PAIRED-UNIFORM SCORING: IMPLEMENTING A BINARIZED SCORING RULE WITH NON-MATHEMATICAL LANGUAGE

ALISTAIR J. WILSON AND EMANUEL VESPA

ABSTRACT. We outline a mechanism for eliciting probabilities using two uniform random numbers that is equivalent to the binarized scoring rule (BSR). Though our implementation is simple to describe and has a non-mathematical explanation, it retains the desirable theoretical features of the BSR. Moreover, we show that a discretized version with evenly-spaced reporting intervals can be implemented in the field with no more equipment than a pair of dice.

The literature on incentivized elicitations of probabilistic beliefs has provided a number of mechanisms to practitioners, where a motivating desiderata has been the development of mechanisms that are incentive compatible for general risk preferences. Early proposals utilized payments with quadratic monetary loss for inaccuracy in the stated belief (Brier, 1950) but were only incentive compatible for risk-neutral respondents. More-recent papers have formalized mechanisms for general risk preferences, such as Hossain and Okui (2013)'s binarized scoring rule (BSR).[1]

While experimental studies have made substantial use of incentive-compatible elicitations—see Nyarko and Schotter (2002) for an example, Schotter and Trevino (2014) and Schlag et al. (2015) for surveys—their use in the field is far less common (Manski, 2004). One reason for this is the complex task of explaining to respondents how incentive-compatible mechanisms map reports into payoffs. Explanations can take substantial time and mathematical literacy. In this short paper, we outline an equivalent mechanism to the BSR for eliciting probabilistic beliefs over a verifiable true/false outcome. Our implementation has the advantage that it can be quickly articulated, does not require more-complex probability distributions than the uniform, and does not make use of mathematical language.

We start by providing two illustrations of our implementation. In the first, representative of a common laboratory experiment, subjects are asked to report a belief over a continuous $[0,1]$ interval. In the second, we illustrate how the mechanism can be used in the field to elicit probabilistic beliefs with two dice.

Illustration 1: Bayesian Updating. There are two bags R and Y each containing three chips, where one of the two bags is selected with equal probability. Two-thirds of the chips are red (yellow) in bag R (Y) and the remaining third are yellow (red). One chip is randomly drawn from the selected bag and shown to the subjects, who are subsequently asked to state their belief that R was the selected bag. Example instructions for our mechanism are as follows:

Date: November, 2016.

Our thanks to David Danz, Charlie Holt, Matthew Raffensberger, Andrew Schotter, Isabel Trevino, Lise Vesterlund, and Georg Weizsäcker. We credit here Karl Schlag and Joël van der Weele with an independent derivation of the mechanism we describe below; they outline the same rule in a set of uncirculated working notes, which were made available to us via private correspondence. Note: The authors work on a number of projects together and have opted to alternate between being first/second author across projects; this order conveys no information on contribution which was entirely equal. Wilson: Dept. of Economics, University of Pittsburgh; alistair@pitt.edu; Vespa: Dept. of Economics, University of California, Santa Barbara; vespa@ucsb.edu;.

[1]Other mechanisms that are compatible with general risk preferences are Karni (2009) and Holt and Smith (2016).

You can make \$5 by telling us how likely you think it is that bag R has been selected. You can pick any number between 0 and 100. The number you select indicates the chance (out of 100) that R is the selected bag. To determine your payment the computer will randomly draw two numbers. For each draw, all numbers between 0 and 100 (including decimal numbers) are equally likely to be selected. Draws are independent in the sense that the outcome of the first draw in no way affects the outcome of the second draw.

- If the selected bag is R and the number you picked is larger than either of the two draws, you will get \$5.
- If the selected bag is Y and the number you picked is smaller than either of the two draws, you will get \$5.

Illustration 2: Field Measurement of Beliefs. Respondents are surveyed on their beliefs that the previous month's unemployment rate is lower than a year ago. Instead of reporting the belief directly, the discretized implementation divides beliefs into a set of equally spaced intervals. Example instructions are as follows:

You can make \$5 by telling us how likely you think it is that the unemployment rate last month was higher than it was a year ago. Please circle the interval that best matches your assessment:

Likelihood (%):	0–5	5–15	15-25	25–35	35–45	45–55	55–65	65–75	75–85	85–95	95-100
Interval Value:	0	1	2	3	4	5	6	7	8	9	10

Your choice indicates the likelihood (expressed as a chance in a 100) that you place on the last month's unemployment rate being higher than it was a year ago. We will assess the accuracy of your answer using data from the Bureau of Labor Statistics, and determine your payment using two 10-sided dice.

- If the unemployment rate last month was higher than it was a year ago and the interval you picked has a value greater than or equal to either of the two die rolls, you will get \$5.
- If the unemployment rate last month was lower than it was a year ago and the interval you picked has a value lower than either of the two die rolls, you will get \$5.

