Master's Thesis

The Heuristic of Subsets: A Model of Probabilistic Discrimination

Author: Aram Grigoryan

Supervisor: Florian Biermann

May 2015

Abstract

I introduce the *heuristic of subsets* and test empirically whether people apply it. According to the heuristic, people estimate the unknown probability of a random event based on the probability of a distinguishable subset of the event space to which this event belongs. In particular, people underestimate the probability of an event if it comes from a subset which has low probability and overestimate it if it comes from a subset which has high probability. I construct a model of probability assessment based on the heuristic and show by experiments conducted at Warsaw and Tbilisi that people behave according to what the model predicts. I also discuss examples of economic decision making where the heuristic would be relevant.

Keywords: heuristic, decision under uncertainty, probabilistic discrimination

Contents

1 Introduction

. Imagine a situation where a person has an option to make an investment in one of the ten most profitable companies. The person wants to invest in the company that provides the highest return to the investment. She knows that eight of those companies operate in the field of information technologies (IT), and the other two in sweets industry. She also possesses information about their profitability records (history of returns), which show that investments in either of the companies on average provide the same amount of return. Given the aforementioned two pieces of information we may consider that the person will be indifferent between investing in either of the companies. However, my papers argues that in this example she will tend to overestimate the profitability (the probability to be the most profitable) of any of the IT companies compared to any of the sweets companies, due to the fact that there are more IT companies in the top ten profitable ones and generally it is more likely than the company that provides the highest return to investment is an IT company.

Understanding how people formulate beliefs about uncertain events is of central importance for sciences that deal with decision making under uncertainty. Economics is no exception in this regard. Many decisions of economic agents are based on subjective assessments of likelihoods of certain events. Examples include decisions made by farmers regarding the type and size of production, decisions made by investors regarding the investment portfolio structure, etc. The psychological mechanisms that people use to evaluate the frequency of classes and likelihoods of events have been the central topic of research for psychologists Amon Tversky and Daniel Kahnemann. Their view on people′s decision making mechanism is that when people face a difficult task of judging probability, they use a limited number of heuristics, which help them to bring the problem to a simple one (1974). For example instead of answering the following question 'How likely is it that this candidate could be tenured in our department?" people may answer the much easier question: 'How impressive was the talk?" and use this answer as a proxy for the probability of the former event (Kahneman and Frederick, 2005). The heuristic way of decision making is examined by Gerd Gigerenzer and his colleagues from the Center of Adaptive Behavior and Cognition at Max Planck Institute, who argue that decision making can be modeled by heuristics that help to make inferences with limited time and knowledge. They are models of bounded rationality (Todd and Gigerenzer, 2000). However, the school does not consider this way of thinking as a bunch of systematic biases and cognitive fallacies, but rather as an adaptive toolbox that helps to make smart conclusions from limited information in a short time (Goldstein and Gigerenzer, 1996).

In my work I suggest a model of decision making that takes into account the mechanism I call the *heuristic of subsets.* The heuristic suggested by my paper is the following: when comparing the probabilities of two events people take into account the probabilities of distinguishable subsets to which these events belong. Particularly, people overestimate the probability of an event that belongs to a subset with high probability compared to the probability of the event that belongs to the low probability subset. The heuristic can be applicable to the economic decision making, as in the above mentioned example of choosing between investment opportunities. The described mechanism of probability judgment is consistent with several other findings in decision theory (e.g. representativeness heuristic) and can be used to explain some phenomena (e.g. low of small numbers, Lindas problem). In regards to the influence of group belonging on probability judgment the heuristic can be applicable for describing situations where the events whose probabilities are being compared come from different subsets or groups (e.g. investment opportunities in the financial market, where the groups are the types of the companies).

The remainder of this paper is organized as follows: in section 2. I discuss the existing literature on the topic, in section 3. I present my model of probability judgment based on the heuristic of subsets, section 4. describes the experiments, section 5. presents the results of the experiments and section 6. concludes.

