ONE OBSERVATION BEHIND TWO PUZZLES

DOV SAMET, IDDO SAMET AND DAVID SCHMEIDLER

1. Two puzzles on the theme "Which is larger?"

In two famous and popular puzzles a participant is required to compare two numbers of which she is shown only one. We show that there is one simple principle behind these puzzles. In particular this principle sheds new light on the paradoxical nature of the first puzzle.

According to this principle the ranking of several random variables must depend on at least one of them, except for the trivial case where the ranking is constant. Thus, in the non-trivial case, there must be at least one variable the observation of which conveys information about the ranking.

A variant of the first puzzle goes back to the mathematician Littlewood (1986) who attributed it to the physicist Schrödinger. See Nalebuff (1989), Brams and Kilgour (1995), and Blackwell (1951) for more detail on the historical background and for further elaboration on this puzzle. Below is the common version of the puzzle as first appeared in Kraitchik (1953).

> **To switch or not to switch?** There are two envelopes with money in them. The sum of money in one of the envelopes is twice as large as the other sum. Each of the envelopes is equally likely to hold the larger sum. You are assigned at random one of the envelopes and may take the money inside. However, before you open your envelope you are offered the possibility of switching the envelopes and taking the money inside the other one. It seems obvious that there is no point in switching: the situation is completely symmetric with respect to the two envelopes. The argument for switching is also simple. Suppose you open the envelope and find a sum x. Then, in the other envelope the sum is either $2x$ or $x/2$ with equal probabilities. Thus, the expected sum is $(1/2)2x + (1/2)x/2 = 1.25x$. This is true for any x, and therefore you should switch even before opening the envelope. Should you or should you not switch?

Solutions to this paradox are discussed in numerous articles. The simplest and most common solution is this. In order to carry out the computation of the expected value, there must be some probability distribution over the two sums. But no probability distribution can have the property that for any envelope, and any given sum x in it, the sum in the other is equally likely to be $2x$ and $(1/2)x$.[1]

Date: May, 2002.
Helpful comments by David Gilat are acknowledged.

[1]The standard argument for the non-existence of the required probability distribution relies on the non-existence of a uniform probability over a countable set. A similar argument can be made when the support is a continuum, Schuss (2000). We provide here an argument that does

The next puzzle is due to Cover (1987). The gist of it appeared already in Blackwell (1951) (see footnote 4 below).

> **Guessing which is larger.** Two different real numbers are each written on a slip of paper facing down. One of the two slips is chosen at random and the number on it is shown to you. You have to guess whether this is the larger or the smaller number. How can you guarantee that the probability of guessing correctly is more than half, *no matter what the numbers are*?

It is somewhat surprising that there is a way to guarantee it. Because it seems that one cannot learn anything about the order of two numbers by observing one of them. But unlike the first puzzle this one is not paradoxical. There is indeed a method, discussed in the last section, guaranteeing that the probability of guessing correctly is larger than half.

In both puzzles the participant faces the same problem: which is the larger number. But there is more to it. The solution of the second puzzle explains the flaw in the first one. It serves as a *proof* that there is no probability distribution as the one assumed in the first puzzle. Indeed, had such a distribution existed, then selecting the two numbers of the second puzzle from this distribution would have frustrated your ability to guess correctly with probability greater than half. Because whichever number x you see, the other number is equally likely to be larger (i.e., $2x$) or smaller (i.e., $(1/2)x$). Thus, the fact that you can make your guess in such a way that you are more likely to be correct than not to be, shows that there is no such distribution. We make this statement more precise in the next section.

The proof in the previous paragraph demonstrates a more general impossibility. It shows that there is no probability distribution over pair of numbers, x_1 and x_2 such that each is equally likely to be larger, and such that observing any one of them does not change this equallity. This conclusion is phrased in terms of conditional probabilities, but it can be rephrased in more elementary terms. What the second puzzle shows is that there is no probability distribution over pair of numbers, x_1 and x_2 such that each is equally likely to be larger, and such that the order of the numbers (as a random variable) is *independent* of each of the numbers.

This principle can be further generalized: the equal likelihood and the restriction to two numbers are not essential. We show here that it is impossible to have n random variables, such that observing any of them conveys no information on their ranking, unless the ranking is constant. In the next section we show how the second puzzle solves the first, and in the last section we state and proof the generalized result.

