Risk and Time: Expected Discounted Utility and Generalizations

- Risk: objective probability distributions. (*can be extended to uncertainty.*)

- Risk and Time: the risks in different periods might be correlated or independent. And the agent might get various signals about future risks.

- Expected utility maximizers need to correctly compute conditional probabilities to be dynamically consistent- let's assume for now that they doe (will later discuss evidence about how and when people make mistakes in inference.)

- The agent receives sequences of outcomes/consumptions that lie in *Z;* for simplicity assume *Z* finite.

- Should require that each $z_t : S \to Z$ is measurable w.r.t. the agent's information but won't spell out the formalism using filtrations.

- The simplest, and standard, model here is expected discounted utility (EDU): Fix a distribution *P* on $Z^{\mathcal{T}}$, and let

$$V = \mathbb{E}_P \sum_{t \in \mathcal{T}} \delta^t u(z_t) \ .$$

Some implications of EDU

1. *u* more concave than *u'* implies that *u* has both more static risk aversion and more preference for consumption smoothing. This link makes it hard to match the representation to some data, which has led to the use of alternatives such as Epstein-Zin preferences.

2. EDU is "indifferent to autocorrelation." For example, in a 2-period model EDU is indifferent between $.5\delta_{(10,5)} + .5\delta_{(5,10)}$ and $.5\delta_{(10,10)} + .5\delta_{(5,5)}$, since they both give (.5,.5) of either 5 or 10 in each period.

   But even if deterministic preferences are additively separable, the agent might want to "hedge" consumption risk over time and be "auto-correlation averse." Or the agent might alternately be "autocorrelation-seeking."

3. EDU predicts convex preferences over the timing of consumption. E.g.

$$.5\delta_{(0,z,0,0)} + .5\delta_{(0,0,0,z)} \succ \delta_{(0,0,z,0)} \text{ because } \frac{1}{2}\delta + \frac{1}{2}\delta^3 > \delta^2;$$

More generally δ^k is convex in *k.*

(for the same reason EDU prefers a 50-50lottery between "now" and period 2 to a sure payoff in period 1, but present bias should make that prediction fit better.)

4. With EDU, people only care about information because it helps them make better decisions. But once we relax EDU the agent may have "non-instrumental" reasons to prefer to get information earlier. More on this in a bit.

- For now let's just look at preferences on probability distributions over outcomes, that is, over elements of $\Delta(Z^{\mathcal{T}})$. This is enough to explore points 1 and 2.

- And set $\mathcal{T} = \{0,1,...,T\}$

- To map this to stochastic processes, let the state $s = (z_0, z_1, ..., z_T)$ and let $\Sigma_t = \{(z_0, z_1, ..., z_t)\}$: all possible realizations of the *z's* up through time *t.*

- Note that the space of all lotteries on time paths $\Delta(Z^{\mathcal{T}})$ is a larger space than the product space of lotteries $\left(\Delta(Z)\right)^{\mathcal{T}}$: for example

  $.5\delta_{(10,5)} + .5\delta_{(5,10)}$ and $.5\delta_{(10,10)} + .5\delta_{(5,5)}$ are distinct elements of $\Delta(Z^{\mathcal{T}})$

  that both map to $\left(.5\delta_5 + .5\delta_{10}, .5\delta_5 + .5\delta_{10}\right) \in \left(\Delta(Z)\right)^{\mathcal{T}}$.

- And we don't want to start out with the smaller preference domain $\left(\Delta(Z)\right)^T$, as that imposes indifference to correlation between periods.

- For any $P \in \Delta(Z^{\mathcal{T}})$, let P_t be its marginal distribution in period *t*, and let P_{-t} be the marginal over all other periods.

Preference $\succsim$ on $\Delta(Z^{\mathcal{T}})$ satisfies *indifference to autocorrelation* if $P_t = Q_t$ for all $t \in \mathcal{T}$ implies $P \sim Q$.

For given $\succsim$, let $\succsim^*$ be the induced preference on $\Delta(Z)$:

$$p \succsim^* q \Leftrightarrow \vec{p} \succsim \vec{q} \quad .$$

Definition: Preference $\succsim$ satisfies *monotonicity* if $P_t \succsim^* Q_t \ \forall t \in \mathcal{T} \rightarrow P \succsim Q$.

Note that monotonicity implies indifference to autocorrelation.