2 Literature Review

The heuristic approach of probability assessment has been revolutionized by Amon Tversky and Daniel Kahneman. In their paper *Judgment Under Uncertainty: Heuristics and Biases (1974)* they introduced systematic processes that determine decisions under uncertainty. People use these tools to estimate unknown probabilities. Among other empirical findings of systematic biases in probability judgment Tversky and Kahneman discovered the *conjunction fallacy (1983).* One of the characteristic rules of the probability space is that the probability of a conjunction can never exceed the probability of its constituents as the former is included in the latter. The classical example of the fallacy is the Lindas Problem. The authors conducted an experiment where the participants consisting of eighty eight undergraduates gave answers to the following description of a girl named Linda and they were asked to rank the probabilities of the below mentioned scenarios:

"Linda is 31 year old, single, outspoken and very bright. She majored in philosophy. As a student she was deeply with issues of discrimination and social justice, and also participated in anti-nuclear demonstrations."

(1) Linda is active in the feminist movement.

(2) Linda is a bank teller.

(3) Linda is a bank teller who is active in the feminist movement.

According to the conjunction rule the probability of (3) can never be more than the probability of (2) or (1). However, eighty five percent of respondents stated the following hierarchy of probabilities: $P(1) > P(3) > P(2)$. Authors explain this phenomenon by the fact that people see the *representativeness* of an event as a means to evaluate its probability. Representativeness heuristic implies that people judge the probability that (3) is true by the degree to which it is representative of the description, that is, to which it resembles the description. From the above mentioned passage about Linda she is believed to be representative for active feminist and unrepresentative of a bank teller, hence (3) should be more appealing to the respondents than (2) in regards to plausibility.

Representativeness heuristic also explains the fact that people tend to follow what Tversky and Kahneman (1971) call the law of small numbers, that is - people believe that random samples of a population will resemble the population more frequently than statistical sampling theory would predict. For example, when asked to imagine fair coin flipping in mind and write down the sequence of coin tosses people tend to make the number of heads and tails equal. According to the law of large numbers when the sequence gets bigger the percentage of flipped heads should equal to fifty percent. However, when the sample size is small the proportion of heads is not as close to fifty percent as people usually imagine. According to the law of small numbers people may believe the sequence $T\ T\ H\ H\ T\ H$ to be more probable than the sequence $H\ H\ H\ H\ H\ H$. Although the probabilities are the same, the former sequence resembles the population (i.e. a long enough sequence will have equal shares of tails and heads), in other words, the former is more representative of a fair coin toss sequence.

Due to this belief that small samples should resemble to populations people avoid writing down long sequences of only heads or tails, hence after a sequence of tossed heads people are more likely to predict a tail. The phenomenon is known to the literature as the *gamblers fallacy*. As people assume that the number of heads and tails should become equal even in a small sample after a sequence of heads people believe that in the subsequent coin flipping a tail should appear to balance the proportion of tails.

Although the representativeness heuristic is informative of peoples judgment under uncertainty and can be used to predict their behavior, the heuristic has not been formalized and modeled successfully. The reason for this is that, although it is perceived in the intuitive level, it is difficult to formalize what it means to be similar or representative. Subset belonging is easier to formalize, however some strong assumptions still need to be made.

Attempts to put probability assessment into a model, have been done previously. One such attempt has been carried out by Einhorn and Hogarth (1985). Their model describes how people make inferences about probabilities in ambiguity by using anchoring and adjustment strategies. People's beliefs over probabilities thus depend on their prior valuations (which can be the best guess of an expert) and some adjustment term. Another model by Goldstein et al. (2002) is based on the so called *recognition heuristic*, which argues that in some settings people evaluate the probability of an event merely based on whether they recognize it or not. The heuristic that the authors discuss is simple and easy to implement and can lead to very accurate estimates for probability assessment given an appropriate setting holds. My model of probability assessment is a more complicated mechanism of bounded rationality where agents make estimations about events probabilities from the subsets they belong to. Provided that we have the appropriate setting the model can be helpful to make judgments regarding people′s beliefs and thus to predict behavior.

