Probability Weighting

Mark Dean

Lecture Notes for Spring 2015 Behavioral Economics - Brown University

1 Introduction

Last lecture we introduced the Allais Paradox and showed that it was incommensurate with the expected utility model of choice under risk. It seems that if our aim is for a descriptive model of risky choice, we are going to need to adapt this model in some way. Many interesting adaptations have been proposed, for example involving the concept of disappointment[1] or salience[2]. In this lecture we are going to consider the most widespread and straightforward modification: Probability weighting.

1.1 Probability Weighting

Let's think back to the Allais Paradox. As a reminder, the standard version of this paradox is stated as

- Lottery a_1: 100% chance of \$16
- Lottery a_2: 10% chance of \$18 , 89% chance of \$16 1% chance of \$0

The second choice is between

- Lottery b_1: 10% chance of \$18, 90% chance of \$0

[1] Gul, Faruk, 1991. "A Theory of Disappointment Aversion," Econometrica, Econometric Society, vol. 59(3), pages 667-86, May.

[2] Pedro Bordalo & Nicola Gennaioli & Andrei Shleifer, 2012. "Salience Theory of Choice Under Risk," The Quarterly Journal of Economics, Oxford University Press, vol. 127(3), pages 1243-1285.

- Lottery b_2: 11% chance of \$16, 89% chance of \$0

with people generally choosing a_1 and b_1.

One natural way of interpreting the paradox is that somehow people do not 'believe' the stated probabilities, or at least, when making decisions they 'weigh' probabilities in a manner that is not quite proportional. For example, many people, when explaining why the exhibit Allais-like behavior, suggest that they really do not like the 1% chance of winning nothing in lottery a_2: this 1% probability weighs heavily in their decisions. In particular, the difference between a 0% chance of getting \$0 (in lottery a_1) and a 1% chance of getting \$0 (in a_2) seems more meaningful that the difference between an 89% chance of getting \$0 (in b_2) and a 90% chance of getting 0 in b_1. This violates the assumption that people's preferences are linear in probabilities, which is at the heart of the expected utility paradigm.

One way of modeling this type of behavior is to allow for a probability weighting function which modifies the weights that different probabilities have. Remember, expected utility requires that there exists a $u : X \to \mathbb{R}$ such that the function

$$U(p) = \sum_{x \in X} p(x)u(x)$$

represents preferences $\succeq$ on $\Delta(X)$. If we allow for a probability weighting function, we would require another function $\pi : [0,1] \to [0,1]$ such that

$$V(p) = \sum_{x \in X} \pi(p(x))u(x)$$

represents $\succeq$.π is the probability weighting function. It takes the true objective probabilities and warps them into what are sometimes called decision weights For example, we could think of a probability weighting function that increases the weight on very low probabilities (so, for example $\pi(0.01) = 0.05$), thus explaining the Allais paradox. If we like, we could demand that these decision weights sum to one and are non negative, and so act like probabilities, but it doesn't really matter. We will call this the 'simple probability weighting model'.

The simple probability weighting model is clearly a generalization of the expected utility model.[3] It should also be obvious that we could pick decision weights that allow for the Allais paradox.

[3] i.e. it allows for everything that the expected utility model allows for (just set $\pi(x) = x$ for all x), and more besides.

So does this make it the right generalization of expected utility? The considered answer of the profession is 'probably not', for two reasons

1. It turns out, that if π is anything other than the identity function, then the simple probability weighting model implies violations of stochastic dominance (assuming preferences are continuous) . By this we mean the following: take two lotteries over monetary outcomes p and q, and let F_p and F_q be the cumulative distribution functions of p and q.[4] We say that p stochastically dominates q if, for every x,

$$F_p(x) \leq F_q(x)$$

   i.e., for any x the probability of getting a prize greater than x is higher under p than under q. If a decision maker can be represented by a simple probability weighting model with a probability weighting function that is not equal to the identity function, then we can find a p and q such that p stochastically dominates q, but q is preferred to p. Unfortunately, the proof is beyond the scope of this course, so you will have to take it on trust, but this is usually taken as a fairly damning criticism of the simple probability weighting model

2. It is not clear that the simple probability weighting model captures quite the decision procedure that we want. One way to think of subjects that fall for the Allais paradox is that they are 'pessimistic': while the objective probability of getting \$0 in a_2 is only 1%, they act as if this probability is larger - in other words they act like they are going to be unlucky. But this implicitly assumes that the 1% probability is upweighted because it is applied to the worst prize. Consider the following thought experiment: First, a decision maker is asked to evaluate a lottery which had a 49% chance of winning \$10, a 49% chance of winning \$0 and a 2% chance of winning \$5. Second, the decision maker is asked to evaluate a lottery which had a 49% chance of winning \$10, a 49% chance of winning \$0 and a 2% chance of losing \$100. Would our agent weigh the 2% chance the same way in both cases? Arguably not. If our decision maker really is pessimistic, then we might expect them to overweight the 2% probability only in the second case. But this is not allowed in the simple probability weighting model described above. 2% is just 2%, and so must always be overweighted or always be underweighted.