<u>Definition:</u> Preference $\succsim$ satisfies *stationarity if* when $P_T = Q_T = P_0^{'} = Q_0^{'}$, $P_t = P_{t+1}^{'}, Q_t = Q_{t+1}^{'}$ for $t = 0,\ldots T-1$, then $P \succsim Q$ iff $P^{'} \succsim Q^{'}$. (*like in Koopmans)*

<u>Example:</u> $P=(a,b,c)$ $Q=(d,e,c)$ $P'=(c,a,b)$, $Q'=(c,d,e)$

$P=(a,b,c)\succsim(d,e,c)=Q$ so $(c,a,b)\succsim(c,d,e)$: (a,b) preferred to (d,e) no matter when these occur.

<u>Definition:</u> $\succsim$ satisfies *mixture continuity* if for all $p,q,r \in \Delta(Z^{\mathcal{T}})$ the sets

$$\{\alpha \in [0,1] : \alpha p + (1-\alpha)q \succsim r\} \text{ and } \{\alpha \in [0,1] : \alpha p + (1-\alpha)q \precsim r\}$$

are closed in [0,1]. (*same as in A-A but on a different domain*).

<u>Definition:</u> $\succsim$ satisfies *independence* if for all $p,q,r \in \Delta(Z^{\mathcal{T}})$ and $\alpha \in (0,1)$, $p \succsim q \Leftrightarrow \alpha p + (1-\alpha)r \succsim \alpha q + (1-\alpha)r$. (*also same as before*).

Theorem (Fishburn [1970]).

1. Suppose that $\mathcal{T}$ is finite, and $\succsim$ satisfies mixture continuity. Then $\succsim$ on $\Delta(Z^{T})$ is represented by $U : Z^{\mathcal{T}} \to \mathbb{R}$ (e.g. the preference is the same as maximizing $E_P U$) iff it satisfies independence.

   This is all we get from the EU axioms; this general form isn't very tractable.

2. The preference satisfies indifference to auto-correlation in addition to the previous iff it is represented by $\sum_{t\in\mathcal{T}} E_{P_t} u_t(z_t)$.

3. It additionally satifies stationarity and monotonicity iff it is represented by $\sum_{t\in\mathcal{T}} \delta^t E_{P_t} u(z_t)$.

Aside: Note that these preferences satisfy the independence axioms over sequences of outcomes.

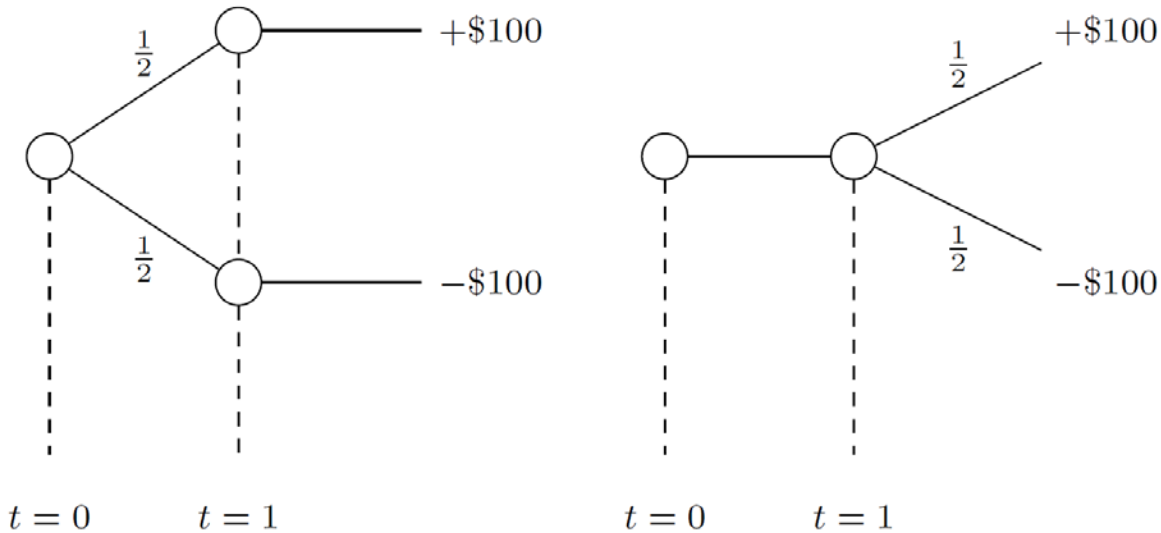
- The “expected discounted utility” representation is easy to work with, but as we saw it links the agent’s preference for consumption smoothing to her static risk aversion.

- So let’s look at some alternatives.

- Kreps-Porteus *Ema* [1978] introduced the idea of “temporal lotteries” to model non-instrumental preferences over the timing of uncertainty.