3 The Model

Let a person know the probabilities of two events A and B, such that $P(A) > P(B)$. The person also receives information that $a \subset A$ and $b \subset B$, for some a and b belonging to the event space Ω (Figure 1). If the person does not have other information to assess the probability relation of a and b than that a belongs to A and b belongs to B, she will estimate a to be more probable than b. The constraint on the subset belonging is a constraint on events likelihood, and rational behavior implies that people should consider the fact that a belongs to a subset with higher probability as a signal for its relatively high probability. However, when the relation of probabilities of a and b can be actually estimated directly, the effect of the subsets probability

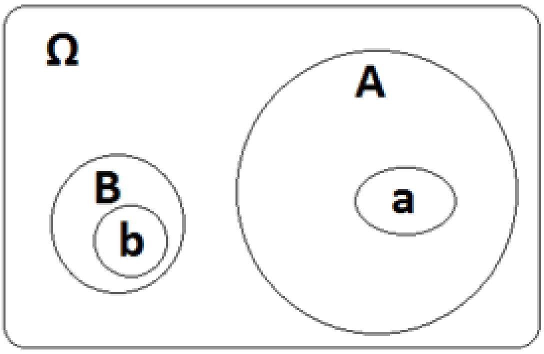

Figure 1:

should be non-existent. The heuristic suggests that the effect can still be existent. A justification for this is that people can never be absolutely sure of the actual probabilities of a and b, hence, they need to take into account the events subset belonging. This for example can happen in our example of investment opportunities. People receive information from which they can infer that the possibility of each of the firms to be the most profitable is the same, and the fact that IT companies among the top profitable are the majority should not play a role on belief formation. However, even a small suspicion over the validity of the profitability records can make people rely on subset belonging.

My model of probability assessment is limited to a situation where people compare the probabilities of two elements (a and b) from the event space Ω, that belong to different subsets ($A, B \subset \Omega$). People receive some information from which they make inferences regarding the probabilities of some events in Ω. For example when people receive information that a fair coin has been flipped four times and during the third toss a head appeared, people may infer that the probability that all tosses are tails is zero and the probability that all tosses are heads is one over eight. I denote by $P(e)$ the *actual probability* of event $e \subset \Omega$, which is the probability that a person will assign to the event if he had infinite time and skills to process the given information. However, as people do not have infinite time and skills their estimations of events probability are not uniquely determined but include a random term by which the estimates differ from the actual probabilities.

First let us discuss the situation where the probability of a subset to which an event belongs does not influence the estimation of events probability. I introduce an operator Est (estimation) such that:

$$Est[P(e)] = P(e) + \mu$$

μ is symmetrically distributed error term with mean 0 and some standard deviation . The standard deviation of the error term depends on the event, whose probability is estimated. We do not require the error to be independent of actual probability. The operator Est has the following properties:

1.

$$Est[P(e)) + P(f)] = Est[P(e)] + Est[P(f)]$$

(for arbitrary e,f $\subset \Omega$)

2.

$$Est[P(e) \times P(f)] = Est[P(e) \times Est[P(f)]$$

These guarantee that the estimations for the sum and multiplication for arbitrary events‘ probabilities are also not biased. Not biasedness condition is the following:

$$E(Est[P(e)]) = P(e)$$

In my model, however, the estimations of probabilities also depend on the probabilities of a distinguishable subset to which these events belong. Suppose that people have received some information regarding a,b,A and B such that, $a \subseteq A \subset \Omega$ and $b \subseteq B \subset \Omega$. In this case the estimation of the probability difference between a and b can be written in the following way;

$$Est[P(a) - P(b)] = P(a) - P(b) + \gamma \times [P(A) - P(B)] + \mu \tag{1}$$

If γ is not equal to zero estimations will be biased, and the bias will equal $\gamma \times [P(A) - P(B)]$. The *heuristic of subsets* argues that γ should be statistically greater than 0. This has been checked by the experiments that will be discussed in the subsequent chapters. Note that γ depends on the operand and on the event whose probability is being estimated. Here I do not talk about a linear effect of subset belonging. The effect of subset belonging will be different when comparing different events' probabilities, however it should always be non-negative.

My model of probability estimation can explain several phenomena that have previously been attributed to the representativeness heuristic. In the Lindas problem we deal with a situation where people compare the probabilities of the following two events:

a - Linda is a bank teller who is active in the feminist movement.

b - Linda is a bank teller.

