Assessing probabilities of unique events: Calibrating qualitative likelihood judgments into a probability distribution

Warren R. Hughes

(Department of Economics, University of Waikato, Hamilton 3240, New Zealand)

Abstract: Multifaceted events in an organizational environment usually need to be assigned probabilities as a prerequisite to analytical decision-making. If the decision situation is unique, a lack of relevant historical frequency data may preclude use of traditional probability models such as the normal, binomial etc. In this case, an individual decision maker (DM) or an informed group of persons can input into a procedure as outlined here to determine a probability distribution that leads to the expected values of alternative actions or fair values of securities. The individual or group member must decide qualitatively on the extent to which one event is "more likely" than another where both events are ranked adjacent (i.e., closest to each other) in terms of likelihood. Even though the individual or group members may lack experience in orthodox probability assessment, these pairwise "more likely" judgments on the relative likelihood of events are not overly demanding for persons familiar with the possible outcomes in the situation under analysis.

Key words: decision analysis; probability assessment; Bayesian revision; strategic planning

1. Introduction

Decision-making under uncertainty can involve sample spaces comprising events that cannot be modeled using traditional stochastic processes such as the normal or binomial processes. The likelihoods of these events in these complex real-world situations can only be assessed or assigned using the intuition or judgment of the decision-maker. The decision-maker (DM) could be an individual or a group of individuals.

This paper outlines a systematic procedure that takes intuitive or qualitative "more likely" judgments about two events at a time, and transforms them into a probability distribution over all possible events that are mutually exclusive and collectively exhaustive. Real-world managers need a simple procedure for expeditiously deriving probabilities required in everyday decision-making such as calculating the expected value of implementing a given action. Such decision-making may sometimes involve events with unique features where judgment must substitute for historical data such as relative frequencies of past outcomes. The procedure may also be useful in asset valuation situations such as the hostile takeover of a target company. Possible multifaceted or multi-sided events in this case would involve a successful outcome at the current offer price, a competing bid from an alternative acquirer, raised bids by all potential acquirers, possible veto by national and/or overseas regulators etc. The current "fair price" of the target company's stock would involve probability judgments on dollar outcomes influenced by these factors which will usually be unique to the situation in question.

The procedure, applicable for $n \geq 3$ events, is first outlined using objective pairwise likelihoods derived from

---

Warren R. Hughes, DBA, honorary fellow in economics, The University of Waikato; research fields: decision theory, financial economics, impact analysis.

the binomial process. Subsequently, derivation of a probability distribution is illustrated for an individual making subjective "more likely" judgments. Where the DM is a group of individuals, modifications to accommodate possible group dynamics in deriving the "more likely" judgments are discussed. Finally, the procedure can also be adapted to instigate Bayesian revision of a prior distribution (however determined) in the light of new information available to the DM.

2. A binomial illustration

The procedure is first illustrated using an example from the binomial process of tossing a "fair" coin five times where $P(\text{Heads}) = P(\text{Tails}) = 0.5$. There are 2^5 or 32 possible elementary events but here the experiment's outcome will be defined by the number of heads observed after five tosses. In this case, the effective sample space shrinks to six events as in 0, 1, …, 5 heads observed. From the initial sample space of 32 events (all of which are equally likely with a "fair" coin) the events 0 heads (5 tails) or 5 heads are the least likely events and each of these can occur in only one way. These least likely events are defined as having a likelihood of unity. The next least likely events are 1 head or 4 heads (1 tail) and these could occur in $5C1$ or $5C4$ or 5 ways making either of these events 5 times "more likely" than the 0 or 5 heads events. Subsequent calculations follow the same procedure and are summarized in Table 1 which shows the resulting probability distribution in the last column.

