Monotone Additive Statistics

Xiaosheng Mu* Luciano Pomatto† Philipp Strack‡ Omer Tamuz§

January 31, 2021

Abstract

We study *statistics*: mappings from distributions to real numbers. We characterize all statistics that are monotone with respect to first-order stochastic dominance, and additive for sums of independent random variables. We explore a number of applications, including a representation of stationary, monotone time preferences, generalizing Fishburn and Rubinstein (1982) to time lotteries.

1 Introduction

How should a random quantity be summarized by a single number? In Bayesian statistics, point estimators capture an entire posterior distribution. In finance, risk measures quantify the risk in a distribution of returns. And in economics, certainty equivalents characterize an expected utility agent's preference for uncertain outcomes.

We use the term *statistic* to describe a map that assigns a number to each real-valued random variable, with the basic requirement that this number depends only on the distribution of the random variable.[1] We study statistics that are monotone with respect to first-order stochastic dominance, and additive for sums of independent random variables. An example of a monotone additive statistic is the expectation. The median is monotone but not additive, while the variance is additive but not monotone.

Monotonicity is a well studied property of statistics (see, e.g., Bickel and Lehmann, 1975a,b), and holds, for example, for certainty equivalents of monotone preferences. Additivity is a stronger assumption. We focus on this property because of its conceptual simplicity and because it serves as a baseline assumption in many settings. In particular,

*Princeton University. Email: xmu@princeton.edu.

†Caltech. Email: luciano@caltech.edu.

‡Yale University. Email: philipp.strack@yale.edu. Philipp Strack was supported by a Sloan fellowship.

§Caltech. Email: tamuz@caltech.edu. Omer Tamuz was supported by a grant from the Simons Foundation (#419427), a Sloan fellowship, and a BSF award (#2018397).

[1] The term "descriptive statistic" usually refers to maps associating a number to observations or to empirical distributions. Because of its simplicity, we apply it here to general distributions. See e.g. Bickel and Lehmann (1975a).

we show below that additivity corresponds to a form of stationarity in the context of preferences over time lotteries.

Beyond the expectation, an additional example of a monotone additive statistic is the map K_a that, given $a \in \mathbb{R}$, assigns to each random variable X the value

$$K_a(X) = \frac{1}{a} \log \mathbb{E}\left[\mathrm{e}^{aX}\right]. \tag{1}$$

In the language of statistics, the map is the (normalized) cumulant generating function evaluated at a. In economics, it corresponds to the certainty equcivalent defined by a CARA preference over gambles. For bounded random variables, the essential minimum and maximum provide further examples of such statistics; as we explain later, they are the limit of $K_a(X)$ as a approaches $\pm\infty$.

Our main result is that these examples, and their weighted averages, are the only monotone additive statistics. That is, we show that every monotone additive statistic Φ is of the form

$$\Phi(X) = \int K_a(X) \, \mathrm{d}\mu(a)$$

for some probability measure μ. This result provides a simple representation of a natural family of statistics, which one may a priori have expected to be much richer.

Our first application is to time lotteries. The starting point for our analysis is the work by Fishburn and Rubinstein (1982), who study preferences over dated rewards: a monetary amount, together with the time at which it will be received. They show that exponential discounting of time arises from a set of axioms, of which the most substantial axiom is *stationarity*: preferences remain invariant when the dated rewards under consideration are shifted by the same amount of time.

We extend the setting of Fishburn and Rubinstein (1982) to that of *time lotteries*: a monetary amount, together with a random time at which it will be received. In this setting, we also introduce a stationarity axiom that requires preferences to be invariant with respect to random shifts in time. As we argue in the main text, this stationarity axiom captures a basic requirement of dynamic consistency.

We show that stationarity, together with a monotonicity and a continuity axiom, imply that the preference admits the representation

$$u(x) \cdot \mathrm{e}^{-r \int K_a(T) \, \mathrm{d}\mu(a)},$$

for each time lottery that delivers a monetary reward x at a random time T. Over deterministic dated rewards, the representation coincides with the one of Fishburn and Rubinstein (1982). General time lotteries are reduced to deterministic ones by a monotone additive statistic that maps the random time T to the value $\int K_a(T) \, \mathrm{d}\mu(a)$. For each parameter a, the term $K_a(T)$ is the certainty equivalent of T under an expected discounted preference with discount factor a. The different certainty equivalents are thus averaged according to the measure μ.

Our representation of these monotone and stationary time preferences has implications for the understanding of risk attitudes toward time. Risk preferences over time lotteries have been studied both theoretically and experimentally (Chesson and Viscusi, 2003; Onay and Öncüler, 2007; Ebert, 2020; DeJarnette et al., 2020). A basic paradox these papers highlight is that most subjects display risk aversion over the time dimension, even though the standard theory of expected utility with exponential discounting predicts that people are risk-seeking with respect to time lotteries. Our analysis shows that expected exponentially discounted utility is only one of many ways to extend exponential discounting (from dated rewards to time lotteries) while maintaining stationarity. In fact, we characterize a class of stationary preferences over time lotteries that exhibit risk aversion over time.

Our second application is to the domain of monetary gambles. In this domain, it is well known that expected utility agents whose preferences are invariant to background risk must have CARA preferences. Our main characterization theorem implies that beyond expected utility, such agents have certainty equivalents that are weighted averages of CARA certainty equivalents. We similarly extend a result of Rabin and Weizsäcker (2009) from the expected utility domain to general monotone preferences. They show that among expected utility maximizers, only CARA agents do not violate stochastic dominance for combined risks. We show that a monotone preference has this property only if it is represented by a monotone additive statistic, i.e., it is represented by an average of CARA certainty equivalents.

1.1 Related literature

Bickel and Lehmann (1975a,b) study location statistics using a similar axiomatic, non-parametric approach, and also consider the monotonicity property that we impose, but not additivity. In contrast, the mathematics literature has studied additive statistics, as homomorphisms from the convolution semigroup to the reals (see Ruzsa and Székely, 1988; Mattner, 1999, 2004), without imposing monotonicity.

In the finance and actuarial sciences literature, the CARA certainty equivalent $-K_{-a}(X)$ shows up and is often called the *entropic risk measure* of X with parameter a (see Föllmer and Schied, 2011). Goovaerts, Kaas, Laeven, and Tang (2004) prove a result that is similar to our Theorem 1, under the stronger assumption that $K_a(X) \geq K_a(Y)$ for all $a \in \bar{\mathbb{R}}$ implies $\Phi(X) \geq \Phi(Y)$. Our monotonicity property only demands this to hold when $X \geq_1 Y$.[2]

In an earlier paper, Pomatto, Strack, and Tamuz (2020) show that on the larger domain of random variables that have all moments, the only monotone additive statistic is the

[2]Note that $X \geq_1 Y$ implies but is not implied by $K_a(X) \geq K_a(Y)$ for all a.

expectation. This result can be reconciled with Theorem 1 by noting that for any $a \neq 0$, the monotone additive statistic $K_a(X)$ that we identify takes infinite value for some unbounded random variables that have all moments. Fritz, Mu, and Tamuz (2020) show that the expectation remains the unique monotone additive statistic on the even larger domain of L^p random variables, for any $p \geq 1$. They additionally show that there are no monotone additive statistics on L^p with $p < 1$, or on the domain of all random variables, where the expectation may not exist.

A strengthening of our additivity condition is the requirement of additivity for all pairs of random variables, rather than just the independent ones. This stronger assumption turns out to be restrictive: The only statistic that satisfies additivity for all random variables is the expectation (see de Finetti, 1970).

As is well known, directly averaging exponential discounting utilities leads to present bias (see Jackson and Yariv, 2020). This phenomenon gives rise to impossibility results regarding the aggregation of stationary individual preferences into a stationary social preference. Our contribution to this literature is to observe that beyond the expected utility framework, aggregating the certainty equivalents of exponentially discounted preferences can maintain stationarity.

Monotone additive statistics also relate to what we called *additive divergences* in a previous paper (Mu, Pomatto, Strack, and Tamuz, 2021). The domain of an additive divergence consists of Blackwell experiments. It satisfies monotonicity with respect to the Blackwell order and additivity for product experiments. Our characterization of additive divergences in that paper is reminiscent of the one we provide here for monotone additive statistics, with Rényi divergences playing the role of the certainty equivalents K_a in the current work.

The remainder of the paper is organized as follows. In §2 we introduce monotone additive statistics, state our main result and provide an outline of its proof. In §3 we apply this result to time lotteries, and in §4 we apply it to monetary gambles. The appendix contains omitted proofs, as well as a study of monotone sub-additive statistics, i.e., statistics which satisfy $\Phi(X+Y) \leq \Phi(X) + \Phi(Y)$ for independent X and Y.

2 Monotone Additive Statistics

2.1 Definition and characterization

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a nonatomic probability space. We denote by L^∞ the collection of bounded real random variables on this space. By a standard abuse of notation we will identify the constant $c \in \mathbb{R}$ with the constant random variable $X(\omega) = c$. Given $X \in L^\infty$, $\max[X]$ and $\min[X]$ denote its the essential maximum and minimum.

We say that a map $\Phi \colon L^\infty \to \mathbb{R}$ is a *statistic* if (i) whenever $X, Y \in L^\infty$ have the same distribution, $\Phi(X) = \Phi(Y)$, and (ii) $\Phi(c) = c$ for every $c \in \mathbb{R}$; that is, Φ assigns c to the constant random variable c. Condition (ii) may appear restrictive, but it amounts to a simple normalization when combined with the monotonicity and additivity assumptions we make below.[3] We are interested in statistics that satisfy two properties: monotonicity with respect to first-order stochastic dominance, and additivity for sums of independent random variables.[4] A statistic $\Phi\colon L^\infty \to \mathbb{R}$ is *additive* if $\Phi(X+Y) = \Phi(X) + \Phi(Y)$ whenever X and Y are independent random variables. It is *monotone* if $X \geq Y$ almost surely implies $\Phi(X) \geq \Phi(Y)$. It is *monotone with respect to first-order stochastic dominance* if $X \geq_1 Y$ implies $\Phi(X) \geq \Phi(Y)$, where $X \geq_1 Y$ denotes first-order stochastic dominance.

Since we assume the statistic depends only on the distribution, monotonicity with respect to first-order stochastic dominance is equivalent to monotonicity. This equivalence is based on the well-known characterization that $X \geq_1 Y$ if and only if there are random variables $\tilde{X}, \tilde{Y}$ such that X and $\tilde{X}$ are identically distributed, Y and $\tilde{Y}$ are identically distributed, and $\tilde{X} \geq \tilde{Y}$ almost surely. Henceforth when we say Φ is monotone, we mean that it is monotone with respect to first-order stochastic dominance.

Let $\bar{\mathbb{R}} = \mathbb{R} \cup \{-\infty, \infty\}$ denote the two point compactification of $\mathbb{R}$. Given $X \in L^\infty$ and $a \in \bar{\mathbb{R}} \setminus \{0, \pm\infty\}$, let

$$K_a(X) = \frac{1}{a} \log \mathbb{E}\left[\mathrm{e}^{aX}\right]. \tag{2}$$

This is the (normalized) cumulant generating function of X, evaluated at a. We additionally define $K_0(X), K_\infty(X), K_{-\infty}(X)$ to be the expectation, essential maximum and essential minimum of X, respectively; this makes $a \mapsto K_a(X)$ a continuous function from $\bar{\mathbb{R}}$ to $\mathbb{R}$.

It is easy to check that each K_a is a monotone additive statistic. Our main result is that these statistics—together with their weighted averages—constitute all of the monotone additive statistics.

Theorem 1. *$\Phi\colon L^\infty \to \mathbb{R}$ is a monotone additive statistic if and only if there exists a Borel probability measure μ on $\bar{\mathbb{R}}$, such that for every $X \in L^\infty$*

$$\Phi(X) = \int_{\bar{\mathbb{R}}} K_a(X)\,\mathrm{d}\mu(a). \tag{3}$$

Moreover, the measure μ is unique.

Theorem 1 holds for other domains of random variables. Denote by L^∞_+ the bounded non-negative random variables, by $L^\infty_{\mathbb{N}}$ the bounded non-negative integer-valued random

[3] Under monotonicity and additivity, any Φ that satisfies (i) and is not identically zero must have $\Phi(1) \neq 0$, and furthermore $\Phi(X)/\Phi(1)$ is a monotone additive statistic that satisfies (ii).

[4] An alternative, equivalent definition is to let the domain of Φ be the set of distributions of the random variables in L^∞. In this domain, additivity would be defined with respect to convolution. We choose to have the domain consist of random variables, as this approach offers some notational advantages.

variables, and by L_M the random variables X for which $K_a(X)$ is finite for all $a \in \mathbb{R}$. The collection L_M contains, in addition to all the bounded random variables, those unbounded ones whose distribution has "sub-exponential" tails, such as the normal distribution.

Theorem 2. *Let L be either L^∞_+, $L^\infty_\mathbb{N}$ or L_M. Then $\Phi \colon L \to \mathbb{R}$ is a monotone additive statistic if and only if it admits a (unique) representation of the form (3). In the case of L_M, the measure μ has to be compactly supported on $\mathbb{R}$.*

To prove Theorem 2 for the cases of $L = L^\infty_+$ and $L = L^\infty_\mathbb{N}$, we show that any monotone additive statistic defined on these smaller domains can be extended to one on L^∞, and then invoke Theorem 1. The case of the larger domain L_M turns out to be more difficult, and the proof requires some additional ideas that are explained in the appendix.

2.2 Proof sketch of Theorem 1

Our approach to the proof of Theorem 1 is via the *catalytic stochastic order*. Given $X, Y \in L^\infty$, we say that X dominates Y in the catalytic stochastic order on L^∞ if there exists a $Z \in L^\infty$, independent of X and Y, such that $X + Z \geq_1 Y + Z$ (i.e., $X + Z$ stochastically dominates $Y + Z$).

The applicability of this order to our problem is immediate: if Φ is monotone and additive, then whenever X dominates Y in the catalytic stochastic order it holds that $\Phi(X) \geq \Phi(Y)$. To see this, note that domination in the catalytic order implies that

$$\Phi(X + Z) \geq \Phi(Y + Z)$$

for some $Z \in L^\infty$, since Φ is monotone. Additivity of Φ implies that $\Phi(X + Z) = \Phi(X) + \Phi(Z)$ and $\Phi(Y + Z) = \Phi(Y) + \Phi(Z)$, and so we have that $\Phi(X) \geq \Phi(Y)$.

Clearly, if $X \geq_1 Y$ then X also dominates Y in the catalytic order, as one can take $Z = 0$ (or in fact any Z). A priori, one may conjecture that this is also a necessary condition. As we show, this is far from true.

Figure 1 gives a simple example of $X, Y \in L^\infty$ that are not ranked with respect to first-order stochastic dominance, but are ranked with respect to the catalytic order.[5] X is Bernoulli, and equals 1 with probability 1/3. Y has the uniform distribution on $[-\frac{3}{5}, \frac{2}{5}]$. As the figure shows, their c.d.f.s are not ranked, and hence they are not ranked in terms of first-order stochastic dominance.[6]

However, if we let Z assign probability half to $\pm\frac{1}{5}$, then $X + Z >_1 Y + Z$. Intuitively, since the c.d.f. of $X + Z$ is the average of the two translations (by $\pm\frac{1}{5}$) of the c.d.f. of X,

[5] We are indebted to the late Kim Border for helping us construct this example.

[6] Pomatto et al. (2020) give examples of random variables X and Y that are not ranked in stochastic dominance, but are ranked after adding an *unbounded* independent Z. In fact, they show that this is possible whenever $\mathbb{E}[X] > \mathbb{E}[Y]$. As we explain below, this result no longer holds when Z is required to be bounded.

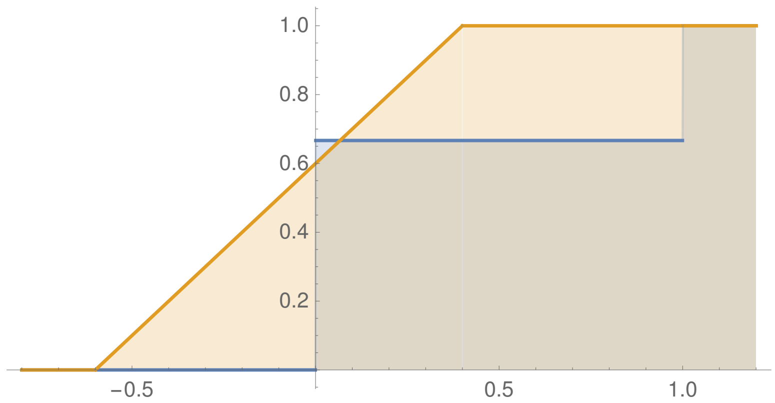

Figure 1: The c.d.f.s of X (blue) and Y (orange).