The empirical evidence has shown that people people assign lower probability for the latter, although the former is a special case of b and it should always hold that $P(a) \leq P(b)$. However, from the experiment results one may conclude that $E(Est[P(a) - P(b)]) > 0$.

Let us look at the distinguishable subsets to which these elements belong. The distinguishable subset to which b belongs is b itself. So we take $B \equiv b$. a belongs to the subset A, which is the following event -Linda is active in the feminist movement (figure 2).

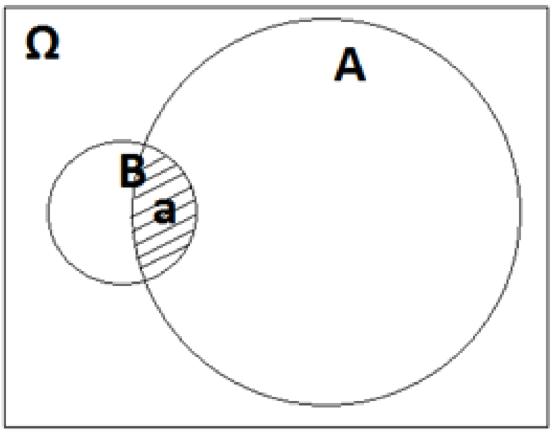

Figure 2:

By the description of Linda it makes sense to assume that the probability of Linda being active in the feminist movement is higher than Linda being a bank teller, as from the passage describing Linda no inference can be made in favor of her working in a bank, however it was stated that she is concerned with issues of discrimination and social justice, which will probably imply participation in the feminist movement. Hence, we can assume that $P(A) > P(B)$, at least, people believe so. Given the positive value of γ our model can explain the conjunction fallacy. The positive bias $\gamma \times [P(A)P(B)]$ can explain the fact that the left hand side of equation (1) is positive, even though $P(a) - P(b)$ is negative. The most questionable part of the explanation is the division into subsets. Here we assume that the distinguishable subset to which a belongs is A, and people indeed use its probability as a proxy for evaluating that of a, however there is another subset (that is B) to which a belongs. Although my model is used to describe probability assessment in a setting where each of the two events belong to one distinguishable subset each, the logic by which the subset influences events probability can be used to give intuition to Lindas Problem also.

Another problem that has been explained by the representativeness heuristic is the phenomenon known as the law of small numbers. The setting of the problem here is similar to the setting of my model and the explanation is more accurate. Imagine a situation where a fair coin is tossed 4 times and a person compares the probabilities of the following events.

a - H T T H

b - H H H H

According to the representativeness heuristic the former is considered more probable as it is more similar and representative to the population (which contains equal percentage of heads and tails). My model gives another explanation to the problem. In the intuitive level it is justified to assume that people distinguish two subsets for a and b. These subsets are;

A The sequence includes two heads and two tails.

B The sequence consists of all heads or all tails.

Given the assumption that people distinguish the above mentioned subsets for a and b and given that is positive our model can explain the fact that people estimate the probability of a to be higher than that of b, even though $P(a) = P(b)$. As $P(A) > P(B)$ in equation (1) we get $Est[P(a) - P(b)] > \mu$, hence $E(Est[P(a) - P(b)]) > 0$. The fact that we use A and B as subsets of interest is because according to low of small numbers each of the events in A (i.e. *HHTT,TTHH,HTHT,THTH,HTTH,THHT*) is overestimated than each of the events in B (i.e. *HHHH,TTTT*). This is very similar to the situation that I simulated in experiments 1 and 2. In a setting where two events probabilities are being estimated and each of them belong to a distinguishable subset the heuristic of subset explains phenomena that have previously been linked with the representativeness heuristic. However, the representativeness heuristic is a broader and less formal concept, whereas the heuristic of subset can be modeled and accurately used to predict behavior given the appropriate setting.

