses (πL if action a_2 is taken and C if action a_1 is taken) implies that action to protect should be taken if $\pi > C/L$.

In particular, the minimal expected expense is

$$E_C = \pi L \text{ if } \pi \leq C/L$$
$$E_C = C \text{ if } \pi \geq C/L \tag{3}$$

A similar approach can be taken for the case of imperfect information. Taking an umbrella is optimal if $Pr\,[\Theta = 1|Z] > C/L$. Clearly, if $p_0 > C/L$ or if $p_1 < C/L$ the optimal policy is identical to that of prior information. Therefore, the minimal expenses are the same. Only when $p_0 < C/L < p_1$ there will be a difference in optimal policies. In this case, an umbrella should only be taken when rain is forecasted ($Z = 1$) and

$$E_F = (1 - p_z)p_0 L + p_z C \tag{4}$$

where p_z is defined as

$$p_z = (\pi - p_0)/(p_1 - p_0) \tag{5}$$

Taking the difference in the expected expenses (4-3)

$$V_F = (1 - p_z)(C - p_0 L) \text{ if } p_0 \leq C/L \leq \pi$$
$$V_F = (\pi - p_0)L - pz(C - p_0 L) \text{ if } \pi \leq C/L \leq p_1 \tag{6}$$

3.2 Quality of information

The quality of a forecast has a broader scope to be examined under two main approaches, measure oriented and distribution oriented approaches (Murphy, 1993). Traditionally, forecast quality consisted of the computation of measures of the overall correspondence between forecasts and observations. The measure oriented approach tended to focus on one or two overall aspects of forecast quality, such as accuracy and skill. The perspective provided by the distribution oriented approach reveals that forecast quality is inherently multifaceted in nature. Correspondence between mean forecast and mean observation, overall strength of linear relationship between individual pairs of forecasts and observations, and average correspondence between individual pairs of forecasts and observations were some of the quality aspects.

The expression (6) for the value of imperfect forecasts depends on two parameters, p_0 and p_1 related to the characteristics of the forecasts. To simplfy the the scientific quality of the imperfect weather forecast, long run relative frequency with which adverse weather is forecast is constrained to equal the long-run relative frequency of rain; that is $PrZ = 1 = \pi$. This constraint implies that

$$p_0 = (1 - p_1)\pi/(1 - \pi) \tag{7}$$

The parameter p_1, $\pi \leq p_1 \leq 1$, completely determines imperfect information. Prior information and perfect information are special, limiting cases in which $p_1 = \pi$ and $p_1 = 1$ respectively.

Linear transformation of p_1 gives

$$q = (p_1 - \pi)/(1 - \pi) \tag{8}$$

where q is a measure of relative distance between p_1 and π, making $0 \leq q \leq 1$ with $q = 0$ for prior information and $q = 1$ for perfect information. q is a correlation coefficient between the weather variable Θ and the forecast of adverse weather Z. Hence, q is a relative measure of quality of imperfect information (forecast). It is important to note that paramaters p_1 and q depend only on the probabilistic characteristics of the weather events and forecasts.

3.3 Quality-Value relationship of information

The value of imperfect information is given by

$$\begin{aligned} V_F &= 0 \text{ if } \pi \leq p_1 \leq p_1* \\ V_F &= \pi L(p_1 - p_1*) \text{ if } p_1* \leq p_1 \leq 1 \end{aligned} \tag{9}$$

where

$$\begin{aligned} p_1* &= 1 - (C/L)[(1 - \pi)/\pi] \text{ if } \pi \geq C/L \\ p_1* &= C/L \text{ if } \pi \leq C/L \end{aligned} \tag{10}$$

Note that in the $\pi \geq C/L$ case, p_1* is the value of p_1 for which $p_0 = C/L$. Equations (9) and (10) express economic value V_F as a simple function of the conditional probability of adverse weather condition p_1.

Figure 3 shows the corresponding general relationship between relative value, $v = V_F/L$, and relative quality q. Up to a certain quality threshold, forecasts are of no economic value because the action that is optimal remains the same as when only climatological information is available. Above this threshold, the economic value of the forecasts increases linearly until the upper limit of perfect information is reached. The quality threshold depends on the relative distance between the cost-loss ratio C/L and the climatological probability of adverse weather π, with no threshold only in the special case of $\pi = C/L$. The slope of the linear increase in relative economic value above the quality threshold is $var(\Theta)$, which has its maximum at $\pi = 0.5$. The relative economic value of perfect information depends on both π and C/L and can never exceed 0.25.

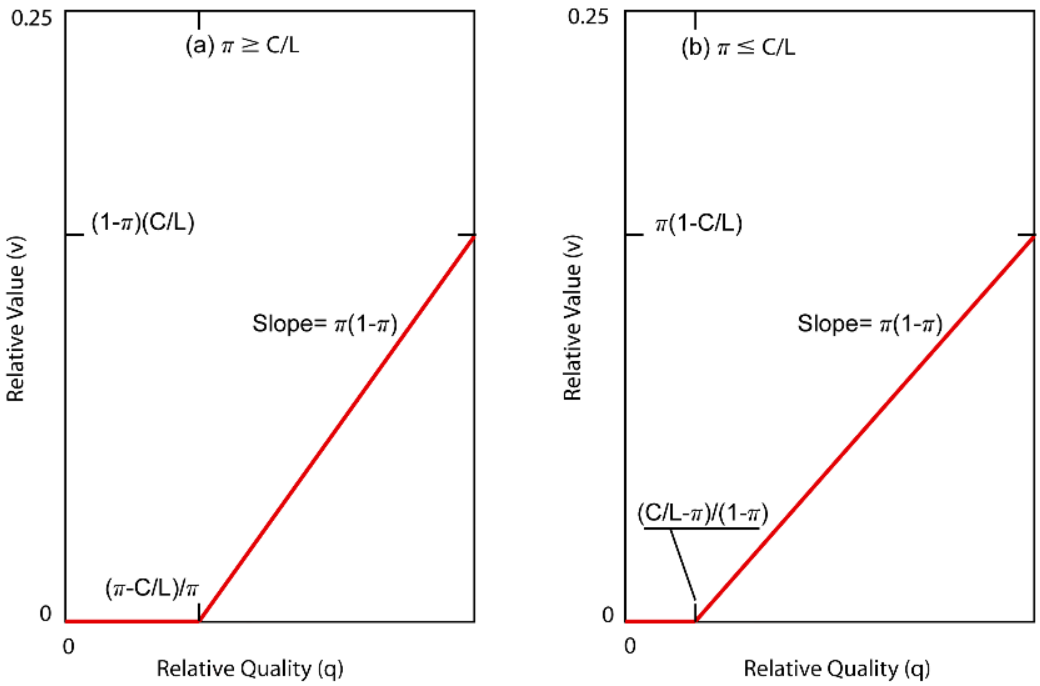

Figure 3: Quality-value relationship

4 Scenario analysis

As a next step, we look at different scenarios for the input variables, C and π, in the model to investigate how the quality-value relationship changes when these variables change. Therefor, we set up the model in Excel and first use a certeris paribus design where either the cost or the probability of rain changes, holding the other variables constant. Moreover, we then also investigate the effect of changes in the distance between π and C/L on the quality threshold as well as the indifference case where $\pi = C/L$, by changing both input variables. In the base case scenario (Figure 4), the following parameters are used: the probability of rain π is 0.3, cost is 40 and loss is 100, resulting in a cost-loss ratio of 0.4. As predicted before, we get a positive, linear relationship between relative quality q and relative value V_f/L, starting from a threshold value of $q = 0.14$ until which the relative value of information is zero.

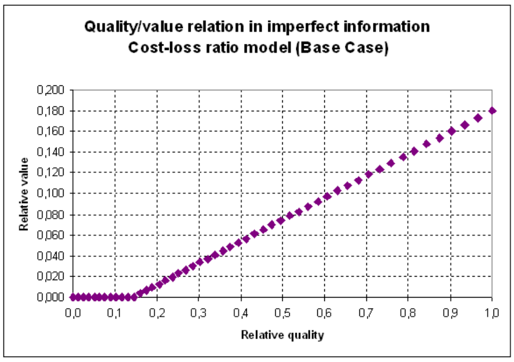

Figure 4: Base Case

Firstly, we investigate how the quality-value relationship changes in case of very high or very low cost C. Therefor, C is set either to 95 or 5. As one would expect, in the high cost scenario information has to be of very high quality in order to possess positive relative value, indicated by a threshold of $q = 0.93$. Similarly, in the low cost scenario this threshold is only slightly lower at $q = 0.84$. So, one can conclude that very high or very low cost leads to a higher quality threshold, keeping the other variables constant. The reason for this is that in case of very high cost of protective action you would almost in all cases not take protective action and in case of very low cost of protective action you would almost in all cases take protective action; only when information is of very high quality, this could change. This is illustrated in the two diagrams of Figure 5.

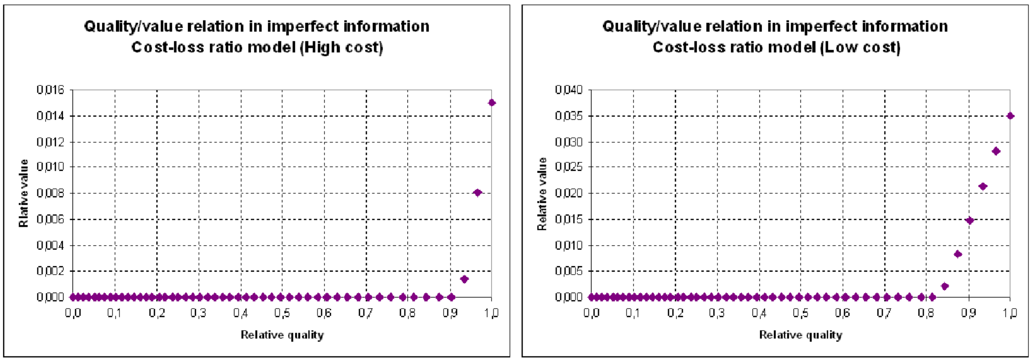

Figure 5: High and low cost scenario

Secondly, we look at the quality-value relationship if the probability of rain is very high or very low. Here, we use values of 0.95 and 0.05 for π. In the high probability case we get a quality threshold $q = 0.58$, which is clearly higher than the threshold of $q = 0.37$ in the low probability case. Therefore, if there is a relatively high probability of rain, information has to be of higher quality in order to be of positive, relative value to the decision-maker. This can be seen in the two diagrams of Figure 6.

Thirdly, we look at the special case where $\pi = C/L$ which means the the prior probability of rain equals the cost-loss ratio. In this case, the decision-maker would be indifferent between taking or not taking an umbrella. As we can see in the following diagram, we get no quality threshold, indicated by the fact that information is always of positive, relative value to the decision maker regardless of its quality. This is shown in Figure 7.

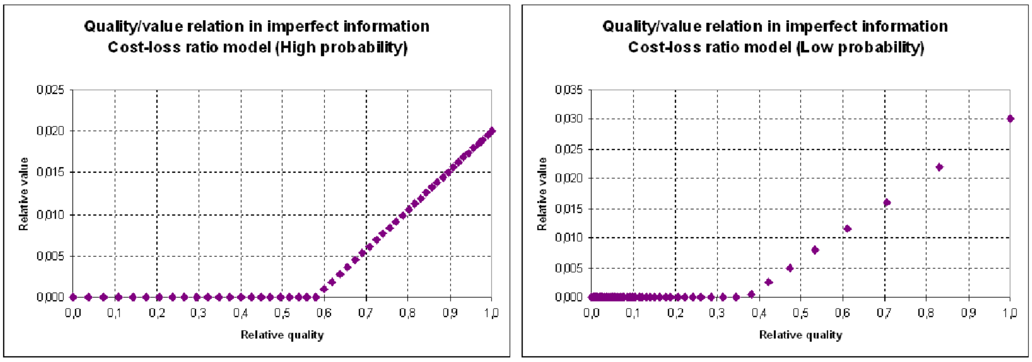

Figure 6: High and low probability scenario

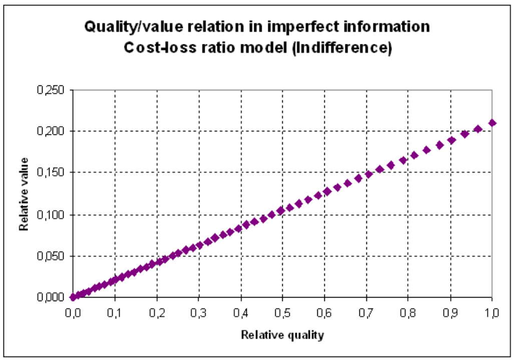

Figure 7: Indifference case

Fourthly, we show that the quality threshold depends on the relative distance between π and C/L. Therefor, we now change both, π and C, and look at a case with a high distance ($\pi = 0.1$, $C = 90$, $C/L = 0.9$) as well as at a case with a low distance ($\pi = 0.3$, $C = 35$, $C/L = 0.35$). It can bee seen in the two graphs in Figure 8 that the high distance between π and C/L leads to a high quality threshold of $q = 0.89$ whereas the low distance leads to a low quality threshold of $q = 0.07$.

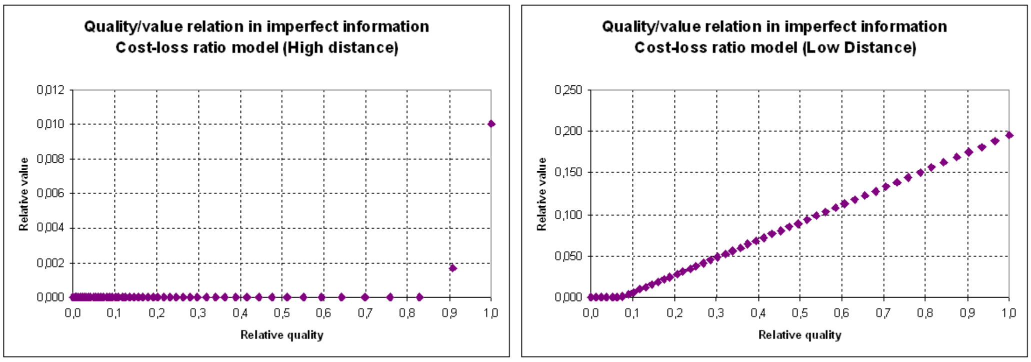

Figure 8: High and low distance scenario

5 Monte Carlo simulation

Finally, using the @Risk simulation extension for Excel, we perform a Monte carlo simulation where π, C and Z were used with uniform distributions as the input variables and relative quality q and relative value V_F/L of information as the output variables.

Figure 9 shows the summary results after 1,000 iterations for relative quality and relative value of information. From this, it can be seen that we get a mean of 0.24 for relative quality and 0.01 for the relative value of information, with a relatively high standard deviation of 0.32 for the relative quality.

6 Conclusion

We have used the framework by Katz and Murphy (1987) to investigate the relationship between quality and value of imperfect information. The relative economic value of imperfect weather forecasts V_F/L was presented as a function of relative quality q. There is a positive, linear relationship between information quality and value, starting from a certain quality threshold ($q = (\pi - C/L)/\pi$ or $q = (C/L - \pi)/1 - \pi$)) until which the relative value of information is zero.

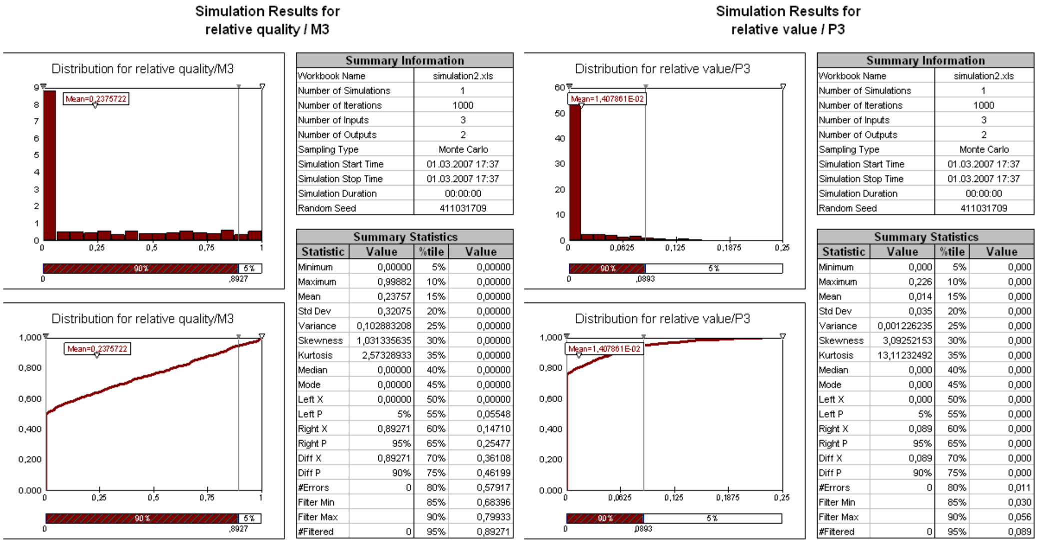

Figure 9: Summary results for relative quality and relative value

This functional relationship was analyzed by looking at different scenarios, where at first either the prior probability of rain π or the cost of information C (and therefore the cost-loss ratio C/L) was changed. Secondly, we have seen that the size of the quality threshold depends on the relative distance between π and C/L and that there is no quality threshold in the special case of $\pi = C/L$. Thirdly, a Monte Carlo simulation was performed where π, C and Z were used as input variables and mean results for relative quality and relative value of information of $q = 0.24$ and $V_F/L = 0.01$ were achieved. Finally, it should be noted that although the analytical framework was successfully tested using scenario analysis and Monte Carlo simulation and plausible results were achieved, there are limitations to this model as it restricts the possible states and actions to two each (rain / no rain; take umbrella / do not take umbrella). Moreover, the theory of decision trees could be used to illustrate the problem faced by the decision maker.

References

Chavas, J. and Pope, R. (1984). Information: Its Measurement and Valuation. *American Journal of Agricultural Economics*, 66:705–716.

Katz, R. and Murphy, A. (1987). Quality/Value Relationship for Imperfect Information in the Umbrella Problem. *The American Statistician*, 41(3):187–189.

Lawrence, D. (1999). *The economic value of information*. Springer, New York.

Lesca, H. and Lesca, E. (1995). *Gestion de l'information, qualitÃ© de l'information et performances de l'entreprise*. Litec, Paris.

Murphy, A. (1993). What is a good forecast? An essay on the nature of goodness in the weather forecasting. *American Meteorological society*, 8:281–293.

Murphy, A. and Winkler, R. (1984). Probability Forecasting in Meterology. *Journal of the American Statistical Association*, 79(387):489–500.

Strong, D., Lee, Y., and Wang, R. (1997). Data Quality in Context. *Communications of the ACM*, 40 (5):103–110.

Thompson, J. (1952). On the operational deficiencies in categorical weather forecasts. *Bull. Amer. Meteor. Soc.*, 33:223–226.