Equivalence to BSR. Hossain and Okui (2013) outline the more-general usefulness of the BSR (three or more outcomes, expectations, etc.), we focus here on the special case of a binary event for which the BSR is equivalent to the mechanism outlined by Allen (1987). Though less general, this special case is particularly common in applications.[2]

Our mechanism elicits a probabilistic belief over a verifiable binary outcome $E \in \{\text{True}, \text{False}\}$, where we assume the respondent possesses a true belief $p = \Pr\{E = \text{True}\}$. To elicit p the mechanism uses two monetary prizes A and B for payment (where $A > B \geq 0$), and two iid draws $X_1, X_2 \sim U[0, 1]$ to determine the outcome. Prizes are assigned as follows:

Mechanism Statement 1 (Paired Uniform)**.** *If E is true, you get the prize A so long as q is **greater** than at **least one** of the two uniform draws. If E is false, you get the prize A so long as q is **less***

[2]Implementing the more-general problem with K possible outcomes is certainly possible with a paired uniform rule (see the concluding remarks in Allen, 1987); however, the explanation does becomes more complex. Given our focus on simplicity of explanation, we omit this generalization.

than ***at least one*** *of the two uniform draws. If you do not get the A prize, you are instead assigned the B prize.*

The mechanism can therefore be written through the three cases as:

$$\text{Prize}(q, E, X_1, X_2) = \begin{cases} A & \text{if } E = \text{True and } q > \min\{X_1, X_2\}, \\ A & \text{if } E = \text{False and } q < \max\{X_1, X_2\}, \\ B & \text{otherwise.} \end{cases}$$

To show that it is incentive compatible to truthfully report, we need to calculate the probability that q is greater than at least one of the two draws, and lower than at least one of two draws. Respectively, these probabilities are $(1-(1-q)^2)$ and $(1-q^2)$. Given the true belief p, the probability of winning the better prize A is given by

$$\pi(p, q) = p \cdot \left(1-(1-q)^2\right) + (1-p) \cdot \left(1-q^2\right),$$

and the effective lottery is $\mathcal{L}(q\,|p) = \pi(p,q) \cdot A \oplus (1-\pi(p,q)) \cdot B$. The best response to the belief is $q^\star(p) = p$ as $\mathcal{L}(q^\star(p)\,|p)$ stochastically dominates any other available lottery $\mathcal{L}(q\,|p)$.[3]

The BSR uses a single uniform draw that is compared to $(1-q)^2$ when the event is true, and to q^2 when it is false, via the following:

Mechanism Statement 2 (BSR). *If E is true, you get the prize A so long as $(1-q)^2$ is less than the uniform draw. If the event E is false, you get the prize A so long as q^2 is less than the uniform draw. If you do not get the A prize you are instead assigned the B prize.*

The BSR mechanism produces an identical prize lottery $\mathcal{L}(q\,|p)$ to our paired-uniform mechanism for all p and q. This is so because the probability of winning the larger prize is $(1-(1-q)^2)$ and $(1-q^2)$ when the event is true or false, respectively. Though equivalent over outcome lotteries, the BSR instructions require the provision of a formula with quadratic terms to respondents in order to articulate how the mechanism awards the prizes.

Another equivalent mechanism suggested by Allen (1987) determines the probability of winning the A prize by drawing random numbers from a non-uniform distribution. The mechanism draws a random variable Y from a linearly decreasing density $f_Y(y) = 2 \cdot (1-y)$ and a random variable Z with a linearly increasing density $f_Z(z) = 2 \cdot z$ (both with a support of $[0,1]$, and where the resulting CDFs are $F_Y(y)$ and $F_Z(z)$). The mechanism is stated as follows:

Mechanism Statement 3 (Allen, 1987). *If E is true, you get the prize A so long as q is larger than a draw from the density $f_Y(y) = 2 \cdot (1-y)$. If the event E is false, you get the prize A so long as q is smaller than a draw from the density $f_Z(z) = 2 \cdot z$. If you do not get the A prize, you are instead assigned the B prize.*

It is similarly easy to show that the Allen (1987) implementation yields the same reduced lottery $\mathcal{L}(q\,|p)$, as the win probabilities are defined by $1 - F_Y(q)$ and $F_Z(q)$ for true and false outcomes, respectively. Though this process has a similar description of payments to our mechanism, and three easy to parse cases ($E =$ True and $q > Y$; $E =$ False and $q < Z$; neither), describing the distributions for Y and Z directly requires longer instructions.[4] One interpretation for the paired-uniform mechanism is therefore as an implementation of the Allen density functions through order

[3] Here the decision maker's preference $(\succeq, \Delta\{A,B\})$ needs to satisfy two meaningful restrictions: (i) independence over the set of simple lotteries $\Delta\{A,B\}$; and (ii) if $\mathcal{L}_1$ stochastically dominates $\mathcal{L}_2$ then $\mathcal{L}_1 \succ \mathcal{L}_2$.

[4] For a laboratory implementation of these densities see Ngangoue and Weizsäcker (2015).

statistics: the random variable $\min\{X_1, X_2\}$ has the same density as Y, while $\max\{X_1, X_2\}$ has the same density as Z.