2. The second puzzle solves the first

To switch or not to switch? Let X_1 and X_2 be two random variables which describe the sums in the the first and the second envelope correspondingly. The puzzle stipulates that for any observation x_1 of X_1,

$$P(X_2 = 2X_1 | X_1 = x_1) = P(X_2 = (1/2)X_1 | X_1 = x_1) = 1/2,$$

and similarly, for any observation x_2 of X_2,

not require this distinction. More importantly, it is formulated in general terms of learning from an observation which seems to lend new insight into this problem.

$$P(X_1 = 2X_2 | X_2 = x_2) = P(X_1 = (1/2)X_2 | X_2 = x_2) = 1/2.$$

This implies that for any measurable set of numbers A,

$$\begin{aligned} P([X_2 = 2X_1] \cap [X_1 \in A]) &= P([X_2 = (1/2)X_1] \cap [X_1 \in A]) \\ &= (1/2)P(X_1 \in A), \end{aligned}$$

and

$$\begin{aligned} P([X_1 = 2X_2] \cap [X_2 \in A]) &= P([X_1 = (1/2)X_2] \cap [X_2 \in A]) \\ &= (1/2)P(X_2 \in A). \end{aligned}$$

These conditions are even weaker then the ones stipulated in the puzzle, and do not require any assumption on conditional probabilities or densities.

Since the event $X_1 = 2X_2$ is the event $X_1 < X_2$ and similarly the event $X_1 = (1/2)X_2$ is the event $X_1 > X_2$ it follows that

$$\begin{aligned} P([X_2 < X_1] \cap [X_1 \in A]) &= P([X_2 > X_1] \cap [X_1 \in A]) \\ &= (1/2)P(X_1 \in A), \end{aligned} \tag{1}$$

$$\begin{aligned} P([X_1 < X_2] \cap [X_2 \in A]) &= P([X_1 > X_2] \cap [X_2 \in A]) \\ &= (1/2)P(X_2 \in A). \end{aligned} \tag{2}$$

These two conditions can be summarized as follows. The probability of each of the events $X_1 > X_2$ and $X_1 < X_2$ is $1/2$ (substitute the real line for A in (1) and (2)), and these two events are independent of X_2 and X_2.

Claim 1. *It is impossible for a pair of random variables X_1 and X_2 to satisfy* 1 *and* 2.

We show that this claim follows from the second puzzle.

Guessing which is larger. It is helpful to present this puzzle as a two-person zero-sum win-lose game. The first player, C, chooses the numbers, while the second, G, makes the guess after observing the number on one of the slips that was chosen at random. Player G wins iff she guesses correctly.

The pure strategies of C are pairs (x_1, x_2) of real, distinct numbers. A mixed strategy of C is a pair of random variables (X_1, X_2) such that $P(X_1 \neq X_2) = 1$. We restrict G's pure strategies to *threshold* strategies. Each $t \in \mathbb{R}$ represents the threshold strategy where the player guesses that the observed number x is the larger if $x \geq t$ and is the smaller otherwise, independently of which slip she observes.[2] Mixed strategies of G are probability distributions over $\mathbb{R}$. We note first,

Claim 2. *If G plays an arbitrary threshold strategy t against any pure strategy (x_1, x_2) of C, she*

- *wins with probability $1/2$ when either $x_1, x_2 < t$ or $x_1, x_2 \geq t$,*
- *wins for sure when either $x_1 < t \leq x_2$ or $x_2 < t \leq x_1$.*

[2]One can think of more general pure strategies in which the guess is any function of the observed number and the slip that was chosen. But in order to guarantee that the probability of winning is higher than $1/2$, threshold strategies suffice.

Indeed, in the first two cases G's guess is the same whether she observes x_1 or x_2. Her guess is correct with probability 1/2. In the last two cases G guesses correctly whether she observes x_1 or x_2.[3] □

Consider a mixed strategy Q of player G such that for each $a < b$, $Q((a,b]) > 0$.

Claim 3. *The strategy Q guarantees that the probability that player G wins is higher than $1/2$ against any pure strategy of C.*

Consider the strategy (x_1, x_2) of C such that $x_1 < x_2$. If $x_1 < t \leq x_2$, which happens with probability $Q((x_1, x_2])$, then G wins for sure. In all other cases G's chance of winning is half. Thus her chances of winning are $1/2 + Q((x_1, x_2]) > 1/2$. The case $x_2 < x_1$ is similar. □

The proof above, that G can guarantee that the probability of winning is higher than 1/2, is due to Cover (1987). We point out that this result, viewed in the context of the zero-sum game, can be used to establish a proof of Claim 1.

Proof of Claim 1: We show that if C is using a mixed strategy (X_1, X_2) that satisfies (1) and (2), then player G cannot guarantee that the probability she wins is higher than 1/2, in contradiction to Claim 3.

The probability that G guesses correctly using the threshold strategy t is

$$\begin{aligned}&\mathrm{Prob}(G \text{ observes } X_1)\big(P([X_1 > X_2] \cap [X_1 \geq t]) + P([X_1 < X_2] \cap [X_1 < t])\big)\\ &+ \mathrm{Prob}(G \text{ observes } X_2)\big(P([X_2 > X_1] \cap [X_2 \geq t]) + P([X_2 < X_1] \cap [X_2 < t])\big).\end{aligned}$$

Since G observes each of X_1 and X_2 with probability 1/2, and by conditions (1) and (2), this probability is

$$\begin{aligned}&\frac{1}{2}\big(\frac{1}{2}P(X_1 \geq t) + \frac{1}{2}P(X_1 < t)\big)\\ &+ \frac{1}{2}\big(\frac{1}{2}P(X_2 \geq t) + \frac{1}{2}P(X_2 < t)\big) = \frac{1}{2}.\end{aligned}$$

Thus, the mixed strategy (X_1, X_2) guarantees player C that the probability she wins is 1/2 against any pure threshold strategy of G, and hence also against Q, which is a contradiction to Claim 3. □

3. Ranking by one observation

The following proposition generalizes Claim 1. Let $X = (X_1, \ldots, X_n)$ be a vector of n real-valued random variables on some sample space. Denote by $\mathbb{W}$ the set of weak orders over $\{1, \ldots, n\}$.[4] The *ranking* of X is a random variable $r(X)$ with values in $\mathbb{W}$. For $W \in \mathbb{W}$, $r(X) = W$ whenever for all $i, j \in \{1, \ldots, n\}$, $X_i \geq X_j$ if and only if $i\,W\,j$.

Proposition 1. *For any vector of random variables $X = (X_1, \ldots, X_n)$, if for each i, X_i, and $r(X)$ are independent then $r(X)$ is constant almost surely.*

[3] Blackwell (1951), in Example 1, introduces a special case of this puzzle. He uses a similar threshold estimate to improve upon the constant estimates, which guarantee that the probability of a correct guess is 1/2 only.

[4] A weak order is a transitive and complete binary relation. Completeness here implies reflexivity.

Claim 1 is a special case of this proposition. It deals with two random variables and assumes that $r(X)$ takes two values each with probability 1/2: one on the event $X_1 > X_2$, the other on $X_1 < X_2$. Thus, $r(X)$ is not constant a.s. and therefore it is impossible, by Proposition 1, that both variables are independent on $r(X)$, i.e., on the events $X_1 > X_2$ and $X_1 < X_2$.

We prove a result which is stronger than Proposition 1, and give an upper bound on the number of variables that can be independent of $r(X)$. We associate with X a graph $G = (V, E)$ with set of vertices $V = \{1, \dots, n\}$. A pair (i, j) is an edges in E iff $r(X_i, X_j)$ is not constant almost surely. A *vertex cover* for G is a set $C \subseteq V$, such that each edge in E has at least one end in C. A *minimal* vertex cover is one with minimal number of nodes. We denote this number by $m(X)$.

Proposition 2. *For any vector of random variables $X = (X_1, \dots, X_n)$ at most $n - m(X)$ of these variables are independent of $r(X)$.*

This proposition implies the previous one. Indeed, if for each i, X_i and $r(X)$ are independent then $m(X) = 0$. Thus, $E = \emptyset$. Therefore, for each $i \neq j$, $r(X_i, X_j)$ is constant a.s., which means that $r(X)$ is constant a.s..

Note, that if $r(X_i, X_j)$ is not constant a.s. for any pair $i \neq j$, then G is a complete graph, $m(X) = n - 1$, and thus there can be at most one of the n variables which is independent on $r(X)$.

Proof of Proposition 2. We first prove the proposition for $n = 2$. In this case we need to show that if $r(X)$ is not constant a.s., then at least one of the variables depends on $r(X)$.

Let $D = \{x \mid x_1 = x_2\}$ be the diagonal of $\mathbb{R}^2$, $A = \{x \mid x_1 < x_2\}$ the set above the diagonal and $B = \{x \mid x_1 > x_2\}$ the set below it. Denote by P be the probability distribution on $\mathbb{R}^2$ induced by X.

Suppose that $r(X)$ is not constant a.s.. Hence, at least two of the sets A, D and B have positive probability. Assume that $P(B) > 0$ and $P(A \cup D) > 0$, and denote $C = A \cup D$ (if $P(B) = 0$ then $P(A) > 0$ and $P(B \cup D) > 0$ and the proof is similar). Suppose, now, that X_i and $r(X)$ are independent for $i = 1, 2$.

Fix a point $a \in R$ and let $H_1 (= H_1^a)$ and $H_2 (= H_2^a)$ be two halves of the plane, $H_1 = \{x \mid x_1 \geq a\}$ and $H_2 = \{x \mid x_2 \geq a\}$. Note that $B \cap H_1$ can be written as the disjoint union $(B \cap H_2) \cup (H_1 \setminus H_2)$. Thus $P(H_1 \setminus H_2) = P(B \cap H_1) - P(B \cap H_2)$. By the independence assumption

$$P(H_1 \setminus H_2) = P(B)P(H_1) - P(B)P(H_2).$$

Analogously,

$$P(H_2 \setminus H_1) = P(C)P(H_2) - P(C)P(H_1).$$

Multiplying the first equality by $P(C)$, the second by $P(B)$ and adding them yields, $P(C)P(H_1 \setminus H_2) + P(B)P(H_2 \setminus H_1) = 0$. As $P(C)$ and $P(B)$ are positive, this implies that $P(H_1 \setminus H_2) = P(H_2 \setminus H_1) = 0$. Thus, the set $\cup_a (H_1^a \setminus H_2^a) \cup (H_2^a \setminus H_1^a)$ where a ranges over all rational numbers has probability zero.[5] But this set is just $A \cup B$, contrary to our assumption that $P(A \cup B) > 0$.

Assume now that $n > 2$. Note that the algebra of events generated by $r(X_i, X_j)$, for $i \neq j$, is contained in the algebra generated by $r(X)$. For each edge $(i, j) \in E$, $r(X_i, X_j)$ is not constant a.s., and hence, as we have shown, at least one of the

[5] This is the only place in the proof that requires countable additivity of the probability on the sample space.

variables X_i and X_j must depend on $r(X_i, X_j)$, and therefore also on $r(X)$. Thus, the indices of the random variables that depend on $r(X)$ form a vertex cover, and hence there are at least $m(X)$ such variables. □

Nothing is assumed in Proposition 2 about the expectations of the variables. If we assume that the differences $X_i - X_j$ have finite means, then a different proof is possible that makes use of the following lemma.

Lemma 1. *Let Y be a random variable with a finite mean. If Y and* $\mathrm{sgn}(Y)$ *are uncorrelated then* $\mathrm{sgn}(Y)$ *is constant almost surely.*

Proof. Let $p = P(Y > 0)$ and $q = P(Y < 0)$. The claim that $\mathrm{sgn}(Y)$ is constant a.s. is equivalent to saying that either $p = 1$ or $q = 1$ or $p = q = 0$. Thus, it is enough to show that if $p < 1$ and $q < 1$ then $p = q = 0$.

Suppose that $p < 1$ and $q < 1$. Then, $|p - q| < 1$. Consider the function $s(Y) = \mathrm{sgn}(Y) - (p - q)$. Obviously, $E(s(Y)) = 0$. As $s(Y)$ and Y are uncorrelated it follows that $E(s(Y)Y) = E(s(Y))E(Y) = 0$. Note that if $Y > 0$ then $s(Y) > 0$, and if $Y < 0$ then $s(Y) < 0$. Thus the product $s(Y)Y$ vanishes only when $Y = 0$, and is positive otherwise. Since the expectation of the product is zero, it must be the case that $Y = 0$ a.s., that is $p = q = 0$ as was required to show. □

This lemma implies Proposition 2 in the case that each of the difference $X_i - X_j$ has a finite mean. As we showed above, it is enough to prove the proposition for $n = 2$. Let $Y = X_1 - X_2$. Then, $r(X_1, X_2)$ and $\mathrm{sgn}(Y)$ partition the sample space into the same three events. Thus if X_i is independent of $r(X_1, X_2)$ it is independent of $\mathrm{sgn}(Y)$ and thus uncorrelated with it. Since the covariance is linear in each argument it follows that Y and $\mathrm{sgn}(Y)$ are also uncorrelated. Thus, by Lemma 1 $r(X_1, X_2)$ is constant a.s.

References

Blachman, N. M. and D. M. Kilgour (2001). Elusive optimality in the box problem, *Mathematics Magazine,* 68, 27-34.

Blackwell, D. (1951). On the translation parameter problem for discrete variable, *Ann. Math. Stat.*, 22, 391-399.

Brams, S. J. and D. M. Kilgour (1995). The box problem: to switch or not to switch?, *Mathematics Magazine,* 74, 171-181.

Cover T., (1987). Pick the largest number, in *Open Problems in Communication and Computation,* T. Cover and B. Gopinath, eds., Springer Verlag, p. 152.

Kraitchik, M., (1953). *Mathematical Recreations*, 2nd ed., Dover, New York.

Nalebuff, B., (1989). The other person's envelope is always greener, *Journal of Economic Perspectives,* 3, 171-181.

Littlewood, J. E., (1986). *Littlewood's Miscellany*, B. Bollobas, ed., Cambridge University Press, Cambridge.

Schuss, Z., (2000). To switch or not to switch, this is the question!, unpublished manuscript.

Winkler, P., (2001), Games people don't play, in *Puzzlers Tribute: A Feast for the Mind*, D. Wolfe, T. Rodgers eds., AK Peters.