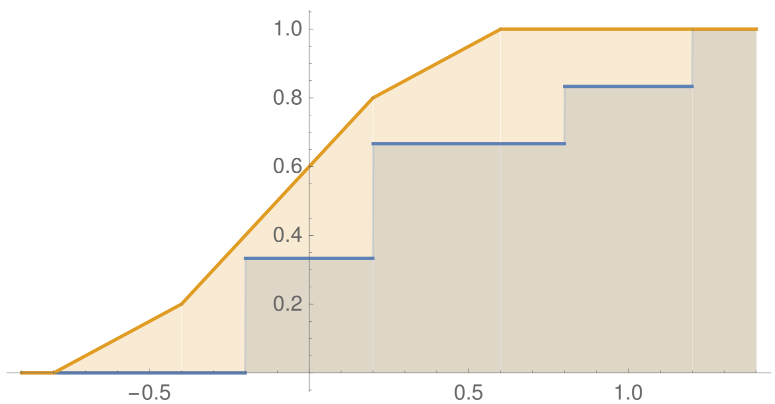

Figure 2: The c.d.f.s of $X + Z$ (blue) and $Y + Z$ (orange).

and since the same holds for the c.d.f. of Y, the result of adding Z is the disappearance of the small "kink" in which the ranking of the c.d.f.s is reversed. This is depicted in Figure 2.

Every monotone additive statistic provides an obstruction to dominance in the catalytic order. That is, if $\Phi(X) < \Phi(Y)$ for some monotone additive statistic Φ, then it is impossible that $X + Z \geq_1 Y + Z$ for some independent Z, since monotonicity would imply that $\Phi(X + Z) \geq \Phi(Y + Z)$, and additivity would then imply that $\Phi(X) \geq \Phi(Y)$. This observation applies in particular to the monotone additive statistics K_a, so that $K_a(X) \geq K_a(Y)$ for all $a \in \bar{\mathbb{R}}$ is necessary for there to exist some Z that makes X stochastically dominate Y.[7]

[7] In fact, except for the trivial case where X and Y have the same distribution, it is necessary to have the strict inequality $K_a(X) > K_a(Y)$ for all $a \in \bar{\mathbb{R}} \backslash \{\pm\infty\}$. This is because $X + Z \geq_1 Y + Z$ implies the strict inequality $K_a(X + Z) > K_a(Y + Z)$ whenever $X + Z$ and $Y + Z$ have different distributions. Thus, a corollary of Theorem 3 below is that for distributions with different minima and maxima, the condition

The following result shows that the statistics K_a are, in a sense, the only obstructions.[8] This constitutes the most important component of the proof of Theorem 1.

Theorem 3. *Let $X, Y \in L^\infty$ satisfy $K_a(X) > K_a(Y)$ for all $a \in \bar{\mathbb{R}}$. Then there exists an independent $Z \in L^\infty$ such that $X + Z \geq_1 Y + Z$.*

To prove Theorem 3 we explicitly construct Z as a truncated Gaussian with appropriately chosen parameters. The idea behind the proof is the following. Denote by F and G the c.d.f.s of X and Y, respectively, and suppose that they are supported on $[-N, N]$. Let $h(x) = \frac{1}{\sqrt{2\pi V}}\mathrm{e}^{-\frac{x^2}{2V}}$ be the density of a Gaussian Z. Then the c.d.f.s of $X + Z$ and $Y + Z$ are given by the convolutions $F * h$ and $G * h$, and their difference is equal to

$$\begin{aligned}[G * h - F * h](y) &= \int_{-N}^{N} [G(x) - F(x)] \cdot h(y - x)\,\mathrm{d}x \\ &= \frac{1}{\sqrt{2\pi V}}\mathrm{e}^{-\frac{y^2}{2V}} \cdot \int_{-N}^{N} \underbrace{[G(x) - F(x)] \cdot \mathrm{e}^{\frac{y}{V}\cdot x}}_{(*)} \cdot \underbrace{\mathrm{e}^{-\frac{x^2}{2V}}}_{(**)}\,\mathrm{d}x\end{aligned}$$

If we denote $a = \frac{y}{V}$, then by integration by parts, the integral of just $(*)$ is equal to $\frac{1}{a}\left(\mathbb{E}\left[\mathrm{e}^{aX}\right] - \mathbb{E}\left[\mathrm{e}^{aY}\right]\right)$, which is positive by the assumption that $K_a(X) > K_a(Y)$ and is in fact bounded away from zero. The term $(**)$ can be made arbitrarily close to 1—uniformly on the intergral domain $[-N, N]$—by making V large. This implies that $[G * h - F * h](y) \geq 0$ for all y, and we further show that the inequality still holds if we modify Z by truncating its tails, ensuring that it is in L^∞.

Theorem 3 allows us to prove the following key lemma:

Lemma 1. *Let $\Phi\colon L^\infty \to \mathbb{R}$ be a monotone additive statistic. If $K_a(X) \geq K_a(Y)$ for all $a \in \bar{\mathbb{R}}$ then $\Phi(X) \geq \Phi(Y)$.*

Proof. Suppose $K_a(X) \geq K_a(Y)$ for all $a \in \bar{\mathbb{R}}$. For any $\varepsilon > 0$, consider $\hat{X} = X + \varepsilon$. Then $K_a(\hat{X}) = K_a(X) + \varepsilon > K_a(Y)$ for all a, and by Theorem 3 there is an independent $Z \in L^\infty$ such that $\hat{X} + Z \geq_1 Y + Z$. Hence, by monotonicity of Φ, $\Phi(\hat{X} + Z) \geq \Phi(Y + Z)$, and by additivity $\Phi(\hat{X}) \geq \Phi(Y)$. This means that $\Phi(X) + \varepsilon \geq \Phi(Y)$ for all $\varepsilon > 0$, and hence $\Phi(X) \geq \Phi(Y)$. □

Once we have established Lemma 1, the remainder of the proof of Theorem 1 uses functional analysis techniques (in particular the Riesz Representation Theorem) to deduce the integral representation for monotone additive statistics.

$K_a(X) > K_a(Y)$ for all $a \in \bar{\mathbb{R}}$ is necessary and sufficient for dominance in the catalytic order on L^∞.

[8] A similar result to Theorem 3 holds if we demand a weaker conclusion that $X + Z$ *second-order* stochastically dominates $Y + Z$. See Proposition 5 in the appendix.

3 Monotone Stationary Time Preferences

3.1 Domain and axioms

We model a *time lottery* by a pair (x, T), which consists of a non-negative payoff $x \in \mathbb{R}_+$ and a bounded non-negative random time $T \in L^\infty_+$ at which this payoff realizes. Thus time is non-negative and continuous in this section. Our primitive is a weak order $\succeq$ on the domain $\mathbb{R}_+ \times L^\infty_+$. We denote by $\sim$ the indifference relation induced by $\succeq$. To avoid notational confusion, in the rest of this section x and y always denote monetary payoffs, t, s and d always denote deterministic times, and capitalized letters $T, S, D, R \in L^\infty_+$ always denote random times.

We impose the following four axioms on $\succeq$:

Axiom 3.1 (More is Better)**.** *If $x > y \geq 0$ then $(x, T) \succ (y, T)$ for all $T \in L^\infty_+$.*

Axiom 3.2 (Earlier is Better)**.** *If $S \geq_1 T$ in first-order stochastic dominance, then $(x, T) \succeq (x, S)$ for all $x \geq 0$. Indifference obtains if $x = 0$, and strict preference obtains if $x > 0$ and $S > T$ are deterministic times.*

Axiom 3.3 (Stationarity)**.** *If $(x, T) \succeq (y, S)$ then $(x, T + D) \succeq (y, S + D)$ for any $D \in L^\infty_+$ that is independent from T and S.*

Axiom 3.4 (Continuity)**.** *For any (y, S), the sets $\{(x, t) : (x, t) \succeq (y, S)\}$ and $\{(x, t) : (x, t) \preceq (y, S)\}$ are closed in the product topology on $\mathbb{R}_+ \times \mathbb{R}_+$.*

3.2 Discussion of the axioms

The first two axioms and the continuity axiom are standard conditions that directly generalize the axioms in Fishburn and Rubinstein (1982). Axioms 3.1 and 3.2 require the decision maker to prefer higher payoffs, and to prefer (stochastically) earlier times. The continuity assumption is similarly standard. Note that it does not require a choice of topology for L^∞_+, the set of random times.

The most substantive condition is stationarity. In the absence of risk, it was shown by Halevy (2015) that stationarity can be understood as the implication of two more basic principles: that preferences are not affected by calendar time, and that the decision maker is dynamically consistent. We now argue that Axiom 3.3 extends the same logic to time lotteries.

First, suppose that the time D, which is a delay added to both S and T, is deterministic. Reasoning as in Halevy (2015), we can consider an enlarged framework where the decision maker is endowed with a profile $(\succeq_t)$ of preferences over time lotteries, with $\succeq_t$ representing the preference the decision maker expresses at time t.

If preferences are not affected by calendar time, then the ranking $(x, T) \succeq_0 (y, S)$ at time zero must imply the same ranking $(x, T + d) \succeq_d (y, S + d)$ at time d. Moreover, dynamic consistency requires that a choice between $(x, T+d)$ and $(x, T+d)$, when evaluated at time zero, must be the same choice the decision maker would in fact make at time d. Hence, $(x, T) \succeq_0 (y, S)$ implies $(x, T+d) \succeq_0 (y, S+d)$, as required by Axiom 3.3.

Suppose now that the delay D is random and independent of S and T. By the above reasoning, dynamic consistency and time invariance imply the ranking $(x, T+d) \succeq_0 (y, S+d)$ for each deterministic time d. We can then imagine that prior to making a choice between $(x, T+D)$ and $(y, S+D)$, the decision maker is informed of the actual realization d of D. Regardless of what the value d is, this information should not change the decision maker's preference of $(x, T+d)$ over $(y, S+d)$, since D is independent of T and S. So, dynamic consistency with respect to this piece of information requires the decision maker to prefer $(x, T+D)$ to $(y, S+D)$. While this latter form of dynamic consistency is suggestive of expected utility, we will in fact derive non-expected utility representations that also satisfy this consistency condition.

3.3 Representation

We say that a preferences $\succeq$ on $\mathbb{R}_+ \times L^\infty_+$ is a *monotone stationary preference* if it satisfies Axioms 3.1, 3.2, 3.3 and 3.4. We say that $\succeq$ is represented by $f \colon \mathbb{R}_+ \times L^\infty_+ \to \mathbb{R}$ if

$$(x, T) \succeq (y, S) \quad \text{if and only if} \quad f(x, T) \geq f(y, S).$$

Our main result in this section is stated as follows:

Theorem 4. *A preference $\succeq$ on $\mathbb{R}_+ \times L^\infty_+$ is a monotone stationary preference if and only if there exists a monotone additive statistic Φ, an $r > 0$, and a continuous and strictly increasing utility function $u : \mathbb{R}_+ \to \mathbb{R}_+$ with $u(0) = 0$, such that $\succeq$ is represented by*

$$f(x, T) = u(x) \cdot \mathrm{e}^{-r\Phi(T)}. \tag{4}$$

The coefficient r represents the discount factor the decision maker applies when evaluating riskless date rewards. As in Fishburn and Rubinstein (1982), it can be chosen arbitrarily by a suitable normalization of u.

By Theorem 2, we can conclude that every monotone stationary preference has a representation of the form

$$f(x, T) = u(x) \cdot \mathrm{e}^{-r \int K_a(T) \, \mathrm{d}\mu(a)}. \tag{5}$$

The result can be interpreted as saying that the decision maker evaluates the pair (x, T) in a multiplicatively separable way, by discounting the utility from x by an appropriate factor that depends only on T. The discount factor can be expressed in terms of the certainty equivalent of T, which we denote by $\Phi(T)$. Furthermore, Φ is a monotone additive statistic, and so its form is pinned down by Theorem 2. This form (5) implies that the random time T is evaluated as the average of certainty equivalents of different exponential discounters.

Regardless of the particular statistic Φ that enters the representation, the preference $\succeq$ when restricted to deterministic dated rewards is represented by exponentially discounted utility. However, Theorem 4 demonstrates there are many ways to extend exponentially discounted utility to the larger domain of time lotteries, while maintaining stationarity. We recover expected exponentially discounted utility if $\Phi(T) = \frac{1}{-r} \log \mathbb{E}\left[\mathrm{e}^{-rT}\right]$, since in that case

$$f(x,T) = u(x) \cdot \mathbb{E}\left[\mathrm{e}^{-rT}\right].$$

As is well known, such preferences are *risk-seeking over time.*

But any monotone additive statistic Φ gives rise to a stationary time preference via the utility representation in Theorem 4, and such a preference need not be either EU or risk-seeking. As an example, if $\Phi(T) = K_r(T) = \frac{1}{r} \log \mathbb{E}\left[\mathrm{e}^{rT}\right]$ then we get a new representation

$$f(x,T) = \frac{u(x)}{\mathbb{E}\left[\mathrm{e}^{rT}\right]},$$

which is in fact *risk-averse over time.* In a later subsection we characterize the precise conditions on the measure μ such that the resulting time preference is risk-seeking, or risk-averse.

The idea behind the proof of Theorem 4 is as follows. For fixed x, the continuity axiom ensures that there is a certainty equivalent function Φ_x such that $(x,T) \sim (x, \Phi_x(T))$ for all $T \in L^\infty_+$. The monotonicity of Φ_x is a simple consequence of the first axiom. To see that Φ_x is additive, we apply stationarity twice. First, stationarity implies that $(x, \Phi_x(T) + \Phi_x(S)) \sim (x, T + \Phi_x(S))$, with the constant $\Phi_x(S)$ playing the role of D. Likewise, stationarity also implies that $(x, \Phi_x(S) + T) \sim (x, S + T)$, where now T plays the role of D. Put together, these imply that $(x, \Phi_x(T) + \Phi_x(S)) \sim (x, \Phi_x(T + S))$, and so $\Phi_x(T + S) = \Phi_x(T) + \Phi_x(S)$. A third application of the stationarity axiom yields that $\Phi_x = \Phi_y$ for every $x, y > 0$, which allows us to write Φ instead of Φ_x. Finally, the representation $u(x)\mathrm{e}^{-r\Phi(T)}$ follows by applying the original result of Fishburn and Rubinstein (1982).

This proof, and the representation in Theorem 4, can be extended to a discrete-time setting. However, one difficulty that arises is that a discrete time lottery need not have a certainty equivalent that is an integer time. Because of this, we need additional work to reduce each time lottery to a deterministic dated reward in order to apply the result of Fishburn and Rubinstein (1982). See Appendix D.2 for details.

3.4 Further axioms

As we mentioned in previous discussion, our representation $f(x,T) = u(x) \cdot \mathrm{e}^{-r\Phi(T)}$ is in general non-expected utility. In this subsection we study the extent to which this representation violates the behavioral assumptions of EU. We begin by investigating the *betweenness* axiom of Dekel (1986), which relaxes expected utility theory by requiring that indifference curves are straight lines (but need not be parallel).

We use the notation $X_\lambda Y$ to denote a random variable that is equal to X with probability $\lambda \in [0,1]$ and equal to Y with probability $1-\lambda$. Equivalently, if the distribution of X is μ and the distribution of Y is ν, then the distribution of $X_\lambda Y$ is $\lambda\mu + (1-\lambda)\nu$.[9]

Axiom 3.5 (Betweenness)**.** *$(x,T) \sim (x,S)$ implies $(x,T_\lambda S) \sim (x,S)$ for all $\lambda \in (0,1)$.*

The next result characterizes monotone stationary preferences that have this property.

Proposition 1. *A monotone stationary preference with representation $f(x,T) = u(x)\mathrm{e}^{-r\Phi(T)}$ satisfies Axiom 3.5 if and only if*

1. *$\Phi(T) = K_a(T)$ for some $a \in \bar{\mathbb{R}}$, or*
2. *$\Phi(T) = \beta \min[T] + (1-\beta)\max[T]$ for some $\beta \in (0,1)$, or*
3. *$\Phi(T) = \frac{-a_1}{a_2-a_1}K_{a_1}(T) + \frac{a_2}{a_2-a_1}K_{a_2}(T)$ for some $a_1 \in (-\infty,0)$ and $a_2 \in (0,\infty)$.*

In fact, our proof shows that Proposition 1 holds under a weaker form of betweenness: $(x,T) \sim (x,t)$ implies $(x,T_\lambda t) \sim (x,t)$. That is, it suffices to require betweenness when mixing with constants.