4 The Experiments

Several experiments have been constructed to check the validity of the presented model which is based on the heuristic of subsets. I have constructed experiments where people make bets on events a and b which have the same probability and belong to subsets A and B respectively, such that $P(A) > P(B)$. Given the $\gamma > 0$ condition, our model will predict that people will estimate higher probability for a. Showing that people tend to bet on a rather than b will confirm that people do so, hence, the model gives realistic prediction, and the heuristic holds in this experiments.

The experiments have been conducted by the participation of students from the International School of Economics at Tbilisi State University in Georgia and Warsaw University of Life Sciences in Poland. The number of participants was 27 and 39 respectively. Several types of experiments have been used to check the hypothesis.

Experiment 1

The following experiment has been conducted. The participants were told that a ball was being randomly chosen from a box containing five balls.The experiment was conducted with real boxes and balls. Each of the balls was unique; each of them had a unique symbol(s) written on it. The balls also differed by color, namely, three of the balls were white and two of the balls were yellow (Figure 3).

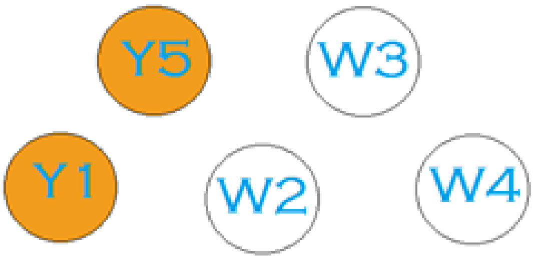

Figure 3:

The participants were asked to bet on a particular ball (i.e. one of the five) being randomly chosen. Let us denote by a and b the events that a particular white and a particular yellow ball (respectively) are chosen. By A and B we denote the events that generally a white ball and a yellow ball is chosen. Hence, we get $a \subseteq A$ and $b \subseteq B$. Systematically avoiding betting on yellow balls implies that people estimate the probability of a to be higher than the probability of b, that is the left hand side of equation (1) is in expectation bigger than zero as the model would predict.

Experiment 2

In the second experiment participants were told that a ball is being taken out from a box containing twelve balls, of which three were yellow and nine were white. Three of the white balls had the same number '27' on them. Similarly, the three yellow balls had the same number '23' on them. Each of the six remaining white balls had a unique number on it (Figure 4).

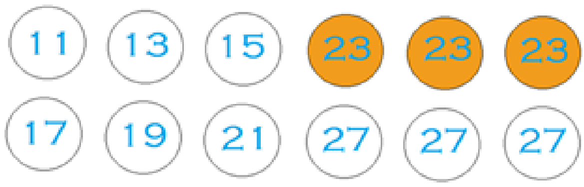

Figure 4:

People should bet on either'27' or '23' as the probabilities of having each of these is three times higher than those of the remaining ones. However, people also should be indifferent between '27' and '23'. The heuristic however suggests that as the probability of having a white ball taken out from the box is in general higher, people will be biased in favor of betting on the number'27', which appears on white balls.

Experiment 3

Participants are told that a ball is taken out from a box containing three balls. Two of the balls are black and one is white. In addition, two balls have letter E on it and the third one has the letter F on it (figure 5).

Participants were told that a ball had been chosen randomly and it appeared to have letter E on it. The participants were asked to bet whether they think the ball is white or black. According to the heuristic people would assign

Figure 5:

higher probability for the event- the ball is black and E, compared to the event- the ball is white and E. Hence, they would tend to bet on black rather than white.

5 Results

The first experiment was conducted at the International School of Economics at Tbilisi State University in Georgia. Thirty eight students of an MA program in economics participated in the experiment. Results of the first experiment are presented in the table below (Table 1).

Table 1:

Color	Actual number of bets	Expected number of bets	t-statistic	p-value
White	30	22.8	2.38**	0.0112
Yellow	8	15.2		
Total	38	38		

As my model of would have predicted the participants tended to bet on white balls. The total number of bets on white balls was 30 out of 38, although if the bets were random that number would have been 22.8. The results are statistically significant with p-value 0.0112. The Null Hypothesis (i.e. people are indifferent between betting on whites and yellows) is rejected.

The second experiment was conducted in the same school, with the same students as the first experiment. As we want to estimate the tendency of choosing between ‘23′ and ‘27′ we consider only people who made bets on either of these numbers. The results of the experiment are shown below (Table 2).

