Mathematical and intuitive concepts of probability in decisions under uncertainty in the context of cancer

Yeung, Nga Wendy
University of Warwick, Department of Statistics
Zeeman Hall
Coventry CV4 7AL, UK
E-mail: N.C.Yeung@warwick.ac.uk

Thomas, James
University of Warwick, Department of Statistics
Zeeman Hall
Coventry CV4 7AL, UK
E-mail: James.Thomas@warwick.ac.uk

Brettschneider, Julia
University of Warwick, Department of Statistics
Zeeman Hall
Coventry CV4 7AL, UK
E-mail: Julia.Brettschneider@warwick.ac.uk

1 Fallacies in dealing with probabilities

While mathematicians have universially settled on a set of axioms for probabilities, empirical studies suggested it can not be assumed that people always follow them. The seminal paper Tverksy and Kahneman (1974) and many more articles by the same authors lead to into this branch of research at the crossroad of cognitive psychology, statistics and economics. Deviations from the mathematical rules have been observed when people deal with simple probability exercises as well as with "real world" situations. Experiments with physicians and lawyers show that professional training does not stop people from incorrectly thinking about probabilities in the respective field.

These themes have been studied in depth and have developed into a subfield of behavioural economics. The starting point is a mathematical point of view from which a deviation is observed which would then become known as a particular "fallacy", "bias" or "misconception". Poulton (1994) gives an overview of the most common fallacies and reviews the related research. Here are some frequently mentioned ones.

- **Conjuction fallacy:** This refers to the overestimation of probabilities belonging to combination of events. While the probability of the corresponding intersection of events should be estimated lower than the probabilities of each individual event, people tend to get mislead if the combined events gives a highly consistent picture. This was initially studied in the experiments around the example about the hypothetical bank teller "Linda".

- **Small sample fallacy:** This refers to the incorrect belief that samples of any size have to be equally representative of the whole population. However, this incorrect for heterogenous populations.

- **Certainty effect:** This describes the tendency of humans to prefer the security of certain outcomes to uncertain outcomes (even if that leads to disadvantages in other respects). It is used to explain behaviours known in game theory under the name Allais paradox and Ellsberg

paradox. Participants tend to avoid those gambles which involve uncertainty, although they yield the same expected payoff.

People's thinking about probabilities is more likely to clash with the laws of probabilities in situations that involve one of more of the following characteristics:

i. events with small or unknown probabilities,

ii. complex or concealed decision and information structures,

iii. outcomes that trigger strong emotions such as joy or fear.

2 Human minds versus mathematical probabilities

The mathematical concept of probability came to its full bloom in the early 1930s in Russia where Kolmogorov published his famous treaty "Grundzüge der Wahrscheinlichkeitstheorie". Its core is a set of axioms strongly influenced by the French and German schools of set theory that had been developed in the preceding decades. There is little doubt that this approach has lead the area of probability calculus to an established mathematical discipline.

As a by-product, the mathematical community has carved out a mathematical model mind for thinking about probability. Applications in mind, if any, were mostly within science and engineering. Stepping back from the mathematical perspective for a moment, we need to realise, that we can can not claim a monopoly on modelling the perception, processing or communication of probabilities. Labelling any discrepancies between a "real world" human mind's thinking about probability from the mathematical model mind's thinking about probability as "fallacy" implies that the latter has been accepted as a standard *a priori*.

Whether such a standard is appropriate for a given field of application, however, needs to be examined before any mathematical conclusions can be implemented there. (Note that this is not even obvious is physics: In the context of quantum mechanics elementary particles may not follow the classical laws of probability.) Calling the discrepancies "fallacies" portraits them as errors which should ultimately made disappear or at least minimised. The field of risk communication has been developing methods around such objectives. Gigerenzer (2007), for example, has been developing a framework in defence of intuitive decision-making arguing that is may be more efficient in real-life situations than a theoretically guided perspective would grant it to be.

3 Studying the discrepancies and their consequences

In this work, we shift the emphasis to study the discrepancies and their effects in more detail rather than trying to eliminate them. In addition, we shed light on potential discrepancies that come from applying probabilistic formulas with parameters calibrated on a particular population to a subpopulation.

We focus on a novel type of medical decision processes brought about by the clinical implementation prognostic tools based on technological innovations from genomics. In this context we describe the nature potential discrepancies in thinking about probabilities may have, quantify them, and gauge their effects on the medical decision process. Focussing on a few fallacies that have been empirically established in comparable situations, we identify both qualitative and quantitative consequences for the decision process. Qualitative differences include structural changes in the probabilistic decision model, while quantitative differences refer to increasing or decreasing initial parameters.

Insights into these matters may open up ways to account for the effect of the discrepancies in decision processes and adjust their outcomes, or at least flag them, especially when it is not possible to prevent the discrepencies in the first place.

4 Cancer treatment decisions

Medical treatment decision are taken by physicians and patients and their families, increasingly in a joint decision-making process. These kinds of decisions often involve high levels of uncertainty:

- Diagnostic tests can not detect everything and are prone to measurement error.
- The disease prognosis depends on the overall state of health. Only partial knowledge is available.
- Disease progression itself is a highly variable process depending on a combination of understood and not (yet) understood future events and behaviour.

Processing, comparing and communicating of probabilities is part of the decision process. As detailed in the last section, human intuitions do not always follow the mathematical laws of probability calculus.

Treatment decisions in medicine need to take into account all available information. However, as the collection of some diagnostic information may be expensive, time-consuming, inconvenient or risky, the question which information is gathered becomes itself part of the decision process. Often, the diagnostic information does not uniquely define an action that would certainly be optimal for a particular patient. Instead, patients are being classified into groups for which certain actions are more likely to be beneficial, as suggested by statistical information about past cases. Sometimes, patients and physicians are facing the dilemma choosing whether or not to undergo a costly, risky and difficult treatment in the hope of better fighting the original disease. This is particularly common in cancer, where initial treatments curing the original cancer are being supplemented with treatments hoping to prevent recurrence or the cancer. Maximal amount of information about the chances for cancer recurrence are desirable.

5 Genomic recurrence tests

A novel type of tests addressing the likelihood for recurrence has been introduced for a number of cancers in recent years. They are based on multidimensional genomic information. The underlying technologies are custom-made microarrays or quantitative real-time PCR. They deliver measurements of the biochemical activity of a few dozen genes for a particular patient. The genes have been selected in larger studies on a wide population, usually based on microarrays involving (nearly) the whole genome. Particular example are MammaPrint (see Glas et al (2006)) and Oncotype DX (see Paik et al (2004)) for breast cancer.

The intention of introducing such tests into clinical practice is to support and improve decision-making about treatment choices by assigning a risk category for recurrence to be taken into account in the decision about further preventive treatments. All the characteris i.-iii. listed in Section 1 as factors magnifying discrepancies apply here.

i. In many cases, cancer recurrence probability estimates are vague; wide ranges are given and additional uncertainty is attached. At the same time, their magnitude is not small enough to allow the patient to simply discard them as negligible. They typically are given in 5- or 10-year survival rates leaving open what comes beyond that time frame.

ii. The genomic information is multidimensional and abstract. It is usually does not provide an immediate interpretation for either the patient nor the physician. It may be incoherent. Additional ambiguity may arise when the genomic based recurrence information contradicts other prognostic factors.

iii. The thought of cancer recurrence typically triggers anxiety; more particularly, fear of death, of painful and lengthy treatments, of the state of uncertainty itself. In addition, there may be shame related to stigmas or guilt (in the case of life style related cancers).

6 Discrepancies in the context of genomic recurrence tests

We looked at three major questions.

6.1 How does the certainty effect play into the integration of two different test scores?

We assume a physician has to integrate the results T_1 of a traditional test and the result T_2 of a novel test. Assume they are normalised so that each tests returns a score belonging to one of three categories: low risk (0-17), inconclusive (18-30), and high risk (31-100). Assume that the physician arrives at a combined scores via a convex combination $\alpha T_1 + (1-\alpha)T_2$ of the two original scores with the weighing depending on the type of physician. We differentiate between different types of physicians.

- Act conservatively: This physician treats based on the worst case scenario of all available information.
- Avoid ambiguity: This physician is driven by keeping the available information consistent. As a result, the physician may avoid doing optional tests or discards some of the available information.
- Wary of medical innovations: This physician is keen on taking up new treatments and new diagnostic tests. More experienced physicians are likely to belong to this group.

These should be thought of hypothetical pure states of mind. Real physicians are likely to be located somewhere between those extremes.

Possible scenarios of ambiguous cases are discussed. For example, in the scenario $T_1 \in [0, 17]$ and $T_2 \in [31, 100]$, the conservative doctor would use a relatively large α leading to an inconclusive or even high risk situation. The wary physician may arrive at a similar value, but for different reasons. The ambiguity avoiding physician may chose α to be either close to 0 or close to 1, depending on his or her personal preference among the tests. In conclusion, we have demonstrated how personal preferences of the decision maker can lead to different treatment decisions.

6.2 What are the sensitivity and the specificity of the combined test?

When combining classical assessment and genomic test we build an overall assessment which itself can be seen as a medical test, hence it can be assigned specificity and sensitivity using the regular definition. How these values relate to the corresponding values of the original tests depends on the way we combine them.

We built a decision tree to understand all different possibilities. For simplicity, we assume there are only two possible outcomes in each test, low risk and high risk. The interesting cases are those of ambiguity. We distinguish two options:

BTP Believe-the-positive Rule: Conclude with a positive result if either test or both tests give positive result. So in the context of the prognostic test, place it in the high risk group if any test classies it as high risk.

BTN Believe-the-negative Rule: The opposite of BTN: place it in the low risk group if any test classies it as low risk.

The BTP rule tends to increase sensitivity at the expense of specificity since increasing high risk rate leads to increase in false positive rate. As a result, false positive results carry more weight than false negative results over multiple tests. Similarly, BTN rule tends to increase specificity at the expense of sensitivity. In fact, whether to use BTP or BTN rule depends on the level of recurrence prevalence We have further worked out recurrence rates for all combinations of low and high risk result from the two tests and used it to derive strategies how to combine the two test results.

6.3 How does the predictive values of the genomic test vary among different subgroups?

The prevalence of recurrence plays a role in determining post-test recurrence probabilities. Some clinical factors are known to affect the recurrence probability and are already being taken into account when making treatment decisions. In addition, there are less obvious factors such as young age, pre-menopausal status, alcohol drinking, obesity, some racial, ethnical, religious and educational backgrounds which are associated with higher prevalence rate of recurrence.

Prevalence among each subgroup of patients has a significant impact on the predictive values of the test. With higher prevalence rate, the risk of recurrence for both results of high risk and low risk would increase. If such increase is so big that even a low risk result gives a post-test probability of recurrence which is higher than the cut-off value for adjuvant treatment, the test would then be meaningless (e.g. young age), similar for sufficiently low prevalence rate. In general, the genomic test results may be misleading for several subgroups of patients. Hence, more individual investigation is needed before making treatment decisions.

REFERENCES

Gigerenzer (2007)
Gut Feelings: The intelligence of the unconscious. New York: Viking.

Glas et al (2006)
Converting a breast cancer microarray signature into a high-throughput diagnostic test, BMC Genomics.

Paik et al (2004)
A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer, New England Journal of Medicine 351(27).

Poulton (1994)
Behavioral decision theory: A new approach, Cambridge University Press.

Tversky and Kahneman (1974)
Judgement under uncertainty: Heuristics and biases, Science 185 (4157).