SUPPLEMENT TO "A DIFFERENTIAL APPROACH TO ANCHORING AND ADJUSTMENT FOR BINARY LOTTERIES"

TSAHI HALYO

1. Alternate Derivation of Beta Heuristic

In this supplement we present an alternate derivation of the Beta heuristics and illustrate some of its relevant properties.

Having defined a subject's perception of the lottery amount, s_l, we may define a subject's subjective probability by:

$$\pi(p) := \frac{\partial c}{\partial s_l} \tag{1}$$

and derive the Beta heuristic through the properties, π, is expected to satisfy from first principles.

1.1. **The Certainty Function.** We begin by defining a certainty function that reflects an experimental subjects' certainty and understanding of a given component of a decision. In the simple lotteries that we will deal with, we are concerned with the certainty/uncertainty that is caused by their understanding (or lack thereof) of the lottery probability and will denote it by $C(p) \in C^{\infty}([0,1])$. It is important to note that our usage of certainty is broader than usual, since various different psychological sources can contribute to it. For instance, the certainty function doesn't distinguish uncertainty caused by a lack of knowledge (e.g. a person who feels uncomfortable with providing a certainty equivalent because he hasn't learned about expected values) and uncertainty that stems from lack of mental effort (e.g. a math PhD. student who is just speeding through a psychology experiment).

The certainty function is a subjective unobservable quantity and accordingly can not be compared between various individuals. In other words, each person perceives their certainty in their own psychological units of uncertainty. What *can* be compared are unitless ratios of certainties. For example, if person A has certainty function $C_A(p)$ and person B has certainty function $C_B(p)$ then we can't meaningfully compare C_A and C_B; but, we can compare the unitless quantities $C_A(p_1)/C_A(p_2)$ and $C_B(p_1)/C_B(p_2)$.

1.2. **The Risk Tolerance.** Assume that the change in risk tolerance given the change $p \to p + dp$ is proportional to the degree of certainty the person feels at probability, p. We summarize this as follows:

Postulate 1. *The more comfortable a person feels with their understanding of the situation at a given probability the more willing they are to adjust the certainty equivalent given a* *small change in the lottery probability. Stated formally, we claim the certainty function satisfies:*

$$d\pi \propto C(p)dp \tag{2}$$

with boundary conditions, $\pi(0) = 0, \pi(1) = 1$.

where the boundary conditions follow from the fact that a certain gain of \$$l$ is worth \$$l$ and a certain receipt of nothing is worthless. Intuitively, we may consider the boundary conditions as a requirement to normalize the certainty function so that it may be compared between individuals as a unitless quantity.

1.3. **Certainty Functions & the Fisher Information.** A probability, p, is intuitively seen as a biased coin toss with probability, p, of landing on heads. The fraction, p, is intuited as the fraction of heads in a reasonable number of tosses due to the Law of Large numbers. However, the stochasticity, that is, the error about the mean gives rise to our feelings of certainty and uncertainty. We will consider two possible ways the mind may encode the error: the variance and the standard deviation. Let us further assume that the certainty function is inversely proportional to the error about the mean. If we denote the Fisher information by $\mathcal{I}(p)$, this may then be expressed as:

$$\begin{aligned} C(p) &\propto \mathcal{I}(p) \quad \text{error as variance} \\ C(p) &\propto \sqrt{\mathcal{I}(p)} \quad \text{error as std. dev.} \end{aligned} \tag{3}$$

1.4. **Risk Tolerance when $C(p) \propto \mathcal{I}(p)$.** We begin with the case in which the error is encoded as the variance so that

$$C(p) \propto \mathcal{I}(p) = 1/\sigma^2(p) = p^{-1}(1-p)^{-1} \tag{4}$$

Mathematically, though, we run into the difficulty of finding a suitable normalization constant since the integral diverges. We therefore follow Haldane and find the improper solution of the form

$$C(p) = 0.5\delta(p) + 0.5\delta(p-1) \tag{5}$$

in which $\delta(x)$ is the Dirac delta function. This yields the solution,

$$\pi(p) = \begin{cases} 0, & p = 0 \\ 1/2, & 0 < p < 1 \\ 1, & p = 1 \end{cases} \tag{6}$$

which describes the primitive intuition of a person that can only categorize probability into three basic states: certainty ($p = 1$); impossibility ($p = 0$) and uncertainty ($p \in (0, 1)$). Each state is mapped to its corresponding risk tolerance: with certainty naturally being mapped to one and impossibility to zero and the state of uncertainty is mapped to half - as expected by one who doesn't grasp probability. This captures the intuition that if a person doesn't understand probability, when faced with a lottery with $p \in (0, 1)$ they act as if they don't know the value of p i.e. act without any prior information, one expect's a coin with unknown bias to act like a fair coin.

1.5. **Risk Tolerance when $C(p) \propto \sqrt{\mathcal{I}(p)}$.** A person whose certainty is proportional to the square root of the Fisher information, that is, inversely proportional to the standard deviation

$$C(p) \propto \sqrt{\mathcal{I}(p)} = 1/\sigma(p) = p^{-1/2}(1-p)^{-1/2} \tag{7}$$

is more sophisticated. In this case the risk tolerance takes the form of

$$\pi(p) = \frac{2}{\pi} \sin^{-1}(\sqrt{p}) \tag{8}$$

which is the arcsine distribution. The value of choosing $C(p) \propto \sqrt{\mathcal{I}(p)}$, is due to the invariance it provides the risk tolerance under a change of coordinates. To clarify the issue, imagine a mental encoding, $f : p \mapsto t$, that maps probabilities to some other values. For instance, consider the encoding to log-odds. Given a certainty function, $C_p(p)$ on the space of probabilities, we're interested in how to choose a certainty function on the transformed space. Naive assignment by the pullback, $C_t = C_p \circ f^{-1}$, creates issues since it will not satisfy $\pi_p = \pi_t \circ f$, where $\pi_t := \int C_t dt$. This would force us to understand the risk tolerance as operating on the privileged coordinate system of the probabilities - something we'd like to avoid given the lack of knowledge about the encoding mechanism. Generally, our issue stems from the fact that

$$dp = d(f^{-1}(t)) = dt/(f' \circ f^{-1})(t) \tag{9}$$

so that coordinate invariance

$$\pi_p = \pi_t \circ f \tag{10}$$

requires that the certainty functions transform according to the rule

$$C_p(p) = C_t(t)(f' \circ f^{-1})(t) \tag{11}$$

To avoid this issue we would like a principle that defines C_t on all coordinate transformations such that: (1) the risk tolerance is preserved under a change in coordinates; (2) no system of coordinates can be meaningfully described as the proper system of coordinates.

Proposition 1. *Let a person encode probabilities by the change of coordinates $T \in C^2([0,1])$ and have $C(p) \propto \sqrt{\mathcal{I}(p)}$, then the risk tolerance is invariant under the change to encoded probabilities. Stated formally:*

$$\pi_t \circ T = \pi_p \tag{12}$$

So that the risk tolerances based on the encoded values, $\pi_t(y) := \int_{t_0}^{y} C_t dt$, and probabilies, $\pi_p(x) := \int_0^x C_p dp$, are equivalent.

Proof. Recall that a probability, p, is understood to map to a biased coin with probability, p. Let $t = T(p)$. Accordingly, this coin has Fisher informations derived from the models in terms of p, $\mathcal{I}_p(p)$, and t, $\mathcal{I}_t(t)$, that satisfy

$$\mathcal{I}_p(p) = \mathcal{I}_t(T(p))|T'(p)|^2 \tag{13}$$

so that we have

$$C_p(p) = (C_t \circ T)(p)T'(p) \tag{14}$$

Accordingly, the risk tolerance is found to satisfy

$$\pi_p(x) = \int_0^x C_p(p)dp \tag{15}$$

$$= \int_0^x (C_t \circ T)(p)T'(p)dp \tag{16}$$

$$= \int_{T(0)}^{T(x)} C_t(t)(T' \circ T^{-1})(t)\frac{dt}{(T' \circ T^{-1})(t)} \tag{17}$$

$$= \int_{T(0)}^{T(x)} C_t(t)dt \tag{18}$$

$$= (\pi_t \circ T)(x) \tag{19}$$

if we note that $t_0 = T(0)$ is required as the lower bound of the domain of integration. □

The value of this proposition is as follows: presently, we do not know the exact mechanism through which numerical probabilities are encoded in the brain. It may be linear as most would have it; logarithmic in line with our exposition concerning base-10 numbers; or be log-odds as Gonzalez & Wu advocate for. The proposition above demonstrates that if the certainty function is the root of the Fisher information then the encoding mechanism is irrelevant to the lottery valuation. Furthermore, by creating a system in which the risk tolerance is equivariant under changes in encoding mechanism, we can understand the risk tolerance as the fundamental subjective measure of probability unbiased by the mental encoding heuristic that they are using.

1.6. **The linear case revisited.** In light of this framework, we may now provide a new explanation of linear adjustment, i.e. $d\pi \propto dp$, as the case in which the certainty function is constant. Since we have the initial conditions $\pi(0) = 0, \pi(1) = 1$, linear adjustment has the unique solution

$$d\pi \propto dp;\ \pi(0) = 0, \pi(1) = 1 \rightarrow \pi = p \tag{20}$$

which implies $V(p, l)$ is the expected utility. In other words, a person who values lotteries as their expected utility using a formula is equally certain/uncertain at each probability since by employing a computational rule each probability is treated the same way.

1.7. **Biases in the Certainty function.** In each of the three cases we dealt with above we found that the solution, π, satisfied a skew-symmetry[1] about $p = 0.5$. However, people in general may be biased towards either certainty ($p = 1$) and/or impossibility ($p = 0$) reflecting their understanding and comfort with that outcome. To accommodate this, we consider a base unbiased certainty function that is scaled by p^α to capture a bias towards certainty and by $(1 - p)^\beta$ to capture a bias towards impossibility with $\alpha, \beta > 0$ and the

[1] More precisely, $\pi(p) - 0.5$ is skew-symmetric about $p = 0.5$.

reverse when they are negative. Formally, we claim that the bias may be captured by generalizing the uncertainty function as:

$$C(p) \propto p^{\alpha-b}(1-p)^{\beta-b} \tag{21}$$

in which $b = 1$, $1/2$, or 0 depending on whether the person's base unbiased certainty function is proportional to $\mathcal{I}, \sqrt{\mathcal{I}}$, or 1. If we normalize the biased function we may see that it is the familiar Beta density. In other words,

Proposition 2. *The general form of the risk tolerance is*

$$\pi(p) = F_B(p; \alpha+1-b, \beta+1-b) \tag{22}$$

in which F_B is the CDF of a Beta distribution.

The three certainty functions we explored in detail all fit within this Beta paradigm since

$$C(p) \propto \mathcal{I} \to \pi = F_B(p,0,0), \quad C(p) \propto \sqrt{\mathcal{I}} \to \pi = F_B(p,0.5,0.5), \quad C(p) \propto dp \to \pi = F_B(p,1,1) \tag{23}$$

2. Linear Approximations to the Beta Theory

The model sketched out suffers from a major practical flaw insofar as it relies upon the use of the incomplete Beta function ($F_B(x,\alpha,\beta)$) - a function notorious for its intractability. Here we sketch out the role of linear approximations in the estimation of this model. As in the earlier section, we plan on taking a linear expansion about $p = 0.5$. We do so, since a Taylor expansion of $F_B(x,\alpha,\beta)$ finds for $\epsilon = \beta - \alpha$ that

$$B(\alpha,\alpha+\epsilon)\left(F_B(x,\alpha,\alpha+\epsilon) - F_B(0.5,\alpha,\alpha+\epsilon)\right) \approx \left(\frac{1}{2}\right)^{2\alpha+\epsilon-2}\left((x-0.5)-\epsilon(x-0.5)^2\right) \tag{24}$$

so that for small ϵ a linear approximation works well away from the boundary points.

Having found the slope (see the Taylor expansion above) we need to find the value of $F(0.5,\alpha,\beta)$. We begin with the change of variables $x \mapsto 1-x$ in the integral that yields:

$$F_B(x,\alpha,\beta) = 1 - F_B(1-x,\beta,\alpha) \tag{25}$$

providing us with

$$F_B(0.5,\alpha,\alpha) = 0.5 \tag{26}$$

More generally, for $\alpha, \beta < 1$ we find that $F(0.5,\alpha,\beta)$ is approximately a function of the ratio $r = \alpha/\beta$. When the ratio takes reasonable values $0.25 < r < 4$ we find that the following approximation holds.

$$F_B(0.5, r\beta, \beta) \approx 0.2 + 0.8e^{-r}, \quad 0.25 < r < 4 \tag{27}$$

so that we may approximate π for the interior probabilities by:

$$\pi(p) = F_B(x,\alpha,\beta) \approx 0.2 + 0.8e^{-\alpha/\beta} + \frac{2^{2-\alpha-\beta}}{B(\alpha,\beta)}(p-0.5) \tag{28}$$

Before proceeding to the next section, it is worth highlighting a few facts: (1) the constant term is only a function of the ratio; (2) the slope is symmetric in α, β;