[4] i.e. for any $x \in \mathbb{R}$, $F_p(x)$ is the probability of getting an outcome less than x according to lottery p.

Because of these two criticisms, the simple probability weighting model is not often used in economics. Instead, people tend to use what we will call the 'cumulative probability weighting model'. This mechanism was introduced by Quiggin [1985], in which it is called the 'Rank Dependent Utility' model. However, it is also at the heart of the more recent versions of prospect theory, called 'cumulative prospect theory'.

The basic idea of the cumulative probability weighting model is that the probability weighting attached to a particular prize should depend on whether it is a good or bad prize, or in other words on its rank. So in the above thought experiment, the weight of the 2% probability depends on whether it is a good or a bad prize. How is this done in practice? Well, the weight applied to a probability is related to that rank of the prize to which the probability is related (hence rank dependent utility). To get at this, we apply the weights to the *cumulative* probability of the distribution (hence cumulative prospect theory).

For simplicity, we will describe the model for a prize space X that has an obvious ranking (for example X is monetary prizes). Thus we know what it means to rank the prizes in a lottery. In fact, it is easy to extend the model to cases where we do not know what this ranking is (we can just extract it by looking at preferences over degenerate lotteries) but we don't need to do this now.

Definition 1 *A decision maker's preferences $\succeq$ over $\Delta(X)$ can be represented by a cumulative probability weighting model if there exists a utility function $u : X \rightarrow \mathbb{R}$ and a cumulative probability weighting function $\psi : [0,1] \rightarrow [0,1]$ such that $\psi(0) = 0$ and $\psi(1) = 1$, such that the function $U : \Delta(X) \rightarrow \mathbb{R}$ represents $\succeq$, where $U(p)$ is constructed in the following way:*

1. *The prizes of p (i.e. $x \in supp(p)$) are ranked $x_1, x_2, \ldots, x_n$ such that $x_1 \succ x_2 \cdots \succ x_n$*
2. *$U(p)$ is determined as*

$$U(p) = \psi(p_1)u(x_1) + \sum_{i=2}^{n} \left(\psi\left(\sum_{j=1}^{i} p_j \right) - \psi\left(\sum_{k=1}^{i-1} p_k \right) \right) u(x_i)$$

Upon first reading, this can look pretty formidable, but it is fairly intuitive. The utility of lottery p is just a weighted average of the utility of its prizes. However, the weights applied to each prize depend on it's ranking compared to the other prizes that p can give. In each case, the weight of the prize is equal to the weighting of the probability of all prizes at least as good as that prize, minus the weighting of all prizes that are better than that prize. Thus, the weight applied to the best prize (received with probability p_1) is

$$\psi(p_1) - 0$$

as p_1 is the probability of getting a prize at least as good as x_1 and 0 is the probability of getting a prize better than x_1

The weight attached to prize x_2 (received with probability p_2) is given by

$$\psi\left(p_1 + p_2\right) - \psi(p_1)$$

as $p_1 + p_2$ is the probability of getting a prize at least as good as x_2 and p_1 is the probability of getting a prize better than x_2.

The weight attached to prize x_3 is given by

$$\psi(p_1 + p_2 + p_3) - \psi\left(p_1 + p_2\right)$$

and so on.

Is should be obvious, but you should convince yourself that if ψ is the identity function, then this reduces to expected utility.