Table 2:

Number	Actual number of bets	Expected number of bets	t-statistic	p-value
‘27‘ 22	15	2.555***	0.008	
‘23‘ 8	15			
Total	30	30		

Here we compare the numbers of bets on ‘27‘ and ‘23‘. From the results we see the bets on the number ‘27‘ appearing on white balls, which constitute the majority, are higher than we would expect if people were indifferent between ‘27‘ and ‘23‘. The results are statistically significant with p-value ¡ 0.01.

The third experiment has been conducted at the Warsaw University of Life Sciences in Poland. The participants were students with different academic levels and fields of studies. The total number of participants was twenty seven. The results are shown below (Table 3).

Even if they are told that the ball was of type E, people tend to bet that the ball was black. The number of bets on black was 22 out of 27, which is higher than the expected number of bets when the participants randomize between black and white.

Table 3:

Color	Actual number of bets	Expected number of bets	t-statistic	p-value
White	5	13.5	3.27***	0.0015
Black	22	13.5		
Total	27	27		

6 Conclusion

The results of the experiment proved that the model can actually predict how people will behave in a certain design. The division into subsets based on one characteristic (i.e. color) made people discriminate between the identical (in the context of likelihood of being chosen) balls.

A possible limitation for generalizing the experiment results is that there may be other factors that make people bet on certain numbers of colors. For example, people may have preferences for certain numbers, symbols or colors. It may happen that peoples bets are not only based on probability estimations but also tastes. In my study I make the assumption that these factors do not play a significant role. Even if it is possible that everybody made bets based on their individual tastes, it is not very likely that their uncorrelated preferences (if such) led to systematic tendency towards betting on particular symbols or colors.

Another limitation to the model is that it can be used to make predictions only in special circumstances, that is, only two events probabilities are being compared and the events should belong to one distinguishable subsets each, with different probabilities. The requirement creates difficulties for the models applications as it is not easy to justify what is the subset that people distinguish for an event. In my experiments I tried to make people distinguish the subsets by one characteristic: the color. However, people may think of other factors for dividing the events into subsets (e.g. odd even numbers). Although there exist limitations in applications the model is important for making predictions in a particular setting and can be used to explain several phenomena in probability judgment.

7 References

Einhorn J. Hillel and Robin M. Hogarth. 1985. Ambiguity and uncertainty in probabilistic inference. *Psychological Review,* Vol 92(4), 433-461

Gigerenzer Gerd and Daniel G. Goldstein. 1996. Reasoning the Fast and Frugal Way: Models of Bounded Rationality Psychological Review, Vol. 103(4), 650-669

Goldstein G. Daniel and Gerd Gigerenzer. 2002. Models of Ecological Rationality. *Psychological Review,* Vol. 109, No. 1, 7590

Kahneman Daniel and Shane Frederick. 2005. A Model of Heuristic Judgment *The Cambridge Handbook of Thinking and Reasoning,* Chapter 12

Todd M. Peter and Gerd Gigerenzer. 2000. Prcis of Simple heuristics that make us smart. *Behavioral and Brain Sciences*, 723-780

Tversky Amos and Daniel Kahneman. 1971. Belief in the law of small numbers. *Hebrew University of Jerusalem Psychological Bulletin,* Vol. 76, No. 2, 105-110

Tversky Amos and Daniel Kahnemann. 1974. Judgment under Uncertainty: Heuristics and Biases. *Science, New Series,* Vol. 185, No. 4157. pp. 1124-1131

Tversky Amos and Daniel Kahneman. 1983. Extensional versus intuitive reasoning: The conjunction fallacy in probability judgment. *Psychological Review*, Vol 90(4), 293-315

8 Appendices

Figure 6:


Rules
Boxes with balls
Random selection
Submit your bets
Correct guess is rewarded by 3 GEL for first two games and 2 GEL for the last game.
No discussions among each other
Part I
5 balls
Submit your bet
Y5
W3
Y1
W2
W4
Part II
12 balls
Write down one of the numbers
Submit your bet
11
13
15
23
23
23
17
19
21
27
27
27

Figure 7: