What are we weighting for?

A mechanistic model for probability weighting

Ole Peters Alexander Adamou Yonatan Berman Mark Kirstein

✉ m.kirstein@lml.org.uk

@nonergodicMark

Full paper at bit.ly/lml-pw-r1

Live paper at bit.ly/lml-pw

MPI MiS, Group Seminar, 03 November 2020

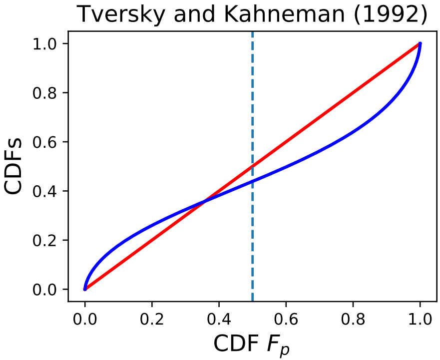
Max Planck Institute for Mathematics in the Sciences

UNIVERSITÄT LEIPZIG Mathematisches Institut

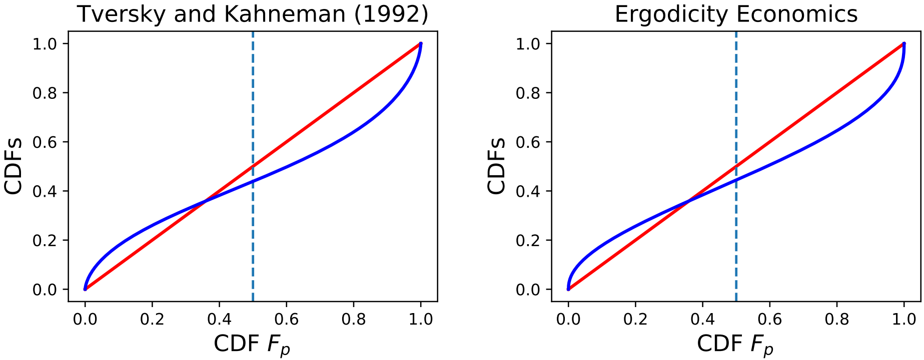
LONDON MATHEMATICAL LABORATORY

Mark Kirstein

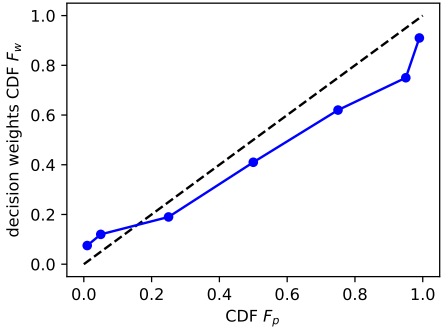
Main result



Main result


Tversky and Kahneman (1992)
Ergodicity Economics
CDFs
CDF F_p
0.0
0.2
0.4
0.6
0.8
1.0

Definition of Probability Weighting (PW)



(Preston and Baratta 1948, p. 188, Fig. 1, relabelled axes)

"The existence [...] a scale of *psychological probability* and its functional relationship to the scale of *mathematical probability*" (p. 186)

- empirical pattern: inverse-S shape
- important component in behavioural economics (Cumulative Prospect Theory)

Classical interpretation of PW:

- as a cognitive process
- maladaptive irrational cognitive bias

↪ How does this pattern emerge?

↪ Can we derive a functional form (rather than fit a function)?

Related Literature

- no motivation of the functional form of weighting curve other than fit
- origin of PW preferences? no stable mappings (Stewart et al. 2015)
- experimental design is a key confounder $\rightarrow$ description-experience gap, *i.e.* less (even under-) overweighting in decisions-from-experience (Hertwig et al. 2004; Hertwig and Erev 2009)
- meta-analyses find all possible weighting curves (Ungemach et al. 2009; Wulff et al. 2018, Tab. 9)
- non-human animals (Andrews et al. 2018; Constantinople et al. 2019; Trimmer et al. 2011)

Statistical explanations

- PW maladaptive, due to biased estimation (Fox and Hadar 2006)
- PW adaptive in sequential learning problems (Seo et al. 2019)
- PW adaptive heuristic as an approximate Bayesian solution for the inference problem (Martins 2006)

$\hookrightarrow$ reproducibility, context dependence, (mal)adaptive

Setup

No *truth*, only difference in models of uncertainty

Task: model payout, x, of a gamble as a random variable.

Disinterested Observer (DO)

DO assigns PDF $p(x)$
$\hookrightarrow$ CDF $F_p(x)$

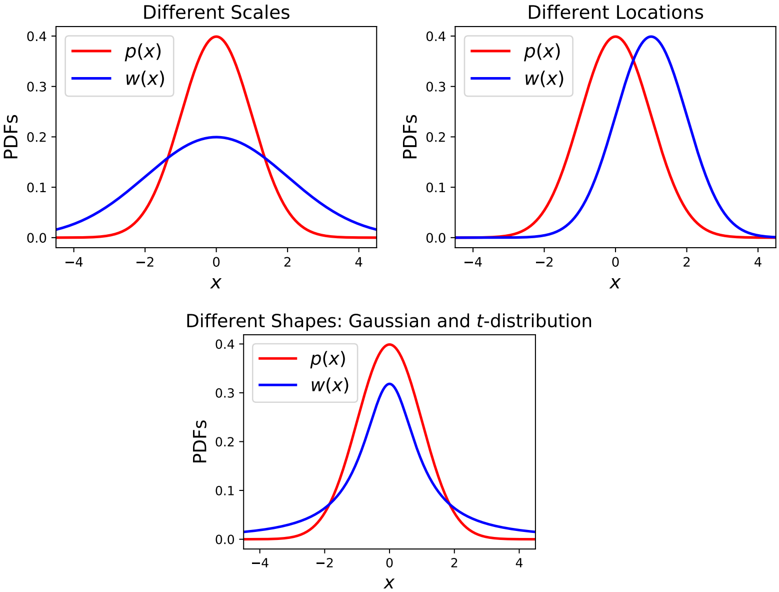
$$F_p(x) = \int_{-\infty}^{x} p(s)\,\mathrm{d}s$$

Decision Maker (DM)

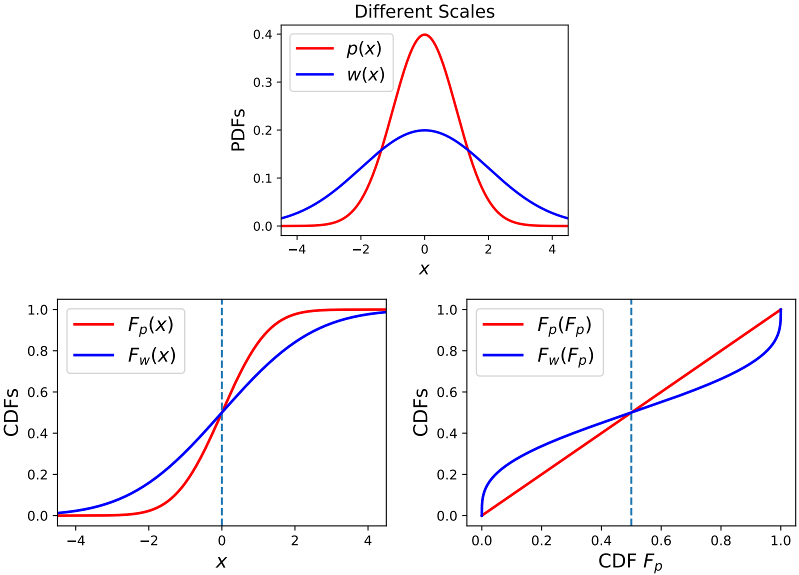
DM assigns different PDF $w(x)$
$\hookrightarrow$ CDF $F_w(x)$

$$F_w(x) = \int_{-\infty}^{x} w(s)\,\mathrm{d}s$$

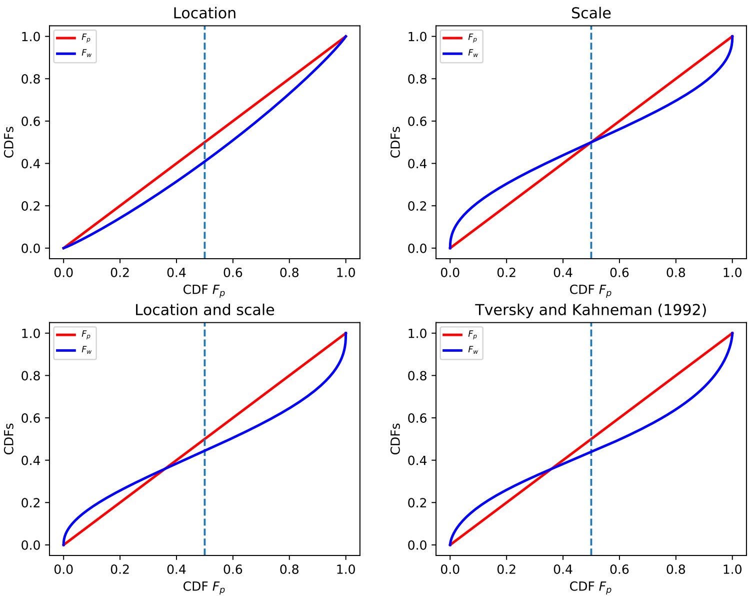
Scales, Locations, Shapes



Thought Experiment: DM assumes greater scale



Asymmetric Inverse-S = diff. in uncertainty + diff. in locations



Numerically feasible for arbitrary distributions:

1. list values of DO's CDF, $F_p(x)$, at set x_i
2. list values of DM's CDF, $F_w(x)$, at same x_i
3. plot $F_w(x)$ *vs.* $F_p(x)$

Functional form of the weighting function

Gaussian case with different scale:

$$w(p) = p^{\frac{1}{\alpha^2}} \underbrace{\frac{\left(2\pi\sigma^2\right)^{\frac{1-\alpha^2}{2\alpha^2}}}{\alpha}}_{\text{normalisation factor}},$$

where

- DO's scale is σ, $X \sim \mathcal{N}\left(\mu, \sigma^2\right)$
- DM's scale is $\alpha\sigma$, $X \sim \mathcal{N}\left(\mu, (\alpha\sigma)^2\right)$
- DO uses greater uncertainty $\alpha < 1 \rightarrow$ regular-S shape
- DM uses greater uncertainty $\alpha > 1 \rightarrow$ inverse-S shape
- uncertainty measured by the standard deviation

Functional Forms

(EE mechanism) Gaussian case:

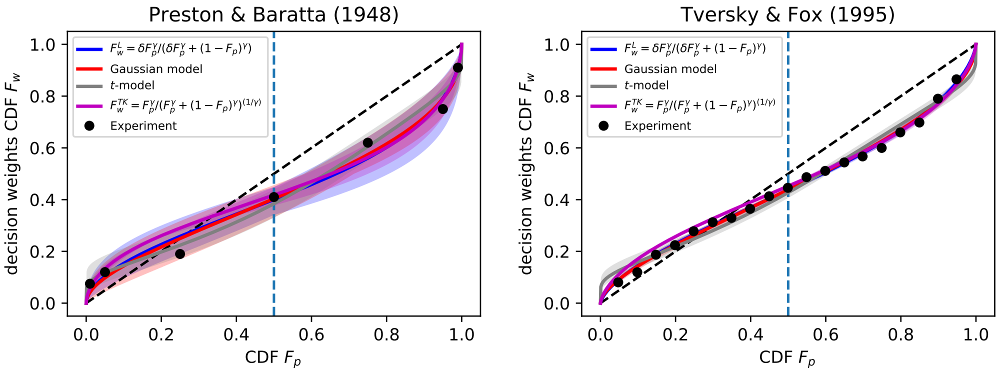
$$\boxed{w(p) = p^{\frac{1}{\alpha^2}} \frac{\left(2\pi\sigma^2\right)^{\frac{1-\alpha^2}{2\alpha^2}}}{\alpha}} \tag{1}$$

Tversky and Kahneman (1992, $\gamma = 0.68$):

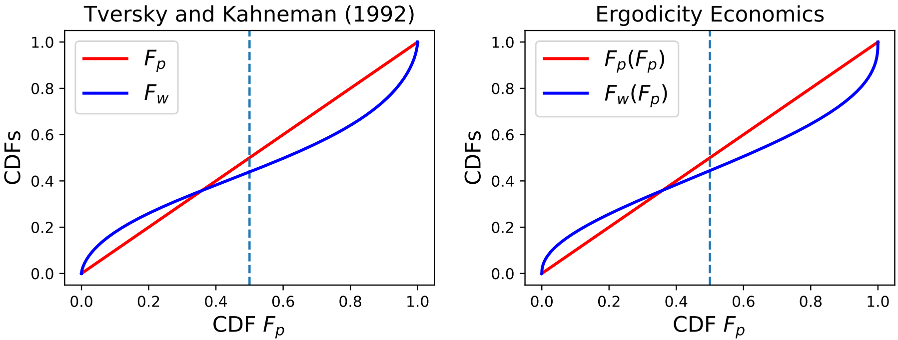
$$F_w^{TK}\left(F_p; \gamma\right) = \left(F_p\right)^{\gamma} \frac{1}{\left[\left(F_p\right)^{\gamma} + \left(1 - F_p\right)^{\gamma}\right]^{1/\gamma}} \tag{2}$$

Lattimore et al. (1992):

$$F_w^{L}\left(F_p; \delta, \gamma\right) = \frac{\delta F_p^{\gamma}}{\delta F_p^{\gamma} + \left(1 - F_p\right)^{\gamma}} \tag{3}$$


Preston & Baratta (1948)
Tversky & Fox (1995)
$F_w^L = \delta F_p^\gamma/(\delta F_p^\gamma + (1-F_p)^\gamma)$
Gaussian model
t-model
$F_w^{TK} = F_p^\gamma/(F_p^\gamma + (1-F_p)^\gamma)^{(1/\gamma)}$
Experiment
decision weights CDF F_w
CDF F_p
0.0
0.2
0.4
0.6
0.8
1.0

Interim conclusion



- DM's greater scale gives inverse-S shape (unimodal distributions)
- difference in locations gives asymmetry
- reproduces observations of probability weighting

Job done. Thank you for your attention ;)

Mark Kirstein

The Ergodicity Question

Typical DO concern

What happens on average to the ensemble of subjects?

$\neq$

Typical DM concern

What happens to me on average over time?

The Ergodicity Question

Typical DO concern $\neq$ **Typical DM concern**

What happens on average to the ensemble of subjects?

What happens to me on average over time?

↪ Broken ergodicity is more fundamental than being just about *correct* behavioural models

- broken ergodicity leads to a collision of ensemble and time perspective in mathematical statements like $F_w(F_p)$

Reasons for different uncertainty models

- DM has no control over experiment
- experiment may be unclear to DM
- DM may not trust DO
- ...

Experiencing probabilities

- "probability" is polysemous
- natural frequencies: "10 out of 100" vs 10% (Gigerenzer 1991, 2018; Hertwig and Gigerenzer 1999)
- probabilities are not observable
- probabilities encountered as
  - known frequencies in ensemble of experiments (DO)
  - frequencies estimated over time (DM)

$\hookrightarrow$ **estimates have uncertainties**

$\hookrightarrow$ **cautious (rational?) DM accounts for these uncertainties**

$\hookrightarrow$ **small probabilities come with large relative uncertainties**

probabilities are not observable, but counts are

Rare Event

- $p(x) = 0.001$
- $T = 100$ observations
- $\sim 99.5\%$ get 0 or 1 events
- $\hat{p}(x) = 0$ or $\hat{p}(x) = 0.01$

$\hookrightarrow$ $\hat{p}(x)$ off by 1000%

Common Event

- $p(x) = 0.5$
- $T = 100$ observations
- $\sim 99.5\%$ get between 35 and 65 events,
- $0.35 < \hat{p}(x) < 0.65$

$\hookrightarrow$ $\hat{p}(x)$ off by 30%

$\boxed{\hookrightarrow \text{ small } p(x)\text{, small count} \rightarrow \text{big uncertainty}}$

Scaling of uncertainty: small count $\rightarrow$ big uncertainty

1. scaling of counts $n(x)$:

$$n(x) \sim p(x)\delta x T$$

Scaling of uncertainty: small count $\rightarrow$ big uncertainty

1. scaling of counts $n(x)$:

$$n(x) \sim p(x)\delta x T$$

Scaling of uncertainty: small count $\rightarrow$ big uncertainty

1. scaling of counts $n(x)$:

$$n(x) \sim p(x)\delta x T \quad \Longleftrightarrow \quad p(x) \sim {}^{n(x)}/_{T\delta x}$$

2. uncertainty in Poisson-distributed counts $\sim \sqrt{n(x)}$

Mark Kirstein

Scaling of uncertainty: small count $\rightarrow$ big uncertainty

1. scaling of counts $n(x)$:

$$n(x) \sim p(x)\delta x T \quad \Longleftrightarrow \quad p(x) \sim {}^{n(x)}/_{T\delta x}$$

2. uncertainty in Poisson-distributed counts $\sim \sqrt{n(x)}$
3. estimate of the asymptotic probability density

$$p(x) \approx \underbrace{\frac{n(x)}{T\delta x}}_{\text{estimate}} \pm \underbrace{\frac{\sqrt{n(x)}}{T\delta x}}_{\text{relative uncertainty}} \tag{4}$$

Mark Kirstein

Scaling of uncertainty: small count $\rightarrow$ big uncertainty

1. scaling of counts $n(x)$:

$$n(x) \sim p(x)\delta x T \quad \Longleftrightarrow \quad p(x) \sim {}^{n(x)}/_{T\delta x}$$

2. uncertainty in Poisson-distributed counts $\sim \sqrt{n(x)}$
3. estimate of the asymptotic probability density

$$p(x) \approx \underbrace{\frac{n(x)}{T\delta x}}_{\text{estimate}} \pm \underbrace{\frac{\sqrt{n(x)}}{T\delta x}}_{\text{relative uncertainty}} \tag{4}$$

4. express the uncertainty in terms of the estimate itself

$$\varepsilon\left[\hat{p}(x)\right] \equiv \frac{\sqrt{n(x)}}{T\delta x} = \sqrt{\frac{\hat{p}(x)}{T\delta x}} \tag{5}$$

Scaling of uncertainty: small count $\rightarrow$ big uncertainty

1. scaling of counts $n(x)$:

$$n(x) \sim p(x)\delta x T \quad \Longleftrightarrow \quad p(x) \sim {}^{n(x)}/_{T\delta x}$$

2. uncertainty in Poisson-distributed counts $\sim \sqrt{n(x)}$
3. estimate of the asymptotic probability density

$$p(x) \approx \underbrace{\frac{n(x)}{T\delta x}}_{\text{estimate}} \pm \underbrace{\frac{\sqrt{n(x)}}{T\delta x}}_{\text{relative uncertainty}} \tag{4}$$

4. express the uncertainty in terms of the estimate itself

$$\varepsilon\left[\hat{p}(x)\right] \equiv \frac{\sqrt{n(x)}}{T\delta x} = \sqrt{\frac{\hat{p}(x)}{T\delta x}} \tag{5}$$

5. standard error $\lim_{p(x)\rightarrow 0} \sqrt{\hat{p}(x)/T\delta x}$ in $\hat{p}$ shrinks

Scaling of uncertainty: small count $\rightarrow$ big uncertainty

1. scaling of counts $n(x)$:

$$n(x) \sim p(x)\delta x T \quad \Longleftrightarrow \quad p(x) \sim {}^{n(x)}/_{T\delta x}$$

2. uncertainty in Poisson-distributed counts $\sim \sqrt{n(x)}$
3. estimate of the asymptotic probability density

$$p(x) \approx \underbrace{\frac{n(x)}{T\delta x}}_{\text{estimate}} \pm \underbrace{\frac{\sqrt{n(x)}}{T\delta x}}_{\text{relative uncertainty}} \tag{4}$$

4. express the uncertainty in terms of the estimate itself

$$\varepsilon\left[\hat{p}(x)\right] \equiv \frac{\sqrt{n(x)}}{T\delta x} = \sqrt{\frac{\hat{p}(x)}{T\delta x}} \tag{5}$$

5. standard error $\lim\limits_{p(x)\to 0} \sqrt{\hat{p}(x)/T\delta x}$ in $\hat{p}$ shrinks
6. relative error in the estimate $\lim\limits_{p(x)\to 0} 1/\sqrt{\hat{p}(x)T\delta x}$ grows,

Scaling of uncertainty: small count → big uncertainty

1. scaling of counts $n(x)$:

$$n(x) \sim p(x)\delta x T \quad \Longleftrightarrow \quad p(x) \sim {}^{n(x)}/_{T\delta x}$$

2. uncertainty in Poisson-distributed counts $\sim \sqrt{n(x)}$
3. estimate of the asymptotic probability density

$$p(x) \approx \underbrace{\frac{n(x)}{T\delta x}}_{\text{estimate}} \pm \underbrace{\frac{\sqrt{n(x)}}{T\delta x}}_{\text{relative uncertainty}} \tag{4}$$

4. express the uncertainty in terms of the estimate itself

$$\varepsilon\left[\hat{p}(x)\right] \equiv \frac{\sqrt{n(x)}}{T\delta x} = \sqrt{\frac{\hat{p}(x)}{T\delta x}} \tag{5}$$

5. standard error $\lim\limits_{p(x)\to 0} \sqrt{\hat{p}(x)/T\delta x}$ in $\hat{p}$ shrinks
6. relative error in the estimate $\lim\limits_{p(x)\to 0} 1/\sqrt{\hat{p}(x)T\delta x}$ grows,
7. **Prediction:** $\lim_{T\to\infty} \hat{p} \to p$

Cautionary principle : DMs don't like surprises

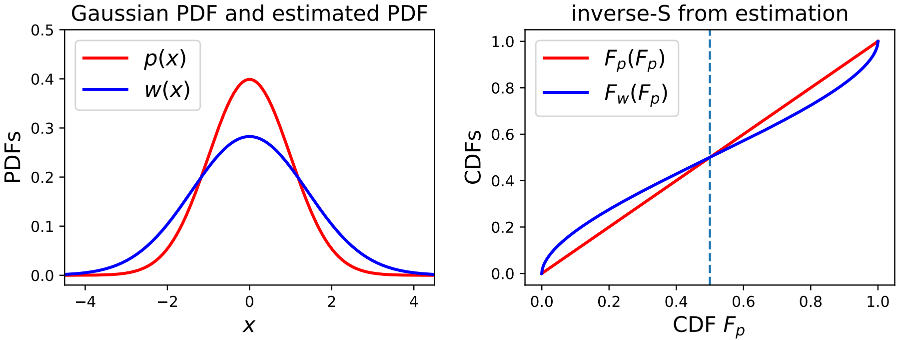
To avoid surprises, DMs add estimation uncertainty $\varepsilon\left[\hat{p}(x)\right]$ to every estimated probability, then normalize, s.t.

$$w(x) = \frac{\hat{p}(x) + \varepsilon\left[\hat{p}(x)\right]}{\int \left(\hat{p}(s) + \varepsilon\left[\hat{p}(s)\right]\right) ds} \tag{6}$$

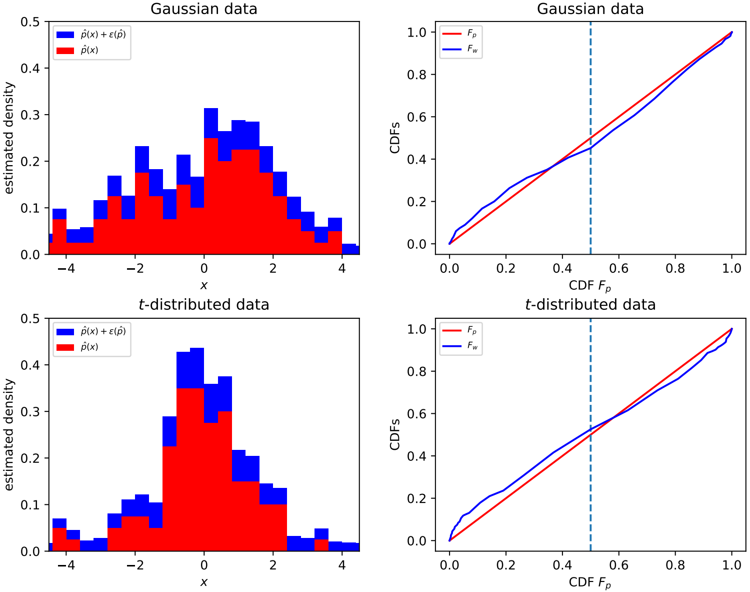
Cautionary principle : DMs don't like surprises

To avoid surprises, DMs add estimation uncertainty $\varepsilon\left[\hat{p}(x)\right]$ to every estimated probability, then normalize, s.t.

$$w(x) = \frac{\hat{p}(x) + \varepsilon\left[\hat{p}(x)\right]}{\int \left(\hat{p}(s) + \varepsilon\left[\hat{p}(s)\right]\right) ds} \tag{6}$$



Interactive code at https://bit.ly/lml-pw-count-b



$T=100$, $\delta x=0.4$, estimates of $\hat{p}(x)$ in red, estimates with one standard error $\hat{p}(x)+\varepsilon\left[\hat{p}(x)\right]$ in blue.

Switch to sketch of experimental design

Conclusion

Classical interpretation of PW

- overestimation of low probability events
- underestimation of high probability events

$\hookrightarrow$ maladaptive irrational cognitive bias

Ergodicity Economics and PW

- inverse-S shape: neutral indicator of different models of uncertainty
- reported observations consistent with DM's extra uncertainty
- may arise from DM estimating probabilities over time

$\hookrightarrow$ Probability weighting is rational cautious behaviour under uncertainty over time

- testable prediction(s) $\rightarrow$ Let's run an experiment!
- Manuscript at `bit.ly/lml-pw-r1`
- Interactive code at `bit.ly/lml-pw-count-b`

Thank you for your attention!

I'm looking forward to the discussion
Comments & questions are very welcome, here or to

✉ m.kirstein@lml.org.uk
@nonergodicMark

Ergodicity Economics Online Conference
18-20 Jan 2021

EE2021

lml.org.uk/ee2021/

Submit an open peer review to this paper on bit.ly/lml-pw-r1

BACK UP

Probability Weighting as an Estimation Issue

"It is important to distinguish overweighting, which refers to a property of decision weights, from the overestimation that is commonly found in the assessment of the probability of rare events. [...] In many real-life situations, overestimation and overweighting may both operate to increase the impact of rare events." (Kahneman and Tversky 1979, p. 281)

↪ distinguish between
- uncertainty estimation and
- "weighting"

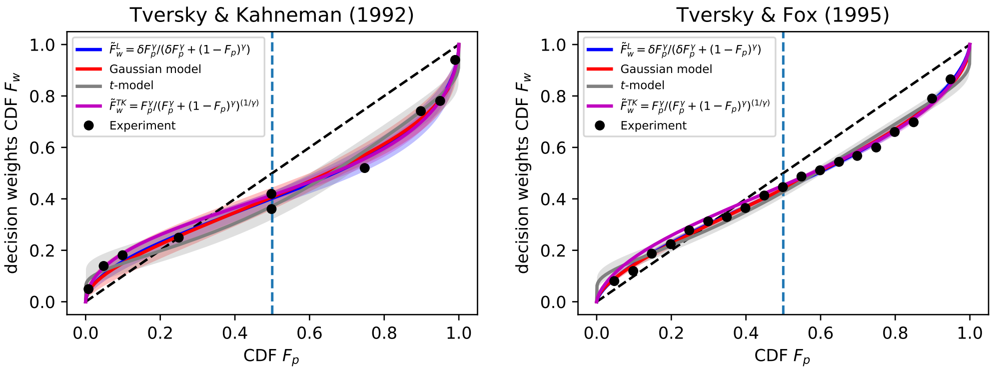
we analyse the former and find very good agreement with the empirical inverse-S pattern

↪ How big is the residual "probability weighting" after accounting for uncertainty estimation?

Main Results

Mark Kirstein

Estimation Error Explains 99% of Probability Weighting



- similar fits of Gaussian & t-distributed model

$\hookrightarrow$ How big is the residual "probability weighting" after accounting for estimation errors?

◂ Main Results

References I

Andrews, Clare et al. (2018). "A marker of biological ageing predicts adult risk preference in European starlings, Sturnus vulgaris". *Behavioral Ecology* 29 (3), pp. 589–597. DOI:10.1093/beheco/ary009 (cit. on p. 5).

Constantinople, Christine M., Alex T. Piet, and Carlos D. Brody (2019). "An Analysis of Decision under Risk in Rats". *Current Biology* 29 (12), 2066–2074.e5. DOI:10.1016/j.cub.2019.05.013 (cit. on p. 5).

Fox, Craig R. and Liat Hadar (2006). ""Decisions from experience" = sampling error + prospect theory: Reconsidering Hertwig, Barron, Weber & Erev (2004)". *Judgement and Decision Making* 1 (2). URL: `http://journal.sjdm.org/06144/jdm06144.htm` (cit. on p. 5).

Gigerenzer, Gerd (1991). "How to Make Cognitive Illusions Disappear: Beyond "Heuristics and Biases"". *European Review of Social Psychology* 2 (1), pp. 83–115. DOI:10.1080/14792779143000033 (cit. on p. 18).

Gigerenzer, Gerd (2018). "The Bias Bias in Behavioral Economics". *Review of Behavioral Economics* 5 (3-4), pp. 303–336. DOI:10.1561/105.00000092 (cit. on p. 18).

Hertwig, Ralph, Greg Barron, Elke U. Weber, and Ido Erev (2004). "Decisions from Experience and the Effect of Rare Events in Risky Choice". *Psychological Science* 15 (8), pp. 534–539. DOI:10.1111/j.0956-7976.2004.00715.x (cit. on p. 5).

References II

Hertwig, Ralph and Ido Erev (2009). "The description–experience gap in risky choice". *Trends in Cognitive Sciences* 13 (12), pp. 517–523. DOI:10.1016/j.tics.2009.09.004 (cit. on p. 5).

Hertwig, Ralph and Gerd Gigerenzer (1999). "The 'conjunction fallacy' revisited: how intelligent inferences look like reasoning errors". *Journal of Behavioral Decision Making* 12 (4), pp. 275–305. DOI:10.1002/(sici)1099-0771(199912)12:4¡275::aid-bdm323¿3.0.co;2-m (cit. on p. 18).

Kahneman, Daniel and Amos Tversky (1979). "Prospect Theory: An Analysis of Decision under Risk". *Econometrica* 47 (2), pp. 263–291. DOI:10.2307/1914185 (cit. on p. 35).

Lattimore, Pamela K., Joanna R. Baker, and A. Dryden Witte (1992). "Influence of Probability on Risky Choice: A Parametric Examination". *Journal of Economic Behavior and Organization* 17 (3), pp. 377–400. DOI:10.1016/S0167-2681(95)90015-2 (cit. on p. 12).

Martins, André C. R. (2006). "Probability biases as Bayesian inference". *Judgement and Decision Making* 1 (2), pp. 108–117. URL: http://journal.sjdm.org/vol1.2.htm (cit. on p. 5).

Preston, Malcolm G. and Philip Baratta (1948). "An Experimental Study of the Auction-Value of an Uncertain Outcome". *American Journal of Psychology* 61 (2), p. 183. DOI:10.2307/1416964 (cit. on p. 4).

References III

- Seo, Daewon, Ravi Kiran Raman, Joong Bum Rhim, Vivek K Goyal, and Lav R. Varshney (2019). "Beliefs in Decision-Making Cascades". *IEEE Transactions on Signal Processing* 67 (19), pp. 5103–5117. DOI:10.1109/tsp.2019.2935865 (cit. on p. 5).
- Stewart, Neil, Stian Reimers, and Adam J. L. Harris (2015). "On the Origin of Utility, Weighting, and Discounting Functions: How They Get Their Shapes and How to Change Their Shapes". *Management Science* 61 (3), pp. 687–705. DOI:10.1287/mnsc.2013.1853 (cit. on p. 5).
- Trimmer, Pete C. et al. (2011). "Decision-making under uncertainty: biases and Bayesians". *Animal Cognition* 14 (4), pp. 465–476. DOI:10.1007/s10071-011-0387-4 (cit. on p. 5).
- Tversky, Amos and Daniel Kahneman (1992). "Advances in Prospect Theory: Cumulative Representation of Uncertainty". *Journal of Risk and Uncertainty* 5 (4), pp. 297–323. DOI:10.1007/BF00122574 (cit. on p. 12).
- Ungemach, Christoph, Nick Chater, and Neil Stewart (2009). "Are Probabilities Overweighted or Underweighted When Rare Outcomes Are Experienced (Rarely)?" *Psychological Science* 20 (4), pp. 473–479. DOI:10.1111/j.1467-9280.2009.02319.x (cit. on p. 5).
- Wulff, Dirk U., Max Mergenthaler-Canseco, and Ralph Hertwig (2018). "A meta-analytic review of two modes of learning and the description-experience gap.". *Psychological Bulletin* 144 (2), pp. 140–176. DOI:10.1037/bul0000115 (cit. on p. 5).