Next, we study the classic independence axiom underlying expected utility theory.

Axiom 3.6 (Independence)**.** *$(x,T) \sim (x,S)$ implies $(x,T_\lambda R) \sim (x,S_\lambda R)$.*

The space of time lotteries is not a mixture space, so we only impose independence for random times associated with the same monetary reward. We do not impose continuity beyond Axiom 3.4.

The following result characterizes monotone stationary preferences that additionally satisfy independence:

Proposition 2. *A monotone stationary preference with representation $f(x,T) = u(x)\mathrm{e}^{-r\Phi(T)}$ satisfies Axiom 3.6 if and only if $\Phi(T) = K_a(T)$ for some $a \in \bar{\mathbb{R}}$.*

This proposition implies that such a preference has one of the following representations:

$$u(x)\cdot \mathrm{e}^{-r\min[T]}, \quad u(x)\cdot \mathbb{E}\left[\mathrm{e}^{-rT}\right], \quad u(x)\cdot \mathrm{e}^{-r\mathbb{E}[T]}, \quad \frac{u(x)}{\mathbb{E}\left[\mathrm{e}^{rT}\right]}, \quad u(x)\cdot \mathrm{e}^{-r\max[T]}$$

The first and last representations correspond to the most extreme forms of risk-seeking and risk-averse time preferences, respectively. Because these extreme preferences do not satisfy "mixture continuity", it is not possible to deduce Proposition 2 directly from the von Neumann Morgenstern theorem. Our proof instead invokes Proposition 1, and uses the independence axiom to further pin down the form of Φ.

[9] Of course, there are many such random variables, but for our purposes this will not be important.

3.5 Risk attitudes toward time

We have shown that monotone stationary preferences over time lotteries admit a representation of the form

$$u(x) \cdot \mathrm{e}^{-r \int K_a(T)\,\mathrm{d}\mu(a)}$$

for some probability measure μ on $\bar{\mathbb{R}}$. In this section we study which measures μ give rise to risk-averse or risk-seeking behavior toward time. For example, we have seen that when μ is a point mass on a, the preference will be risk-averse or risk-seeking depending on whether a is positive or negative.

Formally, we say that a preference $\succeq$ over time lotteries *exhibits risk aversion* if $(x, \mathbb{E}[T]) \succeq (x, T)$ for every $x \in \mathbb{R}_+$ and $T \in L^\infty_+$. If the reverse preference always holds, then $\succeq$ is risk-seeking. The following result generalizes our previous observations regarding point mass measures μ:[10]

Proposition 3. *A monotone stationary preference with representation $f(x,T) = u(x)\mathrm{e}^{-r\Phi(T)}$ is risk-averse (respectively risk-loving) over time if and only if*

$$\Phi(T) = \int_{\bar{\mathbb{R}}} K_a(T)\,\mathrm{d}\mu(a)$$

for a Borel probability measure μ supported on $[0,\infty]$ (respectively $[-\infty,0]$).

Thus, risk aversion over time occurs if and only if the decision maker aggregates the certainty equivalents of exponentially discounting EU agents with discount factors greater than or equal to 1. Likewise, risk seeking occurs if and only if the relevant discount factors in the aggregation are all less than or equal to 1.

More generally, we can compare the risk attitudes of two different monotone stationary preferences. Consider two preferences represented by $u(x)\mathrm{e}^{-r\Phi_\mu(T)}$ and $u(x)\mathrm{e}^{-r\Phi_\nu(T)}$, where Φ_μ and Φ_ν are two different monotone additive statistics with corresponding measures μ and ν. We say that the preference represented by Φ_μ is *more risk-averse than* the preference represented by Φ_ν if $\Phi_\mu(T) \geq \Phi_\nu(T)$ for every $T \in L^\infty_+$. In words, we require the former preference to assign a worse certainty equivalent (i.e., later time) to every random time T.

[10] As a corollary of the analysis in Proposition 3, we know that a statistic Φ is additive and monotone with respect to *second-order* (or any higher-order) stochastic dominance if and only if $\Phi(X) = \int_{\bar{\mathbb{R}}} K_a(X)\,\mathrm{d}\mu(a)$ for a probability measure μ supported on $[-\infty, 0]$. To see why, note that Proposition 3 shows that $\Phi(X) \leq \mathbb{E}[X]$ for all X (which is risk-seeking in time) only if the measure μ associated with Φ is supported on $[-\infty,0]$. Since $\Phi(X) \leq \Phi(\mathbb{E}[X]) = \mathbb{E}[X]$ is a necessary condition for monotonicity with respect to second-order stochastic dominance, μ being supported on $[-\infty,0]$ is also necessary. Conversely, note that for any $a \leq 0$, the function $-\mathrm{e}^{aX}$ is increasing and has alternating derivatives of all orders. So whenever X dominates Y in second-order (or any higher-order) stochastic dominance, it holds that $\mathbb{E}\left[-\mathrm{e}^{aX}\right] \geq \mathbb{E}\left[-\mathrm{e}^{aY}\right]$. From this we obtain $K_a(X) = \frac{1}{a}\mathbb{E}\left[\mathrm{e}^{aX}\right] \geq K_a(Y)$ for any $a \leq 0$, and thus $\Phi(X) = \int K_a(X)\,\mathrm{d}\mu(a)$ is larger than $\Phi(Y)$ whenever μ is supported on $[-\infty,0]$.

Under what conditions on μ and ν is the first preference more risk-averse than the second? That is, when is it the case that $\Phi_\mu(T) \geq \Phi_\nu(T)$ for all T? Since $K_a(T)$ increases in a, first-order stochastic dominance $\mu \geq_1 \nu$ is clearly sufficient, but—as we show—it is not necessary.[11] We provide an exact characterization in the following result.

Proposition 4. *For any two probability measures μ and ν on $\bar{\mathbb{R}}$, the inequality*

$$\int_{\bar{\mathbb{R}}} K_a(T)\,\mathrm{d}\mu(a) \geq \int_{\bar{\mathbb{R}}} K_a(T)\,\mathrm{d}\nu(a)$$

holds for every $T \in L^\infty_+$ if and only if the following two conditions hold:

(i) For every $b > 0$, $\int_{[b,\infty]} \frac{a-b}{a}\,\mathrm{d}\mu(a) \geq \int_{[b,\infty]} \frac{a-b}{a}\,\mathrm{d}\nu(a)$.

(ii) For every $b < 0$, $\int_{[-\infty,b]} \frac{a-b}{a}\,\mathrm{d}\mu(a) \leq \int_{[-\infty,b]} \frac{a-b}{a}\,\mathrm{d}\nu(a)$.

This result can be seen as a generalization of the previous Proposition 3, since a preference exhibits risk aversion (respectively risk seeking) if and only if it is more (respectively less) risk-averse than the risk neutral preference represented by $u(x)\mathrm{e}^{-r\mathbb{E}[T]}$. In the appendix, we explain how to deduce Proposition 3 from Proposition 4.

4 Preferences over Gambles

In this section we consider bounded monetary gambles, and study preferences over these gambles, which we denote by L^∞. As above, we assume that agents' preferences for gambles depend only on their distribution.

4.1 CARA beyond expected utility

Consider an expected utility agent who evaluates a gamble X according to a certainty equivalent $\Psi(X) = u^{-1}\mathbb{E}\left[u(X)\right]$ for some increasing utility function u. As is well known, the assumption that the agent's preferences are not affected by independent background risk implies that the agent has CARA preferences. Formally, if

$$\Psi(X) \leq \Psi(Y) \;\text{ implies }\; \Psi(X+Z) \leq \Psi(Y+Z) \;\text{ for all independent } Z \tag{6}$$

then $\Psi(X) = K_a(X)$ for some $a \in \mathbb{R}$.

A natural question is: how does this result extend beyond expected utility theory? That is, which preferences on L^∞ are monotone with respect to first-order stochastic dominance, and have a certainty equivalent Φ that satisfies (6)? The answer is that such a certainty equivalent Φ must be a monotone additive statistic. To see this, note that

[11] An example is $\mu = \frac{1}{4}\delta_1 + \frac{3}{4}\delta_3$, whereas $\nu = \delta_2$. Condition (ii) in Proposition 4 is trivially satisfied, whereas condition (i) reduces to $\frac{1}{4}(1-b)^+ + \frac{1}{4}(3-b)^+ \geq \frac{1}{2}(2-b)^+$, which holds because the function $(a-b)^+ = \max\{a-b, 0\}$ is convex in a.

$\Phi(X) = \Phi(\Phi(X))$ since the certainty equivalent to the constant $\Phi(X)$ is itself. Thus by (6), we have

$$\Phi(X + Y) = \Phi(\Phi(X) + Y),$$

with Y playing the role of Z there. Likewise, since $\Phi(Y) = \Phi(\Phi(Y))$, (6) gives

$$\Phi(Y + \Phi(X)) = \Phi(\Phi(Y) + \Phi(X))$$

where now the constant $\Phi(X)$ takes the role of Z. Combining the above two equalities yields

$$\Phi(X + Y) = \Phi(\Phi(Y) + \Phi(X)) = \Phi(X) + \Phi(Y),$$

so Φ is additive.

Given this, Theorem 1 implies that any monotone preference that is represented by a certainty equivalent and is invariant to background risk must have a representation of the form

$$\Phi(X) = \int_{\bar{\mathbb{R}}} K_a(X)\,\mathrm{d}\mu(a)$$

for some measure μ on $\bar{\mathbb{R}}$. That is, the certainty equivalent Φ is a weighted average of the certainty equivalents of CARA agents.

4.2 Narrow framing and stochastically dominated choices

Rabin and Weizsäcker (2009) show that for any non-CARA expected utility decision maker, one can construct two pairs of bounded gambles X_1, Y_1 and X_2, Y_2, such that X_1 is chosen over Y_1, X_2 is chosen over Y_2, but the independent sum $X_1 + X_2$ is first-order stochastically dominated by $Y_1 + Y_2$.[12] This result suggests that for "most" EU agents, choosing between risky aspects in isolation can lead to stochastically dominated *combined* choices. In this section we study the extent to which their insight generalizes to non-EU preferences.

Accordingly, our primitive here is a weak order $\succeq$ on L^∞, the space of bounded gambles. As is standard, we write $\succ$ for the strict part of $\succeq$, and $\sim$ for the induced indifference relation. We consider the following axioms on $\succeq$:

Axiom 4.1 (Rabin and Weizsäcker)**.** *Suppose X_1, X_2 are independent and Y_1, Y_2 are independent. If $X_1 \succ Y_1$ and $X_2 \succ Y_2$, then $X_1 + X_2 \not<_1 Y_1 + Y_2$.*

Axiom 4.2 (Responsiveness)**.** *$X + \varepsilon \succ X$ for any X and any $\varepsilon > 0$.*

[12] In Rabin and Weizsäcker (2009) the constructed gambles have binary support. Since we seek to analyze all non-EU preferences, we will allow for bounded gambles.

Axiom 4.3 (Archimedeanity)**.** *If $c + \varepsilon \succ X \succ c - \varepsilon$ for some constant c and all $\varepsilon > 0$, then $X \sim c$.*

Theorem 5. *A preference $\succeq$ on L^∞ satisfies Axioms 4.1, 4.2 and 4.3 if and only if it can be represented by a monotone additive statistic Φ (i.e., $X \succeq Y$ if and only if $\Phi(X) \geq \Phi(Y)$).*

We make a technical remark that for this result to hold, the responsiveness axiom cannot be dropped in general. An example is where $X \succeq Y$ if and only if $\max\{\mathbb{E}[X], 0\} \geq \max\{\mathbb{E}[Y], 0\}$. This preference satisfies the Rabin and Weizsäcker axiom because $X_1 \succ Y_1$ and $X_2 \succ Y_2$ imply $\mathbb{E}[X_1] > \mathbb{E}[Y_1]$ and $\mathbb{E}[X_2] > \mathbb{E}[Y_2]$. So $\mathbb{E}[X_1 + X_2] > \mathbb{E}[Y_1 + Y_2]$ and $X_1 + X_2$ cannot be stochastically dominated by $Y_1 + Y_2$. Archimedeanity is also satisfied, but responsiveness fails.

Archimedeanity (which plays the role of continuity) cannot be dropped either, since it helps ruling out lexicographic preferences. An example is where $X \succeq Y$ if and only if $\max[X] > \max[Y]$, or $\max[X] = \max[Y]$ and $\min[X] \geq \min[Y]$. This preference satisfies the Rabin and Weizsäcker axiom as well as responsiveness, but archimedeanity fails.

References

C. D. Aliprantis and K. Border. *Infinite dimensional analysis: A hitchhiker's guide.* Springer, 2006.

P. Bickel and E. Lehmann. Descriptive statistics for nonparametric models i. introduction. *Annals of Statistics*, 3(5):1038–1044, 1975a.

P. Bickel and E. Lehmann. Descriptive statistics for nonparametric models ii. location. *Annals of Statistics*, 3(5):1045–1069, 1975b.

S. Cerreia-Vioglio, D. Dillenberger, and P. Ortoleva. Cautious expected utility and the certainty effect. *Econometrica*, 83(2):693–728, 2015.

H. W. Chesson and W. K. Viscusi. Commonalities in time and ambiguity aversion for long-term risks. *Theory and Decision*, 54(1):57–71, 2003.

J. H. Curtiss. A note on the theory of moment generating functions. *The Annals of Mathematical Statistics*, 13(4):430–433, 1942.

B. de Finetti. *Theory of Probability.* Wiley, 1970.

P. DeJarnette, D. Dillenberger, D. Gottlieb, and P. Ortoleva. Time lotteries and stochastic impatience. *Econometrica*, 88(2):619–656, 2020.

E. Dekel. An axiomatic characterization of preferences under uncertainty: Weakening the independence axiom. *Journal of Economic Theory*, 40:304–318, 1986.

S. Ebert. Decision making when things are only a matter of time. *Operations Research*, 2020.

P. C. Fishburn and A. Rubinstein. Time preference. *International Economic Review*, 23: 677–694, 1982.

H. Föllmer and A. Schied. *Stochastic finance: an introduction in discrete time.* Walter de Gruyter, 2011.

T. Fritz, X. Mu, and O. Tamuz. Monotone homomorphisms on convolution semigroups, 2020. Working Paper.

M. J. Goovaerts, R. Kaas, R. J. Laeven, and Q. Tang. A comonotonic image of independence for additive risk measures. *Insurance: Mathematics and Economics*, 35(3):581–594, 2004.

Y. Halevy. Time consistency: Stationarity and time invariance. *Econometrica*, 83(1): 335–352, 2015.

M. O. Jackson and L. Yariv. The non-existence of representative agents. 2020. Working Paper.

L. Kantorovich. On the moment problem for a finite interval. In *Dokl. Akad. Nauk SSSR*, volume 14, pages 531–537, 1937.

L. Mattner. What are cumulants? *Documenta Mathematica*, 4:601–622, 1999.

L. Mattner. Cumulants are universal homomorphisms into hausdorff groups. *Probability theory and related fields*, 130(2):151–166, 2004.

X. Mu, L. Pomatto, P. Strack, and O. Tamuz. From Blackwell dominance in large samples to Rényi divergences and back again. *Econometrica*, 89(1):475–506, 2021.

S. Onay and A. Öncüler. Intertemporal choice under timing risk: An experimental approach. *Journal of Risk and Uncertainty*, 34(2):99–121, 2007.

L. Pomatto, P. Strack, and O. Tamuz. Stochastic dominance under independent noise. *Journal of Political Economy*, 128(5):1877–1900, 2020.

M. Rabin and G. Weizsäcker. Narrow bracketing and dominated choices. *American Economic Review*, 99(4):1508–43, 2009.

I. Ruzsa and G. J. Székely. *Algebraic probability theory.* John Wiley & Sons Inc, 1988.

Appendix

In the proofs we often use the notation

$$K_X(a) = K_a(X),$$

so that K_X is a map $\bar{\mathbb{R}} \to \mathbb{R}$. The following lemma is standard.

Lemma 2. *Let $X, Y \in L^\infty$.*

1. *$K_X \colon \bar{\mathbb{R}} \to \mathbb{R}$ is well defined, non-decreasing and continuous.*
2. *If $K_X = K_Y$ then X and Y have the same distribution.*

Proof. See Curtiss (1942). □

A Proof of Theorem 3

First, we can add the same constant b to both X and Y so that $\min[Y + b] = -N$ and $\max[X + b] = N$ for some $N > 0$. Since translating both X and Y leaves the existence of an appropriate Z unchanged (and also does not affect $K_X > K_Y$), we henceforth assume without loss of generality that $\min[Y] = -N$, and $\max[X] = N$. Since $K_X > K_Y$, we know that $\min[X] > -N$ and $\max[Y] < N$.

Denote the c.d.f.s of X and Y by F and G, respectively. Let $\sigma(x) = G(x) - F(x)$. Note that σ is supported on $[-N, N]$ and bounded in absolute value by 1. Moreover, by choosing $\varepsilon > 0$ sufficiently small, we have that $\min[X] > -N + \varepsilon$ and $\max[Y] < N - \varepsilon$. So $\sigma(x)$ is positive on $[-N, -N + \varepsilon]$ and on $[N - \varepsilon, N]$. In fact, there exists $\delta > 0$ such that $\sigma(x) \geq \delta$ whenever $x \in [-N + \frac{\varepsilon}{4}, -N + \frac{\varepsilon}{2}]$ and $x \in [N - \frac{\varepsilon}{2}, N - \frac{\varepsilon}{4}]$. We also fix a large constant A such that

$$\mathrm{e}^{\frac{\varepsilon A}{4}} \geq \frac{8N}{\varepsilon\delta}.$$

Define

$$M_\sigma(a) = \int_{-N}^{N} \sigma(x)\mathrm{e}^{ax}\,\mathrm{d}x.$$

Note that for $a \neq 0$,

$$M_\sigma(a) = \frac{1}{a}\left(\mathbb{E}\left[\mathrm{e}^{aX}\right] - \mathbb{E}\left[\mathrm{e}^{aY}\right]\right),$$

which follows from integration by parts, and that

$$M_\sigma(0) = \mathbb{E}[X] - \mathbb{E}[Y].$$

Therefore, since $K_X > K_Y$, we have that M_σ is strictly positive everywhere. Since $M_\sigma(a)$ is clearly continuous in a, it is in fact bounded away from zero on any compact interval.

We will use these properties of σ to construct a truncated Gaussian density h such that

$$[\sigma * h](y) = \int_{-N}^{N} \sigma(x)h(y-x)\,\mathrm{d}x \geq 0$$

for each $y \in \mathbb{R}$. If we let Z be a random variable independent from X and Y, whose distribution has density function h, then $\sigma * h = (G - F) * h$ is the difference between the c.d.f.s of $Y + Z$ and $X + Z$. Thus $[\sigma * h](y) \geq 0$ for all y would imply $X + Z \geq_1 Y + Z$.

To do this, we write $h(x) = \mathrm{e}^{-\frac{x^2}{2V}}$ for all $|x| \leq T$, where V is the variance and T is the truncation point to be chosen.[13] We will show that $[\sigma * h](y) \geq 0$ holds for each y whenever V is sufficiently large and $T \geq AV + N$ for the constants N and A defined above.

First consider the case where $y \in [-AV, AV]$. In this region, $|y-x| \leq T$ is automatically satisfied when $x \in [-N, N]$. So we can compute the convolution $\sigma * h$ as follows:

$$\int \sigma(x)h(y-x)\,\mathrm{d}x = \mathrm{e}^{-\frac{y^2}{2V}} \cdot \int_{-N}^{N} \sigma(x) \cdot \mathrm{e}^{\frac{y}{V}\cdot x} \cdot \mathrm{e}^{-\frac{x^2}{2V}}\,\mathrm{d}x. \tag{7}$$

Note that $\frac{y}{V}$ in the exponent belongs to the compact interval $[-A, A]$. So for our fixed choice of A, the integral $M_\sigma(\frac{y}{V}) = \int_{-N}^{N} \sigma(x) \cdot \mathrm{e}^{\frac{y}{V}\cdot x}\,\mathrm{d}x$ is uniformly bounded away from zero when y varies in the current region. Thus,

$$\begin{aligned}\int_{-N}^{N} \sigma(x) \cdot \mathrm{e}^{\frac{y}{V}\cdot x} \cdot \mathrm{e}^{-\frac{x^2}{2V}}\,\mathrm{d}x &= M_\sigma(\frac{y}{V}) - \int_{-N}^{N} \sigma(x) \cdot \mathrm{e}^{\frac{y}{V}\cdot x} \cdot (1 - \mathrm{e}^{-\frac{x^2}{2V}})\,\mathrm{d}x \\ &\geq M_\sigma(\frac{y}{V}) - 2N \cdot \mathrm{e}^{AN} \cdot (1 - \mathrm{e}^{\frac{-N^2}{2V}}),\end{aligned} \tag{8}$$

which is positive when V is sufficiently large. So the right-hand side of (7) is positive.

Next consider the case where $y \in (AV, T + N - \varepsilon]$; the case where $-y$ is in this range can be treated symmetrically. Here the convolution can be written as

$$[\sigma * h](y) = \int_{\max\{-N, y-T\}}^{N} \sigma(x) \cdot \mathrm{e}^{\frac{-(y-x)^2}{2V}}\,\mathrm{d}x.$$

We break the range of integration into two sub-intervals: $I_1 = [\max\{-N, y-T\}, N - \varepsilon]$ and $I_2 = [N - \varepsilon, N]$. On I_1 we have $\sigma(x) = G(x) - F(x) \geq -1$, so

$$\int_{x \in I_1} \sigma(x) \cdot \mathrm{e}^{\frac{-(y-x)^2}{2V}}\,\mathrm{d}x \geq -2N \cdot \mathrm{e}^{\frac{-(y-N+\varepsilon)^2}{2V}}.$$

On I_2 we have $\sigma(x) \geq 0$ by our choice of ε, and furthermore $\sigma(x) \geq \delta$ when $x \in [N - \frac{\varepsilon}{2}, N - \frac{\varepsilon}{4}]$. Thus

$$\int_{x \in I_2} \sigma(x) \cdot \mathrm{e}^{\frac{-(y-x)^2}{2V}}\,\mathrm{d}x \geq \frac{\varepsilon}{4} \cdot \delta \cdot \mathrm{e}^{\frac{-(y-N+\frac{\varepsilon}{2})^2}{2V}} \geq 2N \cdot \mathrm{e}^{\frac{-(y-N+\frac{\varepsilon}{2})^2}{2V} - \frac{\varepsilon A}{4}},$$

[13] In general we need a normalizing factor to ensure h integrates to one, but this multiplicative constant does not affect the argument.

where the second inequality holds by the choice of A. Observe that when $y > AV$ and V is large, the exponent $\frac{-(y-N+\frac{\varepsilon}{2})^2}{2V} - \frac{\varepsilon A}{4}$ is larger than $\frac{-(y-N+\varepsilon)^2}{2V}$. Summing the above two inequalities then yields the desired result that $[\sigma * h](y) \geq 0$.

Finally, if $y \in (T+N-\varepsilon, T+N]$, then the range of integration in computing $[\sigma * h](y)$ is from $x = y-T$ to $x = N$, where $\sigma(x)$ is always positive. So the convolution is positive. And if $y > T+N$, then clearly the convolution is zero. These arguments symmetrically apply to $-y \in (T+N-\varepsilon, T+N]$ and $-y > T+N$. We therefore conclude that $[\sigma * h](y) \geq 0$ for all y, completing the proof.

A.1 The catalytic order for second-order stochastic dominance

In this section we point out that the above proof of Theorem 3 also yields an analogous characterization of the catalytic stochastic order for second-order stochastic dominance. Formally, we have

Proposition 5. *Let $X, Y \in L^\infty$ satisfy $K_a(X) > K_a(Y)$ for all $a \in [-\infty, 0]$. Then there exists an independent $Z \in L^\infty$ such that $X + Z \geq_2 Y + Z$.*

Proof. As is well known, X dominates Y in second-order stochastic dominance if and only if their c.d.f.s satisfy

$$\int_{-\infty}^{z} (G(y) - F(y))\,\mathrm{d}y \geq 0$$

for every $z \in \mathbb{R}$. Thus, if we let Z be an independent random variable with density h, then $X + Z \geq_2 Y + Z$ if and only if

$$\int_{-\infty}^{z} [\sigma * h](y)\,\mathrm{d}y \geq 0 \quad \forall z \in \mathbb{R}.$$

Here, as in the proof of Theorem 3, σ denotes the difference $G - F$ and is supported on $[-N, N]$. Since $K_{-\infty}(X) > K_{-\infty}(Y)$, we have $\min[X] > \min[Y]$. So we can choose $\varepsilon, \delta > 0$ such that $\sigma(x) \geq 0$ for $x \in [-N, -N+\varepsilon]$ and $\sigma(x) \geq \delta$ for $x \in [-N+\frac{\varepsilon}{4}, -N+\frac{\varepsilon}{2}]$. We again fix constant A such that $\mathrm{e}^{\frac{\varepsilon A}{4}} \geq \frac{8N}{\varepsilon\delta}$.

Now let $h(x) = \mathrm{e}^{\frac{-x^2}{2V}}$ for $|x| \leq T$, where V is a large variance and $T = AV + N$. Then, as in the proof of Theorem 3, we have

$$[\sigma * h](y) \geq 0 \quad \forall y \leq -AV.$$

This simply uses the fact that σ is positive near the minimum of its support.

Moreover, by assumption $K_a(X) > K_a(Y)$ for $a \leq 0$. So by continuity there exists small $\gamma > 0$ such that $K_a(X) > K_a(Y)$ for $a \leq \gamma$. It follows that

$$M_\sigma(a) = \int_{-N}^{N} \sigma(x)\mathrm{e}^{ax}\,\mathrm{d}x = \frac{1}{a}\left(\mathbb{E}\left[\mathrm{e}^{aX}\right] - \mathbb{E}\left[\mathrm{e}^{aY}\right]\right) > 0 \quad \forall a \leq \gamma.$$

By continuity, we can find $\eta > 0$ such that

$$M_\sigma(a) \geq \eta \quad \forall a \in [-A, \gamma].$$

Thus, when $y \in [-AV, \gamma V]$, we can follow the calculation in (7) and (8) to obtain

$$\begin{aligned}[\sigma * h](y) &= \mathrm{e}^{-\frac{y^2}{2V}} \cdot \int_{-N}^{N} \sigma(x) \cdot \mathrm{e}^{\frac{y}{V}\cdot x} \cdot \mathrm{e}^{-\frac{x^2}{2V}}\,\mathrm{d}x \\ &\geq \mathrm{e}^{-\frac{y^2}{2V}} \cdot \left(\eta - 2N \cdot \mathrm{e}^{AN} \cdot (1 - \mathrm{e}^{\frac{-N^2}{2V}})\right) \\ &\geq \mathrm{e}^{-\frac{y^2}{2V}} \cdot \frac{\eta}{2},\end{aligned}$$

where the last step holds when V is sufficiently large.

Therefore, $[\sigma*h](y) \geq 0$ for all $y \leq \gamma V$, and clearly $\int_{-\infty}^{z}[\sigma*h](y)\,\mathrm{d}y \geq 0$ also holds for $z \leq \gamma V$. Below we consider $z > \gamma V$. The idea here is that $\int_{-\infty}^{\gamma V}[\sigma*h](y)\,\mathrm{d}y$ is sufficiently positive to compensate for the possible negative contribution from $\int_{\gamma V}^{z}[\sigma * h](y)\,\mathrm{d}y$. Specifically, using the above lower bound for $[\sigma * h](y)$, we have

$$\int_{-\infty}^{\gamma V}[\sigma * h](y)\,\mathrm{d}y \geq \int_{-\frac{\gamma V}{2}}^{\frac{\gamma V}{2}}[\sigma * h](y)\,\mathrm{d}y \geq \frac{\eta\gamma V}{2} \cdot \mathrm{e}^{-\frac{\gamma^2 V}{8}}.$$

On the other hand, when $y > \gamma V$ we can bound the magnitude of $[\sigma * h](y)$ as follows:

$$|[\sigma * h](y)| \leq \int_{-N}^{N} |\sigma(x)h(y-x)|\,\mathrm{d}x \leq \int_{-N}^{N} \mathrm{e}^{-\frac{(y-x)^2}{2V}}\,\mathrm{d}x \leq 2N \cdot \mathrm{e}^{-\frac{(\gamma V - N)^2}{2V}}$$

Since σ is supported on $[-N, N]$ and h is supported on $[-AV - N, AV + N]$, we know that $\sigma * h$ is supported on $[-AV - 2N, AV + 2N]$. Thus for $z > \gamma V$,

$$\int_{\gamma V}^{z}[\sigma * h](y)\,\mathrm{d}y \geq -\int_{\gamma V}^{AV+2N} |[\sigma * h](y)|\,\mathrm{d}y \geq -(AV + 2N - \gamma V) \cdot \mathrm{e}^{-\frac{(\gamma V - N)^2}{2V}}.$$

It is easy to see that for sufficiently large V,

$$\frac{\eta\gamma V}{2} \cdot \mathrm{e}^{-\frac{\gamma^2 V}{8}} > (AV + 2N - \gamma V) \cdot \mathrm{e}^{-\frac{(\gamma V - N)^2}{2V}}.$$

Hence the above estimates imply that $\int_{-\infty}^{z}[\sigma * h](y)\,\mathrm{d}y \geq 0$ also holds for $z > \gamma V$. So $X + Z \geq_2 Y + Z$ as we desire to show. □

B Proof of Theorem 1

B.1 Integral representation

Recall that for fixed X, $K_X(a) = K_a(X)$ can be seen as a function of a. Let $\mathcal{L}$ denote the set of functions $\{K_X : X \in L^\infty\}$. If Φ is a monotone additive statistic and $K_X = K_Y$, then X and Y have the same distribution and $\Phi(X) = \Phi(Y)$. Thus there exists some functional $F\colon \mathcal{L} \to \mathbb{R}$ such that $\Phi(X) = F(K_X)$. It follows from the additivity of Φ and the additivity of K_a that F is additive: $F(K_X + K_Y) = F(K_X) + F(K_Y)$. Moreover, F is monotone in the sense that $F(K_X) \geq F(K_Y)$ whenever $K_X \geq K_Y$ (i.e., $K_X(a) \geq K_Y(a)$ for all $a \in \bar{\mathbb{R}}$); this follows from Lemma 1 which in turn is proved by Theorem 3 (see the main text).

Next we show that the monotone additive functional F on $\mathcal{L}$ can be extended to a positive linear functional on the entire space of continuous functions $\mathcal{C}(\bar{\mathbb{R}})$. We first equip $\mathcal{L}$ with the sup-norm of $\mathcal{C}(\bar{\mathbb{R}})$ and establish a technical claim.

Lemma 3. *$F\colon \mathcal{L} \to \mathbb{R}$ is 1-Lipschitz:*

$$|F(K_X) - F(K_Y)| \leq \|K_X - K_Y\|.$$

Proof. Let $\|K_X - K_Y\| = \varepsilon$. Then

$$K_{X+\varepsilon} = K_X + \varepsilon \geq K_Y.$$

Hence, by Lemma 1, $F(K_Y) \leq F(K_{X+\varepsilon})$, and so

$$F(K_Y) - F(K_X) \leq F(K_{X+\varepsilon}) - F(K_X) = F(K_\varepsilon) = \Phi(\varepsilon) = \varepsilon.$$

Symmetrically we have $F(K_X) - F(K_Y) \leq \varepsilon$, as desired. □

Lemma 4. *Any monotone additive functional F on $\mathcal{L}$ can be extended to a positive linear functional on $\mathcal{C}(\bar{\mathbb{R}})$.*

Proof. First consider the rational cone spanned by $\mathcal{L}$:

$$\mathrm{Cone}_{\mathbb{Q}}(\mathcal{L}) = \{qL : q \in \mathbb{Q}_+, L \in \mathcal{L}\}.$$

Define $G : \mathrm{Cone}_{\mathbb{Q}}(\mathcal{L}) \to \mathbb{R}$ as $G(qL) = qF(L)$, which is an extension of F. The functional G is well defined: If $\frac{m}{n}K_1 = \frac{r}{n}K_2$ for $K_1, K_2 \in \mathcal{L}$ and $n, m, r \in \mathbb{N}$, then, using the fact that $\mathcal{L}$ is closed under addition, we obtain $mF(K_1) = F(mK_1) = F(rK_2) = rF(K_2)$, hence $\frac{m}{n}F(K_1) = \frac{r}{n}F(K_2)$. G is also additive, because

$$G\left(\frac{m}{n}K_1\right) + G\left(\frac{r}{n}K_2\right) = \frac{m}{n}F(K_1) + \frac{r}{n}F(K_2) = \frac{1}{n}F(mK_1 + rK_2) = G\left(\frac{m}{n}K_1 + \frac{r}{n}K_2\right).$$

In the same way we can show G is positively homogeneous over $\mathbb{Q}_+$ and monotone.

Moreover, G is Lipschitz: Lemma 3 implies

$$\left|G\left(\frac{m}{n}K_1\right) - G\left(\frac{r}{n}K_2\right)\right| = \frac{1}{n}\left|F(mK_1) - F(rK_2)\right| \leq \frac{1}{n}\|mK_1 - rK_2\| = \left\|\frac{m}{n}K_1 - \frac{r}{n}K_2\right\|.$$

Thus G can be extended to a Lipschitz functional H defined on the closure of $\mathrm{Cone}_{\mathbb{Q}}(\mathcal{L})$ with respect to the sup norm. In particular, H is defined on the convex cone spanned by $\mathcal{L}$:

$$\mathrm{Cone}(\mathcal{L}) = \{\lambda_1 K_1 + \cdots + \lambda_k K_k : k \in \mathbb{N} \text{ and for each } 1 \leq i \leq k, \lambda_i \in \mathbb{R}_+, K_i \in \mathcal{L}\}.$$

It is immediate to verify that the properties of additivity, positive homogeneity (now over $\mathbb{R}_+$), and monotonicity extend, by continuity, from G to H.

Consider the vector subspace $\mathcal{V} = \mathrm{Cone}(\mathcal{L}) - \mathrm{Cone}(\mathcal{L}) \subset \mathcal{C}(\bar{\mathbb{R}})$ and define $I : \mathcal{V} \to \mathbb{R}$ as

$$I(g_1 - g_2) = H(g_1) - H(g_2)$$

for all $g_1, g_2 \in \mathrm{Cone}(\mathcal{L})$. The functional I is well defined and linear (because H is additive and positively homogeneous). Moreover, by monotonicity of H, $I(f) \geq 0$ for any non-negative function $f \in \mathcal{V}$.

The result then follows from the next theorem of Kantorovich (1937), a generalization of the Hahn-Banach Theorem. It applies not only to $\mathcal{C}(\bar{\mathbb{R}})$ but to any Riesz space (see Theorem 8.32 in Aliprantis and Border, 2006).

Theorem. *If $\mathcal{V}$ is a vector subspace of $\mathcal{C}(\bar{\mathbb{R}})$ with the property that for every $f \in \mathcal{C}(\bar{\mathbb{R}})$ there exists a function $g \in \mathcal{V}$ such that $g \geq f$. Then every positive linear functional on $\mathcal{V}$ extends to a positive linear functional on $\mathcal{C}(\bar{\mathbb{R}})$.*

The "majorization" condition $g \geq f$ is satisfied because every function in $\mathcal{C}(\bar{\mathbb{R}})$ is bounded and $\mathcal{V}$ contains all of the constant functions. □

The integral representation in Theorem 1 now follows from Lemma 4 by the Riesz-Markov-Kakutani Representation Theorem.

B.2 Uniqueness of measure

We complete the proof of Theorem 1 by showing that the measure μ in the representation is unique. The following result shows that uniqueness holds even on the smaller domain $L^\infty_{\mathbb{N}}$ of non-negative integer-valued random variables.

Lemma 5. *Suppose μ and ν are two Borel probability measures on $\bar{\mathbb{R}}$ such that*

$$\int_{\bar{\mathbb{R}}} K_a(X)\,\mathrm{d}\mu(a) = \int_{\bar{\mathbb{R}}} K_a(X)\,\mathrm{d}\nu(a).$$

for all $X \in L^\infty_{\mathbb{N}}$. Then $\mu = \nu$.

Proof. We first show $\mu(\{\infty\}) = \nu(\{\infty\})$. For any $\varepsilon > 0$, consider the Bernoulli random variable X_ε that takes value 1 with probability ε. It is easy to see that as ε decreases to zero, $K_a(X_\varepsilon)$ also decreases to zero for each $a < \infty$ whereas $K_\infty(X_\varepsilon) = \max[X_\varepsilon] = 1$. Since $K_a(X_\varepsilon)$ is uniformly bounded in $[0, 1]$, the Dominated Convergence Theorem implies

$$\lim_{\varepsilon \to 0} \int_{\bar{\mathbb{R}}} K_a(X_\varepsilon) \, \mathrm{d}\mu(a) = \mu(\{\infty\}).$$

A similar identity holds for the measure ν, and $\mu(\{\infty\}) = \nu(\{\infty\})$ follows from the assumption that $\int_{\bar{\mathbb{R}}} K_a(X_\varepsilon) \, \mathrm{d}\mu(a) = \int_{\bar{\mathbb{R}}} K_a(X_\varepsilon) \, \mathrm{d}\nu(a)$.

We can symmetrically apply the above argument to the Bernoulli random variable that takes value 1 with probability $1 - \varepsilon$. Thus $\mu(\{-\infty\}) = \nu(\{-\infty\})$ holds as well.

Next, for each $n \in \mathbb{N}_+$ and real number $b > 0$, let $X_{n,b} \in L^\infty_{\mathbb{N}}$ satisfy

$$\begin{aligned} \mathbb{P}\left[X_{n,b} = n\right] &= \mathrm{e}^{-bn} \\ \mathbb{P}\left[X_{n,b} = 0\right] &= 1 - \mathrm{e}^{-bn}. \end{aligned}$$

Then

$$K_a(X_{n,b}) = \frac{1}{a} \log\left[(1 - \mathrm{e}^{-bn}) + \mathrm{e}^{(a-b)n}\right],$$

and so

$$\begin{aligned} \lim_{n \to \infty} \frac{1}{n} K_a(X_{n,b}) &= \lim_{n \to \infty} \frac{1}{n} \frac{1}{a} \log\left[1 - \mathrm{e}^{-bn} + \mathrm{e}^{(a-b)n}\right] \\ &= \begin{cases} 0 & \text{if } a < b \\ \frac{a-b}{a} & \text{if } a \geq b. \end{cases} \end{aligned}$$

This result holds also for $a = 0, \pm\infty$.

Note that $\frac{1}{n} K_a(X_{n,b})$ is uniformly bounded in $[0, 1]$ for all values of n, b, a, since $K_a(X_{n,b})$ is bounded between $\min[X_{n,b}] = 0$ and $\max[X_{n,b}] = n$. Thus, by the Dominated Convergence Theorem,

$$\lim_{n \to \infty} \int_{\bar{\mathbb{R}}} \frac{1}{n} K_a(X_{n,b}) \, \mathrm{d}\mu(a) = \int_{[b,\infty]} \frac{a-b}{a} \, \mathrm{d}\mu(a), \tag{9}$$

and similarly for ν. It follows that for all $b > 0$,

$$\int_{[b,\infty]} \frac{a-b}{a} \, \mathrm{d}\mu(a) = \int_{[b,\infty]} \frac{a-b}{a} \, \mathrm{d}\nu(a).$$

As $\mu(\{\infty\}) = \nu(\{\infty\})$, we in fact have

$$\int_{[b,\infty)} \frac{a-b}{a} \, \mathrm{d}\mu(a) = \int_{[b,\infty)} \frac{a-b}{a} \, \mathrm{d}\nu(a).$$

This common integral is denoted by $f(b)$.

We now define a measure $\hat{\mu}$ on $(0,\infty)$ by the condition $\frac{\mathrm{d}\hat{\mu}(a)}{\mathrm{d}\mu(a)} = \frac{1}{a}$; note that $\hat{\mu}$ is a positive measure, but need not be a probability measure. Then

$$f(b) = \int_{[b,\infty)} \frac{a-b}{a} \,\mathrm{d}\mu(a) = \int_{[b,\infty)} (a-b)\,\mathrm{d}\hat{\mu}(a) = \int_b^{\infty} \hat{\mu}([x,\infty))\,\mathrm{d}x,$$

where the last step uses Tonelli's Theorem. Hence $\hat{\mu}([b,\infty])$ is the negative of the left derivative of $f(b)$ (this uses the fact that $\hat{\mu}([b,\infty])$ is left continuous in b). In the same way, if we define $\hat{\nu}$ by $\frac{\mathrm{d}\hat{\nu}(a)}{\mathrm{d}\nu(a)} = \frac{1}{a}$, then $\hat{\nu}([b,\infty])$ is also the negative of the left derivative of $f(b)$. Therefore $\hat{\mu}$ and $\hat{\nu}$ are the same measure on $(0,\infty)$, which implies that μ and ν coincide on $(0,\infty)$.

By a symmetric argument (with $n - X_{n,b}$ in place of $X_{n,b}$), we deduce that μ and ν also coincide on $(-\infty,0)$. Finally, since they are both probability measures, μ and ν must have the same mass at 0, if any. So $\mu = \nu$. □

C Proof of Theorem 2

C.1 Proof for L^∞_+ and $L^\infty_\mathbb{N}$

It suffices to show that a monotone additive statistic defined on L^∞_+ or $L^\infty_\mathbb{N}$ can be extended to a monotone additive statistic defined on L^∞. First suppose Φ is defined on L^∞_+, the collection of non-negative random variables. Then for any bounded random variable X, we can define

$$\Psi(X) = \min[X] + \Phi(X - \min[X]),$$

where we note that $X - \min[X]$ is a non-negative random variable.

Clearly Ψ is a statistic that depends only on the distribution of X (as Φ does), and $\Psi(c) = c + \Phi(0) = c$ for constants c. When X is non-negative, the additivity of Φ gives $\Phi(X) = \Phi(\min[X]) + \Phi(X - \min[X]) = \min[X] + \Phi(X - \min[X])$, so Ψ is an extension of Φ. Moreover, Ψ is additive because $\min[X+Y] = \min[X] + \min[Y]$, and $\Phi(X + Y - \min[X+Y]) = \Phi(X - \min[X]) + \Phi(Y - \min[Y])$ by the additivity of Φ. Finally, to show Ψ is monotone, suppose X and Y are bounded random variables satisfying $X \geq_1 Y$. Then we can choose a sufficiently large n such that $X + n$ and $Y + n$ are both non-negative, and $X + n \geq_1 Y + n$. Since Φ is monotone for non-negative random variables, $\Phi(X+n) \geq \Phi(Y+n)$. Thus $\Psi(X+n) \geq \Psi(Y+n)$ by the fact that Ψ extends Φ, and $\Psi(X) \geq \Psi(Y)$ by the additivity of Ψ. This proves that Ψ is a monotone additive statistic on L^∞ that extends Φ.

In what follows, we consider the other case where Φ is initially defined on $L^\infty_\mathbb{N}$, the collection of non-negative integer-valued random variables. Given what has been shown above, we just need to extend Φ to a monotone additive statistic on L^∞_+. In this proof and later, we denote by X^{*n} the random variable that is the sum of n i.i.d. copies of X.

We also denote by $\lfloor X \rfloor$ the random variables that equals X rounded down to the nearest (non-negative) integer. Note that

$$\lfloor X+1 \rfloor \geq_1 X \geq_1 \lfloor X \rfloor.$$

We thus also have

$$\lfloor X+Y \rfloor \geq_1 \lfloor X \rfloor + \lfloor Y \rfloor, \tag{10}$$

$$\lfloor X+1 \rfloor + \lfloor Y+1 \rfloor \geq_1 \lfloor X+Y+1 \rfloor. \tag{11}$$

Given a monotone additive statistic $\Phi\colon L^\infty_{\mathbb{N}} \to \mathbb{R}$, define $\Psi\colon L^\infty_+ \to \mathbb{R}$ by

$$\Psi(X) = \lim_{n\to\infty} \frac{1}{n}\Phi(\lfloor X^{*n}+1 \rfloor) = \lim_{n\to\infty} \frac{1}{n}\Phi(\lfloor X^{*n} \rfloor).$$

The first limit exists because $b_n = \Phi(\lfloor X^{*n}+1 \rfloor)$ is a non-negative sequence which is sub-additive by (11) and by monotonicity and additivity of Φ, and thus $\lim_{n\to\infty} b_n/n = \inf_n b_n/n$ is well-known to exist. That the two limits above coincide follows from the additivity of Φ.

Ψ is a statistic because $\Psi(c) = \lim_{n\to\infty} \frac{1}{n}\Phi(\lfloor nc \rfloor) = \lim_{n\to\infty} \frac{1}{n}\lfloor nc \rfloor = c$ for every constant $c \geq 0$. It is also immediate to see that for integer-valued random variables X,

$$\Psi(X) = \lim_{n\to\infty} \frac{1}{n}\Phi(\lfloor X^{*n} \rfloor) = \lim_{n\to\infty} \frac{1}{n}\Phi(X^{*n}) = \lim_{n\to\infty} \frac{1}{n}n\Phi(X) = \Phi(X).$$

So Ψ extends Φ.

Moreover, if $X \geq_1 Y$, then $\lfloor X^{*n} \rfloor \geq_1 \lfloor Y^{*n} \rfloor$ for each n. This implies $\Psi(X) \geq \Psi(Y)$ by the above definition, so Ψ is monotone. Finally, to check Ψ is additive, we suppose X and Y be independent random variables. Then using (11), we have that for each n,

$$\lfloor (X+Y)^{*n}+1 \rfloor \leq_1 \lfloor X^{*n}+1 \rfloor + \lfloor Y^{*n}+1 \rfloor.$$

Together with the monotonicity and additivity of Φ, this implies

$$\begin{aligned}
\Psi(X+Y) &= \lim_{n\to\infty} \frac{1}{n}\Phi(\lfloor (X+Y)^{*n}+1 \rfloor) \\
&\leq \lim_{n\to\infty} \frac{1}{n}\Phi(\lfloor X^{*n}+1 \rfloor) + \lim_{n\to\infty} \frac{1}{n}\Phi(\lfloor Y^{*n}+1 \rfloor) \\
&= \Psi(X) + \Psi(Y).
\end{aligned}$$

Symmetrically, we can use the other definition of $\Psi(X+Y)$ and (10) to show that $\Psi(X+Y) \geq \Psi(X) + \Psi(Y)$. Hence equality holds, and Ψ is a monotone additive statistic that extends Φ. This completes the proof.

C.2 Proof for L_M

We break the proof into several steps below:

C.2.1 Step 1: The catalytic order on L_M

We first establish a generalization of Theorem 3 to unbounded random variables. For two random variables X and Y with c.d.f. F and G respectively, we say that X *dominates* Y *in both tails* if there exists a positive number N with the property that

$$G(x) > F(x) \quad \text{for all } |x| \geq N.$$

That is, we require the stochastic dominance condition between X and Y to hold in the tails. In particular, X needs to be unbounded from above, and Y unbounded from below.

Lemma 6. *Suppose $X, Y \in L_M$ satisfy $K_a(X) > K_a(Y)$ for every $a \in \mathbb{R}$. Suppose further that X dominates Y in both tails. Then there exists an independent random variable $Z \in L_M$ such that $X + Z \geq_1 Y + Z$.*

Proof. We will show that Z can be taken to have a normal distribution, which does belong to L_M. Following the proof of Theorem 3, we let $\sigma(x) = G(x) - F(x)$, and seek to show that $[\sigma * h](y) \geq 0$ for every y when h is a Gaussian density with sufficiently large variance. By assumption, $\sigma(x)$ is strictly positive for $|x| \geq N$. Thus there exists $\delta > 0$ such that $\int_{N+1}^{N+2} \sigma(x)\,\mathrm{d}x > \delta$, as well as $\int_{-N-2}^{-N-1} \sigma(x)\,\mathrm{d}x > \delta$. We fix $A > 0$ that satisfies $\mathrm{e}^A \geq \frac{4N}{\delta}$.

Similar to (7), we have for $h(x) = \mathrm{e}^{-\frac{x^2}{2V}}$ that

$$\mathrm{e}^{\frac{y^2}{2V}} \int \sigma(x) h(y-x)\,\mathrm{d}x = \int_{-\infty}^{\infty} \sigma(x) \cdot \mathrm{e}^{\frac{y}{V} \cdot x} \cdot \mathrm{e}^{-\frac{x^2}{2V}}\,\mathrm{d}x. \tag{12}$$

The variance V is to be determined below.

We first show that the right-hand side is positive if $V \geq (N+2)^2$ and $\frac{y}{V} \geq A$. Indeed, since $\sigma(x) > 0$ for $|x| \geq N$, this integral is bounded from below by

$$\begin{aligned}
& \int_{-N}^{N} \sigma(x) \cdot \mathrm{e}^{\frac{y}{V} \cdot x} \cdot \mathrm{e}^{-\frac{x^2}{2V}}\,\mathrm{d}x + \int_{N+1}^{N+2} \sigma(x) \cdot \mathrm{e}^{\frac{y}{V} \cdot x} \cdot \mathrm{e}^{-\frac{x^2}{2V}}\,\mathrm{d}x \\
\geq\ & -2N \cdot \mathrm{e}^{\frac{y}{V} \cdot N} + \delta \cdot \mathrm{e}^{\frac{y}{V} \cdot (N+1)} \cdot \mathrm{e}^{-\frac{(N+2)^2}{2V}} \\
=\ & \mathrm{e}^{\frac{y}{V} \cdot N} \cdot \left(-2N + \delta \cdot \mathrm{e}^{\frac{y}{V}} \cdot \mathrm{e}^{-\frac{(N+2)^2}{2V}}\right) \\
>\ & 0,
\end{aligned}$$

where the last inequality uses $\mathrm{e}^{\frac{y}{V}} \geq \mathrm{e}^A \geq \frac{4N}{\delta}$ and $\mathrm{e}^{-\frac{(N+2)^2}{2V}} \geq \mathrm{e}^{-\frac{1}{2}} > \frac{1}{2}$. By a symmetric argument, we can show that the right-hand side of (12) is also positive when $\frac{y}{V} \leq -A$.

It remains to consider the case where $\frac{y}{V} \in [-A, A]$. Here we rewrite the integral on the right-hand side of (12) as

$$\int_{-\infty}^{\infty} \sigma(x) \cdot \mathrm{e}^{\frac{y}{V} \cdot x} \cdot \mathrm{e}^{-\frac{x^2}{2V}} \, \mathrm{d}x = M_\sigma(\frac{y}{V}) - \int_{-\infty}^{\infty} \sigma(x) \cdot \mathrm{e}^{\frac{y}{V} \cdot x} \cdot (1 - \mathrm{e}^{-\frac{x^2}{2V}}) \, \mathrm{d}x,$$

where $M_\sigma(a) = \int_{-\infty}^{\infty} \sigma(x) \cdot \mathrm{e}^{ax} \, \mathrm{d}x = \frac{1}{a}\mathbb{E}\left[\mathrm{e}^{aX}\right] - \frac{1}{a}\mathbb{E}\left[\mathrm{e}^{aY}\right]$ is by assumption strictly positive for all a. By continuity, there exists some $\varepsilon > 0$ such that $M_{\sigma(a)} > \varepsilon$ for all $|a| \le A$. So it only remains to show that when V is sufficiently large,

$$\int_{-\infty}^{\infty} \sigma(x) \cdot \mathrm{e}^{ax} \cdot (1 - \mathrm{e}^{-\frac{x^2}{2V}}) \, \mathrm{d}x < \varepsilon \quad \text{for all } |a| \le A. \tag{13}$$

To estimate this integral, note that $M_\sigma(A) = \int_{-\infty}^{\infty} \sigma(x) \cdot \mathrm{e}^{Ax} \, \mathrm{d}x$ is finite. Since $\sigma(x) > 0$ for $|x|$ sufficiently large, we deduce from the Monotone Convergence Theorem that $\int_{-\infty}^{T} \sigma(x) \cdot \mathrm{e}^{Ax} \, \mathrm{d}x$ converges to $M_\sigma(A)$ as $T \to \infty$. In other words, $\int_{T}^{\infty} \sigma(x) \cdot \mathrm{e}^{Ax} \, \mathrm{d}x \to 0$. We can thus find a sufficiently large $T > N$ such that

$$\int_{T}^{\infty} \sigma(x) \cdot \mathrm{e}^{Ax} \, \mathrm{d}x < \frac{\varepsilon}{4}$$

and likewise

$$\int_{-\infty}^{-T} \sigma(x) \cdot \mathrm{e}^{-Ax} \, \mathrm{d}x < \frac{\varepsilon}{4}$$

As $1 - \mathrm{e}^{-\frac{x^2}{2V}} \ge 0$ and $\mathrm{e}^{ax} \le \mathrm{e}^{A|x|}$ when $|a| \le A$, we deduce that

$$\int_{|x| \ge T} \sigma(x) \cdot \mathrm{e}^{ax} \cdot (1 - \mathrm{e}^{-\frac{x^2}{2V}}) \, \mathrm{d}x < \frac{\varepsilon}{2} \quad \text{for all } |a| \le A.$$

Moreover, for this fixed T, we have $\mathrm{e}^{-\frac{T^2}{2V}} \to 1$ when V is large, and thus

$$\int_{|x| \le T} \sigma(x) \cdot \mathrm{e}^{ax} \cdot (1 - \mathrm{e}^{-\frac{x^2}{2V}}) \, \mathrm{d}x < 2T\mathrm{e}^{AT}(1 - \mathrm{e}^{-\frac{T^2}{2V}}) < \frac{\varepsilon}{2} \quad \text{for all } |a| \le A.$$

These estimates together imply that (13) holds for sufficiently large V. This completes the proof. □

C.2.2 Step 2: A perturbation argument

With Lemma 6, we know that if Φ is a monotone additive statistic defined on L_M, then $K_a(X) \ge K_a(Y)$ for all $a \in \mathbb{R}$ implies $\Phi(X) \ge \Phi(Y)$ *under the additional assumption that* X dominates Y in both tails (same proof as for Lemma 1). Below we deduce the same result without this extra assumption. To make the argument simpler, assume X and Y are unbounded both from above and from below; otherwise, we can add to them an independent Gaussian random variable without changing either the assumption or the conclusion. In doing so, we can further assume X and Y admit probability density functions.

The idea is that even if the right tail of X is not uniformly heavier than that of Y, we can add to X a positive random variable with sufficiently heavy tail, such that the resulting sum has heavier tail than Y. We first construct a heavy right-tailed random variable as follows:

Lemma 7. *For any $Y \in L_M$ that is unbounded from above and admits densities, there exists $Z \in L_M$ such that $Z \geq 0$ and*

$$\frac{\mathbb{P}[Z > x]}{\mathbb{P}[Y > x]} \to \infty \qquad \text{as } x \to \infty.$$

Proof. For this result, it is without loss assume $Y \geq 0$ because we can replace Y by $|Y|$ and only strengthen the conclusion. Let $g(x)$ be the probability density function of Y. We consider a random variable Z whose p.d.f. is given by $cxg(x)$ for all $x \geq 0$, where $c > 0$ is a normalizing constant to ensure $\int_{x \geq 0} cxg(x)\,\mathrm{d}x = 1$. Since the likelihood ratio between $Z = x$ and $Y = x$ is cx, it is easy to see that the ratio of tail probabilities also diverges. Thus it only remains to check $Z \in L_M$. This is because

$$\mathbb{E}\left[\mathrm{e}^{aZ}\right] = c \int_{x \geq 0} xg(x)\mathrm{e}^{ax}\,\mathrm{d}x,$$

which is simply c times the derivative of $\mathbb{E}\left[\mathrm{e}^{aY}\right]$ with respect to a. It is well-known that the moment generating function is smooth whenever it is finite. So this derivative is finite, and $Z \in L_M$. $\square$

In the same way, we can construct heavy left-tailed distributions:

Lemma 8. *For any $X \in L_M$ that is unbounded from below and admits densities, there exists $W \in L_M$, such that $W \leq 0$ and*

$$\frac{\mathbb{P}[W \leq x]}{\mathbb{P}[X \leq x]} \to \infty \qquad \text{as } x \to -\infty.$$

The following result constructs perturbed versions of any two random variables X and Y that satisfy "dominance in both tails." For any random variable $Z \in L_M$ and every $\varepsilon > 0$, let Z_ε be the random variable that equals Z with probability ε, and equals zero with probability $1 - \varepsilon$. Note that Z_ε also belongs to L_M.

Lemma 9. *Given any two random variables $X, Y \in L_M$ that are unbounded on both sides and admit densities. Let $Z \geq 0$ and $W \leq 0$ be constructed from the above two lemmata. Then for every $\varepsilon > 0$, $X + Z_\varepsilon$ dominates $Y + W_\varepsilon$ in both tails.*

Proof. For the right tail, we need $\mathbb{P}[X + Z_\varepsilon > x] > \mathbb{P}[Y + W_\varepsilon > x]$ for all $x \geq N$. Note that $W_\varepsilon \leq 0$, so $\mathbb{P}[Y + W_\varepsilon > x] \leq \mathbb{P}[Y > x]$. On other hand,

$$\mathbb{P}[X + Z_\varepsilon > x] \geq \mathbb{P}[X \geq 0] \cdot \mathbb{P}[Z_\varepsilon > x] = \mathbb{P}[X \geq 0] \cdot \varepsilon \cdot \mathbb{P}[Z > x].$$

Since by assumption X is unbounded from above, the term $\mathbb{P}[X \geq 0] \cdot \varepsilon$ is a strictly positive constant that does not depend on x. Thus for sufficiently large x, we indeed have

$$\mathbb{P}[X \geq 0] \cdot \varepsilon \cdot \mathbb{P}[Z > x] > \mathbb{P}[Y > x]$$

by the construction of Z. Thus we do have dominance in the right tail. The left tail works similarly. □

C.2.3 Step 3: Monotonicity with respect to K_a

The next result generalizes the key Lemma 1 to our current setting:

Lemma 10. *Let $\Phi\colon L_M \to \mathbb{R}$ be a monotone additive statistic. If $K_a(X) \geq K_a(Y)$ for all $a \in \mathbb{R}$ then $\Phi(X) \geq \Phi(Y)$.*

Proof. As discussed, we can without loss assume X, Y are unbounded on both sides, and admit densities. Let Z and W be constructed as above, then for each $\varepsilon > 0$, $X + Z_\varepsilon$ dominates $Y + W_\varepsilon$ in both tails, and $K_a(X + Z_\varepsilon) > K_a(X) \geq K_a(Y) > K_a(Y + W_\varepsilon)$ for every $a \in \mathbb{R}$, where the strict inequalities use $Z \geq 0$, $W \leq 0$ and neither is identically zero.

Thus the pair $X + Z_\varepsilon$ and $Y + W_\varepsilon$ satisfy the assumptions in Lemma 6, so we can find an independent random variable $V \in L_M$ (depending on ε), such that

$$X + Z_\varepsilon + V \geq_1 Y + W_\varepsilon + V.$$

Monotonocity and additivity of Φ then imply $\Phi(X) + \Phi(Z_\varepsilon) \geq \Phi(Y) + \Phi(W_\varepsilon)$, after cancelling out $\Phi(V)$. The desired result follows from the lemma below, which shows that our perturbations only slightly affect the statistic value. □

Lemma 11. *For any $Z \in L_M$ with $Z \geq 0$, it holds that $\Phi(Z_\varepsilon) \to 0$ as $\varepsilon \to 0$. Similarly $\Phi(W_\varepsilon) \to 0$ for any $W \in L_M$ with $W \leq 0$.*

Proof. We focus on the case for Z_ε. Suppose for contradiction that $\Phi(Z_\varepsilon)$ does not converge to zero. Note that as ε decreases, Z_ε decreases in first-order stochastic dominance. So $\Phi(Z_\varepsilon) \geq 0$ also decreases, and non-convergence must imply there exists some $\delta > 0$ such that $\Phi(Z_\varepsilon) > \delta$ for every $\varepsilon > 0$.

Let μ_ε be image measure of Z_ε. We now choose a sequence ε_n that decreases to zero very fast, and consider the measures

$$\nu_n = \mu_{\varepsilon_n}^{*n},$$

which is the n-th convolution power of μ_{ε_n}. Thus the sum of n i.i.d. copies of Z_{ε_n} is a random variable whose image measure is ν_n. We denote this sum by U_n.

For each n we choose ε_n sufficiently small to satisfy two properties: (i) $\varepsilon_n \le \frac{1}{n^2}$, and (ii) it holds that

$$\mathbb{E}\left[\mathrm{e}^{nU_n} - 1\right] \le 2^{-n}.$$

This latter inequality can be achieved because $\mathbb{E}\left[\mathrm{e}^{nU_n}\right] = \left(\mathbb{E}\left[\mathrm{e}^{nZ_{\varepsilon_n}}\right]\right)^n$, and as $\varepsilon_n \to 0$ we also have $\mathbb{E}\left[\mathrm{e}^{nZ_{\varepsilon_n}}\right] = 1 - \varepsilon_n + \varepsilon_n \mathbb{E}\left[\mathrm{e}^{nZ}\right] \to 1$ since $Z \in L_M$.

For these choices of ε_n and corresponding U_n, let $H_n(x)$ denote the c.d.f. of U_n, and define $H(x) = \inf_n H_n(x)$ for each $x \in \mathbb{R}$. Since $H_n(x) = 0$ for $x < 0$, the same is true for $H(x)$. Also note that each $H_n(x)$ is a non-decreasing and right-continuous function in x, and so is $H(x)$.

We claim that $\lim_{x\to\infty} H(x) = 1$. Indeed, recall that U_n is the n-fold sum of Z_{ε_n}, which has mass $1-\varepsilon_n$ at zero. So U_n has mass at least $(1-\varepsilon_n)^n \ge (1-\frac{1}{n^2})^n \ge 1-\frac{1}{n}$ at zero. In other words, $H_n(0) \ge 1 - \frac{1}{n}$. By considering the finitely many c.d.f.s $H_1(x), H_2(x), \dots, H_{n-1}(x)$, we can find N such that $H_i(x) \ge 1 - \frac{1}{n}$ for every $i < n$ and $x \ge N$. Together with $H_i(x) \ge H_i(0) \ge 1 - \frac{1}{i} \ge 1 - \frac{1}{n}$ for $i \ge n$, we conclude that $H_i(x) \ge 1 - \frac{1}{n}$ whenever $x \ge N$, and so $H(x) \ge 1 - \frac{1}{n}$. Since n is arbitrary, the claim follows. The fact that $H_n(x) \ge 1 - \frac{1}{n}$ also shows that in the definition $H(x) = \inf_n H_n(x)$, the "inf" is actually achieved as the minimum (since whenever the inf is less than 1, only finitely many $H_n(x)$ matters).

These properties of $H(x)$ imply that it is the c.d.f. of some non-negative random variable U. We next show $U \in L_M$, i.e., $\mathbb{E}\left[\mathrm{e}^{aU}\right] < \infty$ for every $a \in \mathbb{R}$. Since $U \ge 0$, we only need to consider $a \ge 0$. To do this, we take advantage of the following identity based on integration by parts:

$$\mathbb{E}\left[\mathrm{e}^{aU_n} - 1\right] = -\int_{x\ge 0} (\mathrm{e}^{ax} - 1)\,\mathrm{d}(1 - H_n(x)) = a\int_{x\ge 0} \mathrm{e}^{ax}(1 - H_n(x))\,\mathrm{d}x.$$

Now recall that we chose U_n so that $\mathbb{E}\left[\mathrm{e}^{nU_n} - 1\right] \le 2^{-n}$. So $\mathbb{E}\left[\mathrm{e}^{aU_n} - 1\right] \le 2^{-n}$ for every positve integer $n \ge a$. It follows that the sum $\sum_{n=1}^{\infty} \mathbb{E}\left[\mathrm{e}^{aU_n} - 1\right]$ is finite for every $a \ge 0$. Using the above identity, we deduce that

$$a\int_{x\ge 0} \mathrm{e}^{ax} \sum_{n=1}^{\infty} (1 - H_n(x))\,\mathrm{d}x < \infty,$$

where we have switched the order of summation and integration by the Monotone Convergence Theorem. Since $H(x) = \min_n H_n(x)$, it holds that $1 - H(x) \le \sum_{n=1}^{\infty}(1 - H_n(x))$ for every x. And thus

$$\mathbb{E}\left[\mathrm{e}^{aU} - 1\right] = a\int_{x\ge 0} \mathrm{e}^{ax}(1 - H(x))\,\mathrm{d}x < \infty$$

also holds. This proves $U \in L_M$.

We are finally in a position to deduce a contradiction. Since by construction the c.d.f. of U is no larger than the c.d.f. of each U_n, we have $U \geq_1 U_n$ and $\Phi(U) \geq \Phi(U_n)$ by monotonicity of Φ. But $\Phi(U_n) = n\Phi(Z_{\varepsilon_n}) > n\delta$ by additivity, so this leads to $\Phi(U)$ being infinite. This contradiction proves the desired result. $\square$

C.2.4 Step 4: Functional analysis

To complete the proof of the case of L_M in Theorem 2, we also need to modify the functional analysis step in our earlier proof of Theorem 1. One difficulty is that for an unbounded random variable X, $K_a(X)$ takes the value ∞ as $a \to \infty$. Thus we can no longer think of $K_X(a) = K_a(X)$ as a real-valued continuous function on $\bar{\mathbb{R}}$.

We remedy this as follows. Note first that if Φ is a monotone additive statistic defined on L_M, then it is also monotone and additive when restricted to the smaller domain of bounded random variables. Thus Theorem 1 gives a probability measure μ on $\mathbb{R} \cup \{\pm\infty\}$ such that

$$\Phi(X) = \int_{\bar{\mathbb{R}}} K_a(X)\,\mathrm{d}\mu(a)$$

for all $X \in L^\infty$. In what follows, μ is fixed. We just need to show that this representation also holds for $X \in L_M$.

As a first step, we show μ does not put any mass on $\pm\infty$. Indeed, if $\mu(\{\infty\}) = \varepsilon > 0$, then for any bounded random variable $X \geq 0$, the above integral gives $\Phi(X) \geq \varepsilon \cdot \max[X]$. Take any $Y \in L_M$ such that $Y \geq 0$ and Y is unbounded from above. Then monotonicity of Φ gives $\Phi(Y) \geq \Phi(\min\{Y, n\}) \geq \varepsilon \cdot n$ for each n. This contradicts $\Phi(Y)$ being finite. Similarly we can rule out any mass at $-\infty$.

The next lemma gives a way to extend the representation to certain unbounded random variables.

Lemma 12. *Suppose $Z \in L_M$ is bounded from below by 1 and unbounded from above, while $Y \in L_M$ is bounded from below and satisfies $\lim_{a\to\infty} \frac{K_a(Y)}{K_a(Z)} = 0$, then*

$$\Phi(Y) = \int_{(-\infty,\infty)} K_a(Y)\,\mathrm{d}\mu(a).$$

Proof. Given the assumptions, $K_a(Z) \geq 1$ for all $a \in \mathbb{R}$, with $\lim_{a\to\infty} K_a(Z) = \infty$. Let L_M^Z be the collection of random variables $X \in L_M$ such that X is bounded from below, and $\lim_{a\to\infty} \frac{K_a(X)}{K_a(Z)}$ exists and is finite. L_M^Z includes all bounded X (in which case $\lim_{a\to\infty} \frac{K_a(X)}{K_a(Z)} = 0$), as well as Y and Z itself. L_M^Z is also closed under adding independent random variables.

Now, for each $X \in L_M^Z$, we can define

$$K_{X|Z}(a) = \frac{K_a(X)}{K_a(Z)},$$

which reduces to our previous definition of $K_X(a)$ when Z is the constant 1 (in that case L^Z_M is precisely L^∞). This function $K_{X|Z}(a)$ extends by continuity to $a=-\infty$, where its value is $\frac{\min[X]}{\min[Z]}$, as well as to $a=\infty$ by construction. Thus $K_{X|Z}(\cdot)$ is a continuous function on $\bar{\mathbb{R}}$.

Since Φ induces an additive statistic when restricted to L^Z_M, and $K_{X|Z}+K_{Y|Z}=K_{X+Y|Z}$, we have an additive functional F defined on $\mathcal{L}=\{K_{X|Z}:\ X\in L^Z_M\}$, given by

$$F(K_{X|Z})=\frac{\Phi(X)}{\Phi(Z)}.$$

F is well-defined because $Z\geq 1$ implies $\Phi(Z)\geq 1$, and $F(1)=1$. By Lemma 10, F is also monotone in the sense that $K_{X|Z}(a)\geq K_{Y|Z}(a)$ for each $a\in\mathbb{R}$ implies $F(K_{X|Z})\geq F(K_{Y|Z})$.

Likewise we can show F is 1-Lipschitz. Note that $K_{X|Z}(a)\leq K_{Y|Z}(a)+\frac{m}{n}$ is equivalent to $K_a(X)\leq K_a(Y)+\frac{m}{n}K_a(Z)$ and equivalent to $K_a(X^{*n})\leq K_a(Y^{*n}+Z^{*m})$, where we write X^{*n} for the sum of n i.i.d. copies of X. If this holds for all a, then by Lemma 10 we also have $\Phi(X^{*n})\leq\Phi(Y^{*n}+Z^{*m})$, and thus $\Phi(X)\leq\Phi(Y)+\frac{m}{n}\Phi(Z)$ by additivity. Since $\frac{m}{n}$ is an arbitrary positive rational number, we conclude that for any real number $\varepsilon>0$, $K_{X|Z}(a)\leq K_{Y|Z}(a)+\varepsilon$ for all a implies $\Phi(X)\leq\Phi(Y)+\varepsilon\Phi(Z)$. Thus the functional F is 1-Lipschitz.

Given these properties, we can exactly follow the proof of Theorem 1 to extend the functional F to be a positive linear functional on the space of all continuous functions over $\bar{\mathbb{R}}$ (the majorization condition is again satisfied because the constant function n belongs to $\mathcal{L}$ for every positive integer n, since $K_{Z|Z}=1$). Therefore, by the Riesz-Markov-Kakutani Representation Theorem, we obtain a probability measure μ_Z on $\bar{\mathbb{R}}$ such that

$$\frac{\Phi(X)}{\Phi(Z)}=\int_{\bar{\mathbb{R}}}\frac{K_a(X)}{K_a(Z)}\,\mathrm{d}\mu_Z(a)$$

holds for all $X\in L^Z_M$.

In particular, for any X bounded from below such that $\lim_{a\to\infty}\frac{K_a(X)}{K_a(Z)}=0$, it holds that

$$\Phi(X)=\int_{[-\infty,\infty)}K_a(X)\cdot\frac{\Phi(Z)}{K_a(Z)}\,\mathrm{d}\mu_Z(a),$$

where we are able to exclude ∞ from the range of integration (this is important for the change of measure argument below).

If we define the measure $\hat{\mu}_Z$ by $\frac{\mathrm{d}\hat{\mu}_Z}{\mathrm{d}\mu_Z}(a)=\frac{\Phi(Z)}{K_a(Z)}\leq\Phi(Z)$, then since $K_a(X)$ is finite for $a<\infty$, we have

$$\Phi(X)=\int_{[-\infty,\infty)}K_a(X)\cdot\mathrm{d}\hat{\mu}_Z(a).$$

This in particular holds for all bounded X, so plugging in $X=1$ gives that $\hat{\mu}_Z$ is a probability measure. But now we have two probability measures μ and $\hat{\mu}_Z$ on $\bar{\mathbb{R}}$ that lead to the same integral representation for bounded random variables, so Lemma 5 implies that $\hat{\mu}_Z$ coincides with μ and is supported on the standard real line. Plugging in $X = Y$ in the above display then yields the desired result. $\square$

The preceding lemma is useful because, as it turns out, for any $X \in L_M$ bounded from below and unbounded from above, there exists $Z \in L_M$ bounded from below by 1 such that $\lim_{a\to\infty} \frac{K_a(X)}{K_a(Z)} = 0$ (which automatically implies Z is unbounded from above). This is the idea behind the following result:

Lemma 13. *For every $X \in L_M$ that is bounded from below,*

$$\Phi(X) = \int_{(-\infty,\infty)} K_a(X)\,\mathrm{d}\mu(a).$$

Proof. It suffices to consider X that is unbounded from above. Moreover, without loss we can assume $X \geq 0$ without changing the conclusion, since we can add any constant to X. Given the previous lemma, we just need to construct $Z \geq 1$ such that $\lim_{a\to\infty} \frac{K_a(X)}{K_a(Z)} = 0$. Note that $\mathbb{E}\left[\mathrm{e}^{aX}\right]$ strictly increases in a for $a \geq 0$. This means we can uniquely define a sequence $a_1 < a_2 < \cdots$ by the equation $\mathbb{E}\left[\mathrm{e}^{a_n X}\right] = \mathrm{e}^n$. This sequence diverges as $n \to \infty$. We then choose any increasing sequence b_n such that $b_n > n$ and $a_n b_n > 2n^2$.

Consider the random variable Z that is equal to b_n with probability $\mathrm{e}^{-\frac{a_n b_n}{2}}$ for each n, and equal to 1 with remaining probability. To see that $Z \in L_M$, we have

$$\mathbb{E}\left[\mathrm{e}^{aZ}\right] \leq \mathrm{e}^a + \sum_{n=1}^{\infty} \mathrm{e}^{-\frac{a_n b_n}{2}} \cdot \mathrm{e}^{a b_n} = \mathrm{e}^a + \sum_{n=1}^{\infty} \mathrm{e}^{(a - \frac{a_n}{2})\cdot b_n}.$$

For any fixed a, $\frac{a_n}{2}$ is eventually greater than $a+1$. This, together with the fact that $b_n > n$, implies the above sum converges.

Moreover, for any $a \in [a_n, a_{n+1})$, we have

$$\mathbb{E}\left[\mathrm{e}^{aZ}\right] \geq \mathbb{E}\left[\mathrm{e}^{a_n Z}\right] \geq \mathbb{P}[Z = b_n] \cdot \mathrm{e}^{a_n b_n} \geq \mathrm{e}^{\frac{a_n b_n}{2}} > \mathrm{e}^{n^2},$$

whereas $\mathbb{E}\left[\mathrm{e}^{aX}\right] \leq \mathbb{E}\left[\mathrm{e}^{a_{n+1}X}\right] \leq \mathrm{e}^{n+1}$. Thus

$$\frac{K_a(X)}{K_a(Z)} = \frac{\log \mathbb{E}\left[\mathrm{e}^{aX}\right]}{\log \mathbb{E}\left[\mathrm{e}^{aZ}\right]} \leq \frac{n+1}{n^2},$$

which converges to zero as a (and thus n) approaches infinity. $\square$

C.2.5 Step 5: Completing the proof

By a symmetric argument, the representation $\Phi(X) = \int_{(-\infty,\infty)} K_a(X)\,\mathrm{d}\mu(a)$ also holds for all X bounded from above. In the remainder of the proof, we will use an approximation argument to generalize this to all $X \in L_M$. We show a technical lemma that facilitates the argument:

Lemma 14. *The measure μ is supported on a compact interval of the standard real line.*

Proof. Suppose not, and without loss assume the support of μ is unbounded from above. We will construct a non-negative $Y \in L_M$ such that $\Phi(Y) = \infty$ according to the integral representation. Indeed, by assumption we can find a sequence $2 < a_1 < a_2 < \cdots$ such that $a_n \to \infty$ and $\mu([a_n, \infty)) \geq \frac{1}{n}$ for all large n. Let Y be the random variable that equals n with probability $\mathrm{e}^{-\frac{a_n \cdot n}{2}}$ for each n, and equals 0 with remaining probability. Then similar to the above, we can show $Y \in L_M$. Moreover, $\mathbb{E}\left[\mathrm{e}^{a_n Y}\right] \geq \mathrm{e}^{\frac{a_n \cdot n}{2}}$, implying that $K_{a_n}(Y) \geq \frac{n}{2}$. Since $K_a(Y)$ is increasing in a, we deduce that for each n,

$$\int_{[a_n, \infty)} K_a(Y) \, \mathrm{d}\mu(a) \geq K_{a_n}(Y) \cdot \mu([a_n, \infty)) \geq \frac{n}{2} \cdot \frac{1}{n} = \frac{1}{2}.$$

The fact that this holds for $a_n \to \infty$ contradicts the result that $\Phi(Y) = \int_{(-\infty,\infty)} K_a(Y) \, \mathrm{d}\mu(a)$ is finite. $\square$

Thus we can take N sufficiently large so that μ is supported on $[-N, N]$. To finish the proof, consider any $X \in L_M$ that may be unbounded on both sides. For each positive integer n, let $X_n = \min\{X, n\}$ denote the truncation of X at n. Since $X \geq_1 X_n$, we have

$$\Phi(X) \geq \Phi(X_n) = \int_{[-N,N]} K_a(X_n) \, \mathrm{d}\mu(a)$$

Observe that for each $a \in [-N, N]$, $K_a(X_n)$ converges to $K_a(X)$ as $n \to \infty$. Moreover, the fact that $K_a(X_n)$ increases both in n and in a implies that for all a and all n,

$$|K_a(X_n)| \leq \max\{|K_a(X_1)|, |K_a(X)|\} \leq \max\{|K_{-N}(X_1)|, |K_N(X_1)|, |K_{-N}(X)|, |K_N(X)|\}.$$

As $K_a(X_n)$ is uniformly bounded, we can apply the Dominated Convergence Theorem to deduce

$$\Phi(X) \geq \lim_{n \to \infty} \int_{[-N,N]} K_a(X_n) \, \mathrm{d}\mu(a) = \int_{[-N,N]} K_a(X) \, \mathrm{d}\mu(a).$$

On the other hand, if we truncate the left tail and consider $X^{-n} = \max\{X, -n\}$, then a symmetric argument shows

$$\Phi(X) \leq \lim_{n \to \infty} \int_{[-N,N]} K_a(X^{-n}) \, \mathrm{d}\mu(a) = \int_{[-N,N]} K_a(X) \, \mathrm{d}\mu(a).$$

Therefore for all $X \in L_M$ it holds that

$$\Phi(X) = \int_{[-N,N]} K_a(X) \, \mathrm{d}\mu(a).$$

This completes the entire proof of the case of L_M in Theorem 2.

D Additional Proofs

D.1 Proof of Theorem 4

In the first step, we fix any reward $x > 0$. Then by monotonicity in time and continuity, for each (x, T) there exists a (unique) deterministic time $\Phi_x(T)$ such that

$$(x, \Phi_x(T)) \sim (x, T).$$

Clearly, when T is a deterministic time, $\Phi_x(T)$ is simply T itself. Note also that if S first-order stochastically dominates T, then

$$(x, \Phi_x(T)) \sim (x, T) \succeq (x, S) \sim (x, \Phi_x(S)),$$

so that $\Phi_x(S) \geq \Phi_x(T)$. We next show that for any T and S that are independent, $\Phi_x(T+S) = \Phi_x(T)+\Phi_x(S)$. Indeed, by stationarity, $(x, \Phi_x(T)) \sim (x, T)$ implies $(x, \Phi_x(T)+S) \sim (x, T+S)$ and $(x, \Phi_x(S)) \sim (x, S)$ implies $(x, \Phi_x(T)+\Phi_x(S)) \sim (x, \Phi_x(T)+S)$. Taken together, we have

$$(x, \Phi_x(T) + \Phi_x(S)) \sim (x, T + S).$$

Since $\Phi_x(T) + \Phi_x(S)$ is a deterministic time, the definition of Φ_x gives $\Phi_x(T) + \Phi_x(S) = \Phi_x(T + S)$ as desired.

In the second step, note that our preference preference $\succeq$ induces a preference on $\mathbb{R}_+ \times \mathbb{R}_+$ consisting of deterministic dated rewards. By Theorem 2 in Fishburn and Rubinstein (1982), for any given $r > 0$ we can find a continuous and strictly increasing utility function u with $u(0) = 0$ such that for deterministic times $t, s \geq 0$

$$(x, t) \succeq (y, s) \quad \text{if and only if} \quad u(x) \cdot \mathrm{e}^{-rt} \geq u(y) \cdot \mathrm{e}^{-rs}.$$

By definition, $(x, T) \sim (x, \Phi_x(T))$ for any random time T. Thus we obtain that the decision maker's preference is represented by

$$(x, T) \succeq (y, S) \quad \text{if and only if} \quad u(x) \cdot \mathrm{e}^{-r\Phi_x(T)} \geq u(y) \cdot \mathrm{e}^{-r\Phi_y(S)}.$$

While $\Phi_0(T)$ was not defined before, it will not matter because $u(0) = 0$.