- Non-instrumental: in a temporal lottery, the agent’s only choice is at the start. But different temporal lotteries can give him more or less advance information about future consumption.

- Temporal lotteries are defined by backwards recursion: $H_T := \Delta(Z), H_t := \Delta\left(Z \times H_{t+1}\right)$: a probability distribution on what you get today and the situation next period.



(a) early resolution (b) late resolution

Here option a) is $\frac{1}{2}\delta_{(0,\delta_{100})}+\frac{1}{2}\delta_{(0,\delta_{-100})}$ and option b) is $\delta_{\left(0,\frac{1}{2}\delta_{100}+\frac{1}{2}\delta_{-100}\right)}$.

One reason information earlier could be better is if the agent could use it to guide her decisions. Here we suppose she can't, but conceivably she still prefers to hear early.

The agent has *Kreps-Porteus preferences* if she maximizes the expected value of the function U_0 which is defined recursively by $U_T = W_T$ and $U_t = W_t(z_t, E_{|} U_{t+1})$, with each W_t bounded and continuous increasing in its second argument.

- Note that if $W_t(z,\gamma) = u(z) + \delta\gamma$ we have EDU.
- More generally $W_t(z,\gamma) = u(z) + \delta(u(z))\gamma$ is Uzawa preferences.
- With KP preferences, the past doesn't matter for how the agent views the future- no history dependence.

- Recall the two choices $\frac{1}{2}\delta_{(0,\delta_{100})} + \frac{1}{2}\delta_{(0,\delta_{-100})}$ vs. $\delta_{\left(0,\frac{1}{2}\delta_{100}+\frac{1}{2}\delta_{-100}\right)}$.

- A KP agent's value for the early-resolving option is

$$.5W_0(0,W_1(100)) + .5W_0(0,W_1(-100)).$$

- For the late resolving lottery it is $W_0\left(0,.5W_1(100)+.5W_1(-100)\right)$.

- So if W_0 is convex in its second argument, the agent prefers the early-resolving lottery.

For any $h_{t+1}, h_{t+1}^{'} \in H_{t+1}$ and any z_t define

$$h = \left(z^{t-1}, \alpha \delta_{(z_t, h_{t+1})} + (1-\alpha) \delta z_{(z_t, h_{t+1}^{'})} \right)$$

and

$$h^{'} = \left(z^{t-1}, z_t, \alpha h_{t+1} + (1-\alpha) h_{t+1}^{'} \right).$$

These two paths have the same outcomes through period *t* and the same lotter over continuations; they differ only in whether the agent learns h_{t+1} in period *t* or period *t*+1.

Preference $\succsim$ displays preference for earlier/later resolution of uncertainty if for all *t* $h_t \succsim h_t^{'}$ or $h_t \precsim h_t^{'}$ iff $h \precsim h^{'}$.

Theorem (Kreps and Porteus) A Kreps-Porteus preference has preference for earlier/later resolution iff W_t is convex/concave for all *t.*

- KP agents reduce compound lotteries that occur "at the same time" but needn't reduce those one period apart. Does this make sense for short period lengths? How long is a period?

- Epstein-Zin preference $V_t(z) = \left((1-\delta)z_t^{\rho} + \delta V_{t+1}(z)^{\rho/\alpha}\right)^{\alpha/\rho}$ is a special case of Kreps-Porteus.

- On deterministic payoff streams this is equivalent to additive discounting with $u(z) = z^{\rho}$, so that $V_t = (1-\delta)z_t^{\rho} + \delta V_{t+1}$.

- Here the agent need not be neutral w.r.t. autocorrelation.

- And the Kreps-Porteus aggregator is $W(z,x) = \left((1-\delta)z_t^{\rho} + \delta x^{\rho/\alpha}\right)^{\alpha/\rho}$ so the agent prefers earlier resolution if $\alpha > \rho$.

- The elasticity of intertemporal substitution is $1/(1-\rho)$, and the coefficient of relative risk aversion for wealth gambles is $1-\alpha$.

- So committing to values for those two tradeoffs pins down how the agent feels about early resolution.

- Bansal and Yaron *J Finance* [2004] use these preferences (and specifications for the process driving stock dividends) to explain the equity premium, low risk-free rate, and the realized volatilities of market return, risk free rate, and the price-dividend ratio.

- Epstein, Farhi, and Strzalecki *AER* [2014] show that the BY estimates imply the agent would give up 25% of a month's consumption to learn it a month earlier (in a hypothetical world where the agent couldn't act on the information.) Plausible?