Table 1   Likelihoods and probabilities for the number of heads in five tosses of a fair coin

Event	No. of ways event can occur	Pairwise likelihood	Compound likelihood	Probability
0 Heads	1	1 (by definition)	1 = 1	1/32 = 0.03125
5 Heads	1	1/1 = 1	1*1 = 1	1/32 = 0.03125
1 Head	5	5/1 = 5	1*1*5 = 5	5/32 = 0.15625
4 Heads	5	5/5 = 1	1*1*5*1 = 5	5/32 = 0.15625
2 Heads	10	10/5 = 2	1*1*5*1*2 = 10	10/32 = 0.3125
3 Heads	10	10/10 = 1	1*1*5*1*2*1 = 10	10/32 = 0.3125
Total	32		Likelihood total = 32	1.00

In this well-structured problem, the number of ways an event (number of heads) can occur effectively determines the probability of that event. Note that the pairwise likelihood relates to the immediately preceding event where all events are ordered from least to most likely. The least likely event (first event) is defined to have a likelihood of unity. In the above illustration, the pairwise likelihoods (column three) in this example are objectively determined from the binomial coefficients (column two). With some events equally likely, the resulting pairwise likelihood in these cases will be unity. The initial ordering of events from least to most likely allows the decision-maker to determine the pairwise likelihood for events of adjacent likelihood. In particular, this minimizes the problem of estimating the extent to which the likelihood of a very unlikely event is exceeded by that of a very likely event. For example, the compound likelihoods in Table 1 show the most likely events to be 10 times "more likely" than the least likely events. This relative likelihood is objectively determined here by the binomial coefficients, but could be a difficult judgment for a DM to make in circumstances where objective frequencies were not available and qualitative judgments alone are the basis for determining probabilities. The procedure outlined here minimizes the need for such difficult judgments.

The validity of this procedure in a general setting (i.e., no normal, binomial etc. process to generate objective probabilities) can be illustrated with a simple example using three events A, B and C. Let the "true" probabilities

of these events be a, b and c satisfying the usual probability axioms. Of course, these "true" probabilities can only be derived with a fair coin, dice, card set etc. as demonstrated above for the coin tossing example. However, as a thought experiment, we can imagine that "true" probabilities exist and are given as above. With a, b and c the "true" probabilities, then the "true" pairwise likelihoods are correspondingly b/a and c/b where A, B, C rank the events from the least likely event A to the most likely event C. Starting with these pairwise likelihoods, the procedure deriving the resulting probability distribution is outlined in Table 2.

Table 2   Validation of the procedure for a three event problem

Event	"true" pairwise likelihood	"true" compound likelihood	derived = "true" probabilities
A	1	1	1/(1/a) = a
B	b/a	1*(b/a) = b/a	(b/a)/(1/a) = b
C	c/b	(b/a)*(c/b) = c/a	(c/a)/(1/a) = c
Total		1/a	1.00

Given the "true" pairwise likelihoods, the procedure correctly reproduces the "true" probabilities. The pairwise likelihoods such as c/b are reported as a single value, although the DM may determine this value as a ratio of imagined probabilities c and b. This means compensating errors are possible. If b, c are the "true" probabilities and $b*$, $c*$ the imagined probabilities in the reported pairwise likelihood value, it is possible that c/b = $c*/b*$. As a result, the "true" pairwise likelihood is reported with the procedure consequently reproducing the "true" probabilities, even if the imagined probabilities do not collectively satisfy the usual probability axioms. Imagined probabilities may be most useful in evaluating low probability events as in a 1% chance event followed in the ranking by a 5% chance event for a pairwise likelihood value of 5.0.

3. Unique events and a single decision-maker

Unique events may preclude use of probability models based on the binomial, normal and other stochastic processes. Yet, an individual decision-maker may have intuitive ideas as to the relative likelihoods of possible events, and after a sequence of such pairwise judgments, these ideas can be transformed into a probability distribution over all events following the procedure outlined above. The use of pairwise judgments to establish relative benefits, costs etc. of different options are a well-established practice in decision-making especially for the Analytic Hierarchy Process or Analytic Network Process as originally formulated by Saaty (1980, 2005).

All possible events/outcomes in the situation are first ranked in order of increasing likelihood with the least likely event(s) assigned a likelihood of unity. The next least likely event must then be rated as to how much "more likely" it is relative to the least likely (immediately preceding) event. For example, a relative likelihood value of 1.33 for the second least likely event indicates this event to be one third "more likely" than the least likely event. Unlike the events in the coin-tossing illustration above, this judgment of 1.33 would be clearly subjective and may be idiosyncratic of an individual DM. In deciding on the definitive pairwise likelihood value, it may be advantageous for the DM to also consider the reciprocal value of the current "more likely" judgment. This gives a likelihood for the less likely event as a percentage of the likelihood of the more likely event. For example, an initial pairwise "more likely" value of 1.33 (4/3) means the less likely event has only 75% (3/4) of the likelihood of the more likely event. The DM, on reflection, may consider this to be too low with an 80% (4/5) or 83% (5/6) percentage better reflecting the DM's definitive pairwise judgment. As a result, the "more likely" value used in the calculations would then reduce to 1.25 or 1.20 as the case may be. For any pairwise judgment, any relative

frequency information available from past history concerning the two events could be employed in making the definitive "more likely" judgment.

A procedure for mapping qualitative judgments into quantitative values must be pragmatic and simple to follow. The first task in assessing pairwise likelihoods is the "order of magnitude". Assume the DM assesses the "more likely" event of two events to be less than twice as likely as the other event. Following this first judgment, the next step is to assess the extent of the "more likely" magnitude as one of slightly more likely, significantly more likely or substantially more likely. Table 3 outlines a suggested range of quantitative pairwise likelihood values for each of these qualitative judgments.

Table 3 Nomenclature for quantifying qualitative likelihood judgments

Qualifier of "likely" in pairwise comparison of likelihood for the "more likely" event	Pairwise likelihood range	Mid-point
Equally	1.0	1.00
Slightly more	$1.0^+ - 1.2^-$	1.10
Significantly more	$1.2 - 1.8^-$	1.50
Substantially more— just less than twice as	$1.8 - 2.0^-$	1.90
Twice as	2.0	2.00

The above qualitative to quantitative mapping is only suggestive, and perhaps a more detailed nomenclature or taxonomy is possible with associated quantitative values. After the slightly, significantly, substantially "more likely" decision, the DM must assess the appropriate quantitative value within the suggested pairwise likelihood range. These quantitative judgments, which typically will be unique to the situation under analysis, determine the resulting probability distribution.

If the DM initially determines the order of magnitude to be two to three times "more likely", the mid-points of the three appropriate ranges then become 2.10, 2.50 and 2.90 respectively. Ranges for higher orders of magnitude are similarly defined.

As the order of magnitude increases, graduations between the end points may be harder for the DM to discern. Small graduations as in "slightly more than five times more likely" reflect a precision of enumeration that may not be credible given the intuitive nature of these likelihood judgments. Possibly the only meaningful graduation at higher levels of magnitude would be a number such as 5.5 reflecting the judgment of "somewhere between five and six times more likely".

One qualification to this last conjecture is that if the DM can judge that in 20 trials of the situation where one of the two events under comparison occur, the relative frequencies would be say 17 and 3, the likelihood value for the more likely event should be exactly 17/3 or 5.667 to three decimals. It may be that some pairwise comparisons can be evaluated in this way whereas other comparisons can only be evaluated qualitatively in the first instance. Table 3 is a (not necessarily unique) guide for evaluating a likelihood value based only on the DM's initial qualitative assessment on the relative likelihood of two events.

The following example illustrates the procedure based on a sample space of six events, as shown in Table 4, using the nomenclature of Table 3. The least likely event is defined as *A*, and the most likely event is defined as *F*. Mid-points of the appropriate ranges are used simply to illustrate the procedure. Real-world DMs could choose other values in the appropriate ranges better reflecting their subjective likelihood judgments. Probabilities are derived as compound likelihood for the event in question divided by total compound likelihood over all events.

Table 4   Sample calculations for probability assessment over six events *A, B, C, D, E, F*

Event	Qualifier for "likely" over preceding event	Pairwise likelihood	Compound likelihood	Probability	Percentage probability (%)
A	No qualifier for the least likely event at 1.0	1.00	1.0000	0.051118	5.1
B	Slightly more	1.10	1.1000	0.056230	5.6
C	Significantly more than twice as	2.50	2.7500	0.140575	14.1
D	Equally	1.00	2.7500	0.140575	14.1
E	Significantly more than equally	1.50	4.1250	0.210863	21.1
F	Just under twice as	1.90	7.8375	0.400639	40.0
Total			19.5625	1.000000	100

Using the compound likelihood, the above judgments imply that event *F* is just over 7.8 times "more likely" than event *A*. Given the imprecision inherent in the "more likely" judgments for unique events, it is suggested that a three decimal or percentage probability to one decimal (as in only a 5.1% chance for event *A* above) would be an appropriate level of precision for the resulting probabilities generated by the procedure.

4. Probability assessment by a group

The first task of the group is to rank the possible events from least to most likely. If this can be achieved unanimously, then the procedure evolves as for an individual DM. For a group, the "more likely" judgment for each succeeding event after the least likely event is determined as an average (mean, median, mode, geometric mean etc.) of the judgments of the individual group members.

One caveat for group decision-making is that all members should understand the nomenclature of Table 3 as to the meaning of "slightly more likely" etc. First, individuals will differ in their interpretation of what is "slightly more likely" versus "significantly more likely". That is, they could differ in their interpretations as to the meaning of "slightly" versus "significantly". Secondly, for two given events, individuals may have different judgments about whether one event is "slightly" or "significantly" more likely than the other event. Of course, the averaging process naturally mitigates the effects of either kind of variation in interpretations and judgments across the group.

If there is not unanimity in the ranking of events, the above procedure can still be followed. However, resulting "more likely" values for at least one pairwise comparison will involve an average of values above and below unity. Adjustments to eliminate the influence of outliers may be necessary, for example, discarding the lowest and highest values. Alternatively, best practice in this situation may be to use an average such as the mode, median or geometric mean that eliminates or minimizes the influence of outliers.

Another option in the absence of group unanimity as to which of two events is the "more likely" is to use ordinal judgments only. This option may be preferred if group dynamics result in individual "more likely" assessments widely dispersed about unity and/or there is serious discord in the group over the likelihood ranking of all events. This initial ranking of events could be determined by majority vote at each point in the ranking. Suppose *X* and *Y* are two events involved in a disputed pairwise comparison with m group members where *x* members favor *X*, *y* favor *Y* as to which event is "more likely" with *z* favoring an equally likely judgment. For $z \geq 0$, if $x = y$, then events *X* and *Y* should be judged equally likely with a "more likely" value of unity for the resulting pairwise likelihood value. If $x > y$, then the "more likely" value could be determined as *(m-y)/(m-x)*. This value shows the ratio of members not favoring *Y* (the majority) to those not favoring *X* and necessarily exceeds unity for

$x > y$. For $m = 10$, $x = 2$, $y = 0$ and $z = 8$ the pairwise likelihood value will be 1.25 in favor of X over Y. For $m = 10$, $x = 5$, $y = 3$ and $z = 2$, the resulting pairwise likelihood value for X over Y is 1.40, which is less than (say) 5/3 or 1.67, which ignores some information on the pairwise likelihoods.

5. Extremely unlikely events

Some decision situations may involve events that are extremely unlikely as in a "once in a 100-year flood". The consequences of such events may involve extreme value outcomes that should form part of the expected value calculus, even though the probability of such an event obtaining is very low. The probabilities of such events could be assessed using known, objective odds with remaining events assessed as above and renormalized to reflect a proper probability distribution summing to unity over all events. Inclusion of extremely unlikely events may require a final probability distribution using more than three decimals in order to calculate an accurate expected or "fair" value.

6. Bayesian revision

In some cases after derivation of an initial probability distribution, new evidence necessitates a revision of probabilities. In this case, the initial probability distribution (possibly derived as above) becomes the prior distribution. The consistency of the new evidence with each of the possible events (or states of the world) can then be qualitatively assessed using the procedure outlined here. That is, the event (or state) least consistent with the new evidence is assigned a likelihood of unity with pairwise likelihoods assigned to remaining events ranked according to increased consistency with the new evidence. Clearly, this ranking may be completely different to that used in deriving the initial or prior distribution as above. Using the above nomenclature, an event or state "slightly more consistent with the new evidence" than the preceding event or state in the ranking would be assigned a pairwise likelihood in the range 1.0^+ to 1.2^- as outlined in Table 3. The resulting compound likelihood (column 4 in Table 4) becomes the "likelihood" in the usual Bayesian revision procedure where the posterior distribution is proportional to prior times likelihood and normalized to unity. An illustration of this procedure is contained in Hughes (2009).

7. Conclusions

The procedure outlined here demonstrates how a sequence of intuitive or qualitative judgments of the relative likelihoods of two events can be rendered precise resulting in a probability distribution consistent with these judgments over all possible events. As noted, such precision should be limited to percentage probabilities using one decimal (at most) to avoid criticisms of spurious accuracy. Precise probability judgments using established stochastic models as in the binomial, poisson, normal etc. should be utilized if they are appropriate for the situation under analysis. In some cases where both methodologies can be employed, the procedure outlined here will result in the same probability distribution as illustrated above for the binomial example. The procedure is also applicable to Bayesian revision of a prior probability distribution (however derived) in the light of new evidence.

The recent development of so-called prediction markets as in Wolfers and Zitzewitz (2008) also leads to objective probability assessment reflecting the judgments of many participants. The procedure outlined here is not seen as a competitor to these betting/prediction markets. However, betting markets are usually limited to picking a

winner from two or more candidates. Real-world decision problems, on the other hand, typically require judgments on the relative likelihoods of a number of diverse outcomes each having a different consequence or payoff for the DM. The procedure is seen as especially useful for "in-house" probability derivation for a set of possibly complex events that may be unique and/or confidential to an individual or to a small group in the organization analyzing the situation. The business of the organization could involve engineering, legal, medical , mining etc. decision-making. In such cases, the professionals involved in the decision are typically untrained in making probability judgments, but may have good intuition as to the extent to which one outcome is "more likely" than another based on their past experience and expertise. The procedure outlined here results in the derivation of a probability distribution over these events consistent with a sequence of pairwise judgments that do not require sophisticated probability judgments of decision-makers. Individuals need only assess qualitatively the extent to which one of two events or outcomes is "more likely" than the other (or if both events are "equally likely") without justifying their choice in any way. A defined nomenclature or taxonomy from qualitative judgments to numerical pairwise likelihoods allows determination of a probability distribution consistent with those judgments. For Bayesian revision, "consistent with the new evidence" replaces "likely" above.

The procedure does not purport to derive objective probabilities as may be possible with a fair coin etc. and probability models such as the binomial process. The procedure will, however, determine probabilities allowing calculation of expected values etc. that reflect the current intuition of the DM for the possible events in question. At a minimum, the procedure can deliver a benchmark probability distribution that could be further refined to better reflect the DM's final judgments on the probability distribution over all events.

References:
Hughes, W. R.. (2009). A statistical framework for strategic decision making with AHP: Probability assessment and Bayesian revision. *OMEGA, 37*, 463-470.
Saaty, T. L.. (1980). *The analytic hierarchy process.* New York: McGraw-Hill.
Saaty, T. L.. (2005). *Theory and applications of the analytic network process.* Pittsburg: RWS Publications.
Wolfers J. & Zitzewitz, E.. (2008). Prediction markets. In: Durlauf, S. N. & Blume, L. E. (Eds.), *The new palgrave dictionary of economics (2nd ed.).* London: Palgrave McMillan, volume 6, 599-605.

(Edited by Ruby and Chris)