An example will help to clarify the procedure, and show how this model can explain the Allais paradox. Let us assume that a subject has a utility function for money $u(x) = x$, and has a cumulative probability weighting function $\psi(p(x)) = p(x)^m$. Would this make the decision maker exhibit Allais-type behavior? In order to avoid calibration issues, we will ask: what value of x would make

$$\bar{a}_1 = 1 \times 1,000,000$$

indifferent to the lottery

$$\bar{a}_2 = 0.01 \times 0 + 0.89 \times 1,000,000 + 0.1 \times x$$

and the y that would make

$$\bar{b}_2 = 0.89 \times 0 + 0.11 \times 1,000,000$$

indifferent to

$$b_1 = 0.90 \times 0 + 0.10 \times y$$

Allais - type behavior is indicated by $x > y$

In order to figure this out, note that

- $U(\bar{a}_1) = \psi(1)u(1,000,000) = 1,000,000$
- For lottery $\bar{a}_2$

$$\begin{aligned} U(\bar{a}_2) &= \psi(0.1)x \\ &\quad + (\psi(0.99) - \psi(0.1))\, 1,000,000 \\ &\quad + (\psi(1) - \psi(0.99))\, 0 \end{aligned}$$

(assuming $x > 1,000,000$)

- For lottery $\bar{b}_2$

$$U(\bar{b}_2) = \psi(0.11)1,000,000 + (1 - \psi(0.11))\, 0$$

- and for lottery $\bar{b}_1$

$$U(\bar{b}_1) = \psi(0.10)y + (1 - \psi(0.10))\, 0$$

Thus, if $U(\bar{a}_1) = U(\bar{a}_2)$ it must be that

$$\begin{aligned} & 1,000,000 \\ = \; & \psi(0.1)x \\ & + (\psi(0.99) - \psi(0.1))\, 1,000,000 \end{aligned}$$

implying

$$\begin{aligned} x &= \frac{1 - (\psi(0.99) - \psi(0.1))}{\psi(0.1)} 1,000,000 \\ &\Rightarrow \frac{1 - (0.99^m - 0.1^m)}{0.1^m} 1,000,000 \end{aligned}$$

Similarly, if $U(\bar{b}_1) = U(\bar{b}_2)$ we have

$$\begin{aligned}
\psi(0.11)1,000,000 + (1 - \psi(0.11))\,0 &= \psi(0.10)y + (1 - \psi(0.10))\,0 \\
&\Rightarrow \; y = \frac{\psi(0.11)}{\psi(0.10)} 1,000,000 \\
&= \frac{0.11^m}{0.10^m} 1,000,000
\end{aligned}$$

So, for Allais type behavior, we require that

$$\begin{aligned}
\frac{0.11^m}{0.10^m} &< \frac{1 - (0.99^m - 0.1^m)}{0.1^m} \\
&\Rightarrow \; 0.99^m - 0.1^m < 1^m - 0.11^m
\end{aligned}$$

In order to get some idea when this might be, let's try a couple of different values for m. First, for $m = 2$ we have

$$0.99^m - 0.1^m \approx 0.97$$

while

$$1^m - 0.11^m \approx 0.99$$

So we do get the Allais effect. For $m = 1$

$$0.99^m - 0.1^m = 0.89$$

and

$$1^m - 0.11^m = 0.89$$

So independence holds (as we would expect).

For $m = 0.5$ we have

$$0.99^m - 0.1^m \approx 0.68$$

while

$$1^m - 0.11^m \approx 0.67$$

So we get the *opposite* of the Allais effect.

This shows us that the cumulative probability weighting model is consistent with both the Allais paradox and it's opposite, depending on the shape of the weighting function. So what type of cumulative probability weighting functions are consistent with the common consequence effect (i.e. the Allais paradox)? The clue is in the example above. While we will not prove it, it can be shown that a probability weighting function implies a generalized form of the common consequence at all probabilities if and only if it is convex. Such weighting functions are also thought of as 'pessimistic', as they put more weight on low-probability outcomes.

However, weighting functions that are concave everywhere are not popular for other reasons. For one thing, convexity does not necessarily guarantee the common ratio effect (as you will see for homework). In fact, you can have concave probability weighting functions that give the opposite of the common ratio effect. For another, in order to generate the 'four fold pattern of risk' that we described previously, the probability weighting function needs intersect the diagonal from above.[5] For these reasons, the probability weighting function is usually assumed to be 'S' shaped - initially concave, then convex - so small probabilities of winning are overweighted, while large probabilities are underweighted. S shaped functions also embody the principle of diminishing sensitivity: a change in probability of 0.1 has a bigger impact when it changes the probability of winning from 0 to 0.1 or 0.9 to 1 than when it changes it from 0.6 to 0.7.

Prelec [1998] comes up with a set of axioms that guarantee that the weighting function has the following parametric form:

$$w(p) = \exp(-(-\ln(p)^{\alpha}))$$

which is often used in practical applications, as it requires the estimation of a single parameter.

[5] risk-seeking over low-probability gains, risk-averse over high-probability gains, risk-averse over low-probability losses, and risk-seeking over high-probability losses.