- Epstein-Zin preferences motivated by the desire to unlink the agent’s static risk aversion from her taste for consumption smoothing over time.

- But there is (at least) one other aspect of preference to worry about- taste for non-informative information- and Epstein Zin links it to the other two.

- It would be nice to have a tractable alternative, but since indifference to timing in KP forces W to be linear, we may need to drop the KP form and let current flow utility depend on some aspects of past consumption.

Heuristics and Biases Tversky-Kahnemann *Science* [1974] surveys their past findings.

Base Rate Neglect: When making statistical inferences people tend to ignore the role of the base rate (prior probability) and rely too strongly on the signal (new information).

Taxicab example from TK [1982] "Evidential Impact of Base Rates":

"A cab was involved in a hit and run accident at night. Two cab companies, the Green and the Blue, operate in the city. 85% of the cabs in the city are Green and 15% are Blue, and a witness identified the cab as Blue. The court tested the reliability of the witness under the same circumstances that existed on the night of the accident and concluded that the witness correctly identified each one of the two colors 80% of the time and failed 20% of the time. What is the probability that the cab involved in the accident was Blue rather than Green?"

$$\Pr(\text{Blue}|\text{blue report}) = \frac{.15*.8}{.15*.8+.85*.2} = \frac{12}{29}$$, but many subjects said .8.

Lawyer-engineer example:

> *"Jack has been drawn from a population which is x% engineers and 1-x % lawyers. Jack wears a pocket protector. What is the probability Jack is an engineer?"*

- Subjects weren't told the fractions of engineers and lawyers that use pocket protectors, so we don't know what the "right" answer is.

- If beliefs about these fractions aren't affected by the claimed base rate, Bayes rule yields a much higher answer when x=.7 than when it is .3. Yet peoples' answers are about the same.

Representativeness

"Linda is 31 years old, single, outspoken, and very bright. She majored in philosophy. As a student, she was deeply concerned with issues of discrimination and social justice, and also participated in anti-nuclear demonstrations. Please rank the following statements by their probability, using 1 for the most probable and 8 for the least probable."

1. Linda is a teacher in elementary school.
2. Linda works in a bookstore and takes Yoga classes.
3. Linda is active in the feminist movement.
4. Linda is a psychiatric social worker.
5. Linda is a member of the League of Women Voters.
6. Linda is a bank teller.
7. Linda is an insurance salesperson.
8. Linda is a bank teller and is active in the feminist movement.

Depending on the subject population, 80%-90% rank item 8 as more likely than item 6. (*non-incentivized, but still!)*

Insensitivity to Sample Size (and the “law of small numbers”)

Kahneman and Tversky *Cog. Pysch.* [1972]

The average heights of adult males and females in the US are, respectively, 5’10” and 5’4”. Both distributions are approximately normal with a standard deviation of about 2.5 in. An investigator has selected one population by chance and has drawn from it a random sample. What do you think are the odds that he has selected the male population if:

- *the sample consists of a single person whose height is 5’10”?*
- *the sample consists of 6 persons whose average height is 5’8”?*

- KT asked 115 students from the University of Michigan, all of whom had had at least one course in statistics.

- Asked for male/female odds ratio in both cases.

- To get numerical values from Bayes rule here we need to numerically evaluate the standard normal.

KT's 'correct' answers: 16 and 29; Harry's are 17.8 and 317.16.

- Median reports by the subjects were 8 for the first sample and 2.5 in the second: so a) too little impact of the data, and b) the larger sample was misjudged to be less informative- people ignore the way the empirical cdf concentrates on the theoretical one.

- Potential explanation: the first sample is "representative" of a male student, while the second looks a bit mixed (but more male than female).

- Also from the 1972 *Cog Psych* paper: subjects –do- pay attention to sample size in the one posterior odds calculation where they shouldn't, namely with a two-point prior.

2 urns, one has 2/3 red balls and 1/3 white and the other has 2/3 whiet 1/3 red. An urn is picked at random. One person draws 5 balls from the urn, 4 red and 1 white; another person observe 20 balls, 12 red and 8 white. Which of the two individuals should feel more confident that the urn contains 2/3 red balls and 1/3 white balls, rather than the opposite?

- The posterior odds ratio for mostly red to mostly white after a sample of *r* Red and *w* White is $\frac{(2/3)^r(1/3)^w}{(1/3)^r(2/3)^w}\frac{.5}{.5} = 2^{r-w}$, <u>independent of sample size *r*+*w*.</u>

- After (4,1) the odds ratio is 8 and after (12,8) it is 16, but most people said the smaller sample was stronger evidence, presumably because its ratio of r/w was larger. Replicated by Griffin Tversky *Cog. Psych.* [1992] with a $20 prize for being most accurate (*should we expect this to induce truthful reports?*)

Not clear if this is general critique of Bayesian updating or a critique of two-point priors, because 2-point priors have very unintuitive implications.