Interval Reporting with Die Rolls. The equivalence between the paired-uniform mechanism and the BSR/Allen mechanisms is shown above for a continuous elicitation. While analytically tractable, and capable of eliciting beliefs to an arbitrary degree of resolution, many implementations will discretize the action space. To that end, we now show that the procedure can be modified to use two fair die-rolls, and that the paired uniform procedure has another desirable property: it can be described as an elicitation over a series of equally spaced intervals.

Rather than eliciting particular point beliefs, when discretized the process partitions the set of feasible beliefs (the unit interval) into $N+1$ sub-intervals, $\mathcal{Q}_N(\boldsymbol{\theta}) = \big\{[0, \theta_1), [\theta_1, \theta_2), \ldots, [\theta_{N-1}, \theta_N), [\theta_N, 1]\big\}$. Defining $\theta_0 = 0$ and $\theta_{N+1} = 1$, we enumerate the intervals as $\kappa([\theta_k, \theta_{k+1})) = k$. Given the interval value assignment $\kappa : \mathcal{Q}_N(\boldsymbol{\theta}) \to \{0, 1, \ldots, N\}$, the mechanism asks participants to report $Q \in \mathcal{Q}_N$, with outcomes determined through the verifiable event *E* and two discrete *iid* uniform draws X_1 and X_2 over $\{1, \ldots, N\}$ as:

$$\text{Prize}_N(Q, E, X_1, X_2) = \begin{cases} A & \text{if } E = \text{True and } \kappa_N(Q) \geq \min\{X_1, X_2\}, \\ A & \text{if } E = \text{False and } \kappa_N(Q) < \max\{X_1, X_2\}, \\ B & \text{otherwise.} \end{cases}$$

A reported interval Q therefore leads to the lottery $\mathcal{L}(Q\,|p) = \pi(p, Q) \cdot A \oplus (1 - \pi(p, Q)) \cdot B$ for a respondent with belief p, where the probability of the larger A prize is

$$\pi(p, Q) = p \cdot \left(1 - \left(1 - \tfrac{\kappa(Q)}{N}\right)^2\right) + (1 - p) \cdot \left(1 - \left(\tfrac{\kappa(Q)}{N}\right)^2\right).$$

An interval partition $\mathcal{Q}_N(\boldsymbol{\theta})$ is truthfully implementable if for every $Q \in \mathcal{Q}_N(\boldsymbol{\theta})$ and every belief $p \in Q$, the lottery $\mathcal{L}(Q\,|p)$ weakly stochastically dominates all other choices. The discrete paired-uniform rule has a unique truthfully implementable division $\mathcal{Q}_N(\boldsymbol{\theta}^\star)$ defined by $\theta^\star_j = \frac{1}{2N}(1 + 2(j-1))$.

For a ten-sided die, the procedure produces intervals as in our second illustration: two terminal-intervals of width 5 percent, and nine interior intervals of width 10 percent. Using this interval structure, an incentivized elicitation can be carried out where an interviewer can state to respondents that whichever interval their true belief lies in, they will maximize their chance of winning the prize by reporting that interval.

Conclusion. Incentivized elicitations, both for the laboratory and the field, will ideally have a number of properties. The literature has focused on incentive compatibility, and a number of novel mechanisms have been developed. However, alongside incentive compatibility, another desirable feature is that the implementation does not rely on mathematical language to explain how reports are mapped to payoffs. Our paper outlines an equivalent mechanism to binarized quadratic scoring for simple true/false events, and so is incentive compatible for fairly general risk preferences. However, the advantage of the paired-uniform rule is that is can be described without mathematical language, and implemented with field equipment as minimal as a pencil and a pair of dice.

References

Allen, Franklin, "Discovering Personal Probabilities When Utility Functions are Unknown," *Management Science*, 1987, *33* (4), 542–544.

Brier, Glenn W, “Verification of forecasts expressed in terms of probability,” *Monthly Weather Review*, 1950, *78* (1), 1–3.

Holt, Charles A and Angela M Smith, “Belief Elicitation with a Synchronized Lottery Choice Menu That Is Invariant to Risk Attitudes,” *American Economic Journal: Microeconomics*, 2016, *8* (1), 110–139.

Hossain, Tanjim and Ryo Okui, “The binarized scoring rule,” *Review of Economic Studies*, 2013, *80* (3), 984–1001.

Karni, Edi, “A mechanism for eliciting probabilities,” *Econometrica*, 2009, *77* (2), 603–606.

Manski, Charles F, “Measuring expectations,” *Econometrica*, 2004, *72* (5), 1329–1376.

Ngangoue, Kathleen and Georg Weizsäcker, “Learning from unrealized versus realized prices,” June 2015. DIW Berlin discussion paper.

Nyarko, Yaw and Andrew Schotter, “An experimental study of belief learning using elicited beliefs,” *Econometrica*, 2002, *70* (3), 971–1005.

Schlag, Karl H, James Tremewan, and Joël J Van der Weele, “A penny for your thoughts: a survey of methods for eliciting beliefs,” *Experimental Economics*, 2015, *18* (3), 457–490.

Schotter, Andrew and Isabel Trevino, “Belief elicitation in the laboratory,” *Annual Review of Economics*, 2014, *6* (1), 103–128.