Specifically, sample size does matter with a non-doctrinaire prior (such as the beta distribution) over the probability of *Red*- this is the case if you think any ratoi of Red to Black balls is *a priori* possible.

In this case the posterior density converges to a point mass on *r/w*.

- Koehler *Behavioral and Brain Sciences* [1996] surveys various attempted replications of the KT experiments.
- In 7 “lawyer/engineer” experiments, 3 had more or less complete base rate neglect, in 4 the base rate had a “strong” (not necessarily correct) effect.
- Koehler also points out that base rates have more impact when people learn through trial by trial feedback (in an experiment where subjects were told the base rate and also made guesses).
- In many field settings, subjects aren’t told the base rate.

Hertwig, Barron, Weber, and Erev *Pysch. Science* [2004]:

"*When people have access to information sources such as newspaper weather forecasts…they can make decisions from description. When people must decide whether to back up their computer's hard drive, cross a busy street, or go out on a date, however, they typically do not have any summary description of the possible outcomes or their likelihoods. For such decisions, people can call only on their own encounters with such prospects, making decisions from experience. Decisions from experience and decisions from description can lead to dramatically different choice behavior. In the case of decisions from description, people make choices as if they overweight the probability of rare events…in the case of decisions from experience, in contrast, people make choices as if they underweight the probability of rare events…two possible causes of this underweighting are reliance on relatively small samples… and overweighting of recently sampled information.*"

- “Recency bias” is fairly common in the lab but it can depend on instructions and on the availability of calculators and summary statistics.
- But there isn’t a consensus on how best to model learning across a range of problem domains- and very little work on this so far by economists.
- Another problem people have is in dealing with autocorrelation or lack thereof.
- “Gambler’s fallacy”: false belief in negative autocorrelation. Croson and Sundali *J Risk Uncertainty* [2005]: roulette gamblers in a casino bet more on a color after a long streak of the other. Suettens et al *JEEA* [2016]: “underbetting” on lottery numbers that have recently won; this is costly due to the pari-mutuel payoff structure. Chen at al *QJE* [2016]: judges and referees tend to alternate between decisions in randomly ordered decision problems, as if they thought the right decision is negatively autocorrelated. (effect is 5% of decisions. Data is refugee court asylum decisions- looking only at decisions on same or consecutive days, umpire calls of balls and strikes in MLB, and a field experiment on loans in India.)

- In the roulette example people should know that the rolls are independent and what the probabilities are. In the lottery example people at least should know that all numbers are equally likely. With the judges the cases are assigned randomly.

Reading for next time: Strzalecki 9.1 (value of information), Augenblink et al *QJE* [2015], Fudenberg and Levine *AER* [2006], *Ema* [2012] (*you can skip the proofs in both of these and the second half of the 2012 paper where it introduces a stock of cognitive resources*)

If the agent learns over time, and can adjust to the information, preference needn't be "linear in probabilities." Dynamic choice+ learning is outside of the formal setup of the lecture but worth looking at.

<u>Example:</u>

- Two periods, 0 and 1, $V = u(z_0) + u(z_1) = \ln(z_0 + 1) + \ln(z_1 + 1)$.

- Consumer faces fixed prices of 1 in each period and has fixed income *I.*

- Has to choose period-0 consumption z_0 before knowing *I.*

- Suppose that the consumer faces the income lottery $p = .5\delta_{98} + .5\delta_{178}$.

- Can show the optimal choice is $z_0^* = 59$, with expected utility $.5\big(\ln(60)+\ln(30)\big)+.5\big(\ln(60)+\ln(120)\big):=V^*$

- Define I^{**} to be the certain equivalent of this lottery: $V^* = 2\ln(\frac{I^{**}}{2}+1)$.

- Then the consumer is indifferent between income lottery *p* and a sure income I^{**}.

- The independence axiom on income lotteries would say the consumer is indifferent between a lottery between *p* and I^{**} and either of them, *e.g.* $q := \alpha p + (1-\alpha)\delta_{I^{**}} \sim p$

- But can show *q* is strictly worse.
Reason: the optimal z_0 for I^{**} is different than for *p.* So can't guarantee V^* given a lottery over *p* and I^{**} - the consumer has less information here.