It remains to show that for all $x, y > 0$, Φ_x and Φ_y are the same statistic. For this we choose deterministic times t and s such that $(x, t) \sim (y, s)$, i.e.,

$$u(x) \cdot \mathrm{e}^{-rt} = u(y) \cdot \mathrm{e}^{-rs}.$$

For any random time T, stationarity implies $(x, t + T) \sim (y, s + T)$, so that

$$u(x) \cdot \mathrm{e}^{-r\Phi_x(t+T)} = u(y) \cdot \mathrm{e}^{-r\Phi_y(s+T)}.$$

Using the additivity of Φ_x and Φ_y, we can divide the above two equalities and obtain $\Phi_x(T) = \Phi_y(T)$ as desired. Since this holds for all T and all $x, y > 0$, we can write

$$\Phi_x(T) = \Phi(T)$$

for a single monotone additive statistic Φ. This completes the proof.

D.2 Time lotteries in discrete time

In this section we consider the domain $\mathbb{R}_+ \times L^\infty_{\mathbb{N}}$ of discrete time lotteries. The original axioms 3.1, 3.2 and 3.3 for continuous time directly carry over to discrete time, except that in some of their statements we now restrict to integer-valued random times. However, it turns out that we need a strengthening of the continuity axiom 3.4:

Axiom D.1 (Strong Continuity)**.** *Consider any sequence of discrete time lotteries* $\{(x_n, T_n)\}$ *such that* $x_n \to x$*, the distributions of* T_n *weakly converge to that of* T*, and* $\{\max[T_n]\}$ *is uniformly bounded. Then for any discrete time lottery* (y, S)*,* $(x_n, T_n) \succeq (y, S)$ *for every* n *implies* $(x, T) \succeq (y, S)$*, and* $(x_n, T_n) \preceq (y, S)$ *for every* n *implies* $(x, T) \preceq (y, S)$*.*

A feature of the above continuity axiom is that it rules out extreme risk aversion (or risk-seeking) over time. Thus, in the following analogue of Theorem 4, the monotone additive statistic Φ is generated by a measure μ supported on $\mathbb{R}$ rather than the extended real line $\bar{\mathbb{R}}$. We call such Φ *strongly monotone.*[14]

Proposition 6. *A preference* $\succeq$ *on* $\mathbb{R}_+ \times L^\infty_{\mathbb{N}}$ *satisfies Axioms 3.1, 3.2, 3.3 and D.1 if and only if there exists a strongly monotone additive statistic* Φ*, an* $r > 0$*, and a continuous and strictly increasing utility function* $u : \mathbb{R}_+ \to \mathbb{R}_+$ *with* $u(0) = 0$*, such that* $\succeq$ *is represented by*

$$f(x, T) = u(x) \cdot \mathrm{e}^{-r\Phi(T)}.$$

Proof. We first check that the representation satisfies the strong continuity Axiom D.1 (the other axioms are straightforward to check). Indeed, suppose

$$\Phi(T) = \int_{\mathbb{R}} K_a(T)\, \mathrm{d}\mu(a)$$

for some probability measure μ supported on $\mathbb{R}$. Then whenever $T_n \to T$ (in terms of their distributions) and $\max[T_n]$ is uniformly bounded, we can deduce from the Dominated Convergence Theorem that $\Phi(T_n) \to \Phi(T)$. This implies $u(x_n) \cdot \mathrm{e}^{-r\Phi(T_n)} \to u(x) \cdot \mathrm{e}^{-r\Phi(T)}$, and thus strong continuity holds.

Turning to the opposite direction, we assume the preference $\succeq$ satisfies the axioms. We first prove the following stronger version of stationarity:

$$(x, T) \succeq (y, S) \quad \text{if and only if} \quad (x, T + D) \succeq (y, S + D)$$

whenever D is independent from T and S. The "only if" direction is assumed, so we focus on the "if". It suffices to show that the strict preference $(x, T) \succ (y, S)$ also implies the strict preference $(x, T + D) \succ (y, S + D)$. Since $(0, T) \preceq (y, S)$, by strong continuity there

[14]We do not use the terminology "strictly monotone" because it suggests a weaker requirement that $\Phi(X) > \Phi(Y)$ whenever X is strictly larger than Y in first-order stochastic dominance. That would correspond to μ being not entirely supported on $\{\pm\infty\}$, whereas here we require μ to have no mass at $\pm\infty$.

exists $\tilde{x} \in [0, x)$ such that $(\tilde{x}, T) \sim (y, S)$. Thus by the assumed version of stationarity, $(\tilde{x}, T + D) \sim (y, S + D)$. Monotonicity in money then yields $(x, T + D) \succ (\tilde{x}, T + D) \sim (y, S + D)$. This gives the desired result.

Next, as in the proof of Theorem 4, we fix $x > 0$ and define a "certainty equivalent" $\Phi_x(T)$ for every T. However, since $\Phi_x(T)$ will not be an integer in general, we cannot define it using the indifference relation induced by $\succeq$. We instead proceed as follows. For each $T \in L^\infty_{\mathbb{N}}$, define

$$B_x(n, T) = \max\{m \in \mathbb{N} : \ (x, m) \succeq (x, T^{*n})\}$$

Note that for fixed T, $B_x(n, T)$ is a non-negative super-additive sequence in n. This is because if $(x, m_1) \succeq (x, T^{*n_1})$ and $(x, m_2) \succeq (x, T^{*n_2})$, then applying stationarity twice yields

$$(x, m_1 + m_2) \succeq (x, T^{*n_1} + m_2) \succeq (x, T^{*n_1} + T^{*n_2}) = (x, T^{*n_1+n_2}).$$

Note also that by monotonicity in time, $B_x(n, T) \leq \max[T^{*n}] = n \max[T]$. So we have a well-defined finite limit

$$\Phi_x(T) = \lim_{n\to\infty} \frac{1}{n} B_x(n, T).$$

It is easy to see that Φ_x is a monotone statistic. It is also super-additive because for each n, $(x, m) \succeq (x, T^{*n})$ and $(x, m') \succeq (x, S^{*n})$ imply $(x, m + m') \succeq (x, (T + S)^{*n})$ by two applications of stationarity. Moreover, using

$$B_x(n, T) = \min\{m \in \mathbb{N} : \ (x, m) \prec (x, T^{*n})\} - 1,$$

we can also show Φ_x is sub-additive. Thus Φ_x is a monotone additive statistic.

We next show that $(x, T) \succeq (x, S)$ if and only if $\Phi_x(S) \geq \Phi_x(T)$. Suppose $\Phi_x(S) > \Phi_x(T)$ holds strictly, then by definition we have $B_x(n, S) > B_x(n, T)$ for sufficiently large n. Thus, for this n, the integer $m = B_x(n, S)$ satisfies

$$(x, T^{*n}) \succ (x, m) \succeq (x, S^{*n}).$$

This implies $(x, T) \succ (x, S)$, because by repeated application of stationarity $(x, S) \succeq (x, T)$ would imply $(x, S^{*n}) \succeq (x, T^{*n})$.

It remains to show that $\Phi_x(S) = \Phi_x(T)$ implies $(x, T) \sim (x, S)$. By symmetry it suffices to show $(x, T) \succeq (x, S)$. Let S_ε be equal to S with probability $1 - \varepsilon$, and equal to $\max[S] + 1$ with probability ε. Then $\Phi_x(S_\varepsilon) > \Phi_x(S) = \Phi_x(T)$, so that $(x, T) \succ (x, S_\varepsilon)$ for every $\varepsilon > 0$. By strong continuity, we thus obtain $(x, T) \succeq (x, S)$ as desired.

We now further show Φ_x is strongly monotone. Suppose for contradiction that the measure μ associated with Φ_x puts mass at least $\frac{1}{N}$ on $a = \infty$, for some large positive integer N. Consider the random variable T_ε which equals N with probability ε and equals 0 otherwise. Note that $\Phi_x(T_\varepsilon) \geq \frac{1}{N} \max[T_\varepsilon] = 1$. Thus for any $\varepsilon > 0$, $(x, T_\varepsilon) \preceq (x, 1)$ by what we showed above. But since $T_\varepsilon \to 0$ and are uniformly bounded, strong continuity implies $(x, 0) \preceq (x, 1)$, contradicting monotonicity in time. A similar contradiction obtains if μ puts mass at least $\frac{1}{N}$ on $a = -\infty$, by considering the time lotteries $N - T_\varepsilon$ versus the deterministic time $N - 1$.

Hence, for every $x > 0$ we have constructed a strongly monotone additive statistic Φ_x, such that $(x, T) \succeq (x, S)$ if and only if $\Phi_x(T) \leq \Phi_x(S)$. What remains to be done is to relate the preferences for different rewards x. This is however another new difficulty relative to the proof of Theorem 4. The issue is that we cannot directly reduce the time lottery (x, T) to the deterministic reward $(x, \Phi_x(T))$ by indifference, since the latter need not be in *discrete* time.

To address this problem, we introduce an auxiliary preference $\succeq^*$ defined on the set of deterministic dates rewards $\mathbb{R}_+ \times \mathbb{R}_+$ in *continuous* time. Specifically, consider any (x, t) and (y, s), where $x, y > 0$ and t and s need not be integers. By the fact that Φ_x, Φ_y satisfy strong continuity, we can find integer-valued bounded random times T, S such that $\Phi_x(T) = t$ and $\Phi_y(S) = s$. We then define $(x, t) \succeq^* (y, s)$ if and only if $(x, T) \succeq (y, S)$. Since we have shown that $(x, T) \sim (x, T')$ whenever $\Phi_x(T) = \Phi_x(T')$, this definition of $\succeq^*$ does not depend on the specific choice of T and S. In addition, it is easy to see that $\succeq^*$ is complete and transitive. We can further include the zero reward by defining $(x, t) \succ (0, s)$ for any $x > 0$ and $(0, t) \sim (0, s)$.

Below we show that the preference $\succeq^*$ satisfies the axioms in Fishburn and Rubinstein (1982). We introduce a key technical lemma that we prove at the end of this section:

Lemma 15. *Let Φ and Ψ be two strongly monotone additive statistics defined on $L^\infty_\mathbb{N}$. Then for any real number $d > 0$, there exist two random variables $D, D' \in L^\infty_\mathbb{N}$ such that*

$$\Phi(D) - \Phi(D') = d = \Psi(D) - \Psi(D').$$

We use this lemma to prove the stationarity property of $\succeq^*$, namely $(x, t) \succeq^* (y, s)$ if and only if $(x, t + d) \succeq^* (y, s + d)$. Let $T, T', S, S' \in L^\infty_\mathbb{N}$ satisfy $\Phi_x(T) = t$, $\Phi_x(T') = t + d$, $\Phi_y(S) = s$, $\Phi_y(S') = s + d$. Also let $D, D' \in L^\infty_\mathbb{N}$ be given by Lemma 15, such that

$$\Phi_x(D) - \Phi_x(D') = d = \Phi_y(D) - \Phi_y(D').$$

Suppose $(x, t) \succeq^* (y, s)$, then by definition $(x, T) \succeq (y, S)$. This implies, by stationarity of $\succeq$, that

$$(x, T + D) \succeq (y, S + D).$$

Now observe that

$$\Phi_x(T + D) = \Phi_x(T) + \Phi_x(D) = t + d + \Phi_x(D') = \Phi_x(T' + D').$$

Thus $(x, T + D) \sim (x, T' + D')$ and likewise $(y, S + D) \sim (y, S' + D')$. It follows that

$$(x, T' + D') \succeq (y, S' + D').$$

By stationarity of $\succeq$ again, we conclude that $(x, T') \succeq (y, S')$ and thus $(x, t+d) \succeq^* (y, s+d)$. Moreover, if we have the strict preference $(x,t) \succ^* (y,s)$ to begin with, then the above steps and the conclusion $(x, t+d) \succ^* (y, s+d)$ are also strict. This proves the stationarity of $\succeq^*$.[15]

We now use stationarity to show $\succeq^*$ is monotone in money. Suppose $x > y > 0$, then $(x, 0) \succ (y, 0)$ when viewed as discrete time lotteries, and by definition $(x, 0) \succ^* (y, 0)$ when viewed as dated rewards. Thus stationarity implies $(x,t) \succ^* (y,t)$ for every $t \geq 0$. As for monotonicity in time, suppose $\Phi_x(S) = s > t = \Phi_x(T)$. Then $(x,T) \succ (x,S)$ by the fact that Φ_x represents the preference $\succeq$ restricted to the reward x. So by definition $(x,t) \succ^* (x,s)$ also holds.

It remains to check $\succeq^*$ is continuous in the sense that if $(x_n, t_n) \to (x,t)$ and $(x_n, t_n) \succeq^* (y,s)$ for every n, then $(x,t) \succeq^* (y,s)$ (and that the same holds for the preferences reversed). To show this, note that $t_n < \lfloor t \rfloor + 1$ for every large n. By strong monotonicity (thus continuity) of Φ_x, we can find a binary integer random variable T_n supported on 0 and $\lfloor t \rfloor + 1$ such that $\Phi_x(T_n) = t_n$. Passing to a sub-sequence if necessary, we can assume T_n has a limit T. Since $(x_n, t_n) \succeq^* (y,s)$, we know by definition that $(x_n, T_n) \succeq (y, S)$ for any S with $\Phi_y(S) = s$. Thus by strong continuity of $\succeq$, we deduce $(x,T) \succeq (y,S)$. Since $\Phi_x(T) = \lim \Phi_x(T_n) = \lim t_n = t$, we have $(x,t) \succeq^* (y,s)$ as desired.

Hence we can apply Theorem 2 in Fishburn and Rubinstein (1982) to deduce that

$$(x,t) \succeq^* (y,s) \quad \text{if and only if} \quad u(x) \cdot \mathrm{e}^{-rt} \geq u(y) \cdot \mathrm{e}^{-rs},$$

for some continuous and strictly increasing function $u : \mathbb{R}_+ \to \mathbb{R}_+$ with $u(0) = 0$. Since by definition $(x,T) \succeq (y,S)$ if and only if $(x, \Phi_x(T)) \succeq^* (y, \Phi_y(S))$, we obtain

$$(x,T) \succeq (y,S) \quad \text{if and only if} \quad u(x) \cdot \mathrm{e}^{-r\Phi_x(T)} \geq u(y) \cdot \mathrm{e}^{-r\Phi_y(S)}.$$

Once we have this representation, for any $x, y > 0$ we can find $T, S \in L^\infty_{\mathbb{N}}$ such that $u(x) \cdot \mathrm{e}^{-r\Phi_x(T)} = u(y) \cdot \mathrm{e}^{-r\Phi_y(S)}$, so $(x,T) \sim (y,S)$. Then for any independent D, we also have $(x, T+D) \sim (y, S+D)$ so that $u(x) \cdot \mathrm{e}^{-r\Phi_x(T+D)} = u(y) \cdot \mathrm{e}^{-r\Phi_y(S+D)}$. Dividing the two equalities thus yields $\Phi_x(D) = \Phi_y(D)$ for every D. We can therefore write $\Phi_x(T) = \Phi(T)$ for a single strongly monotone additive statistic Φ, which completes the proof. $\square$

Proof of Lemma 15. Suppose for the sake of contradiction that the result is not true. We claim there cannot exist $X, Y, X', Y' \in L^\infty_{\mathbb{N}}$ such that $\Phi(Y) - \Phi(X) = d < \Psi(Y) - \Psi(X)$ and $\Phi(Y') - \Phi(X') = d > \Psi(Y') - \Psi(X')$. Indeed, given such random variables, we may add a large constant to X', Y' so that $X' >_1 X$ and $Y' >_1 Y$, without affecting the assumption. Then as λ varies in $[0,1]$, the statistic value $\Phi(X'_\lambda X)$ increases continuously

[15] This proof would be a little simpler if there exists D such that $\Phi_x(D) = \Phi_y(D) = d$, which would be a stronger statement than Lemma 15. But such integer-valued D might not exist when d is not an integer, and Φ_x is larger than Φ_y in the sense of Proposition 4.

in λ (where $X'_\lambda X \in L^\infty_{\mathbb{N}}$ is the $(\lambda, 1-\lambda)$-mixture between X' and X). Likewise $\Phi(X'_\lambda X)$ increases continuously in λ. So for any $\lambda \in [0,1]$, there exists a unique $h(\lambda) \in [0,1]$ such that $\Phi(Y'_{h(\lambda)}Y) - \Phi(X'_\lambda X) = d$. This function $h(\lambda)$ is strictly increasing and continuous, and satisfies $h(0) = 0, h(1) = 1$. Note that $\Psi(Y'_{h(\lambda)}Y) - \Psi(X'_\lambda X)$ is larger than d when $\lambda = 0$, but smaller than d when $\lambda = 1$. Thus by continuity, there exists λ such that $\Psi(Y'_{h(\lambda)}Y) - \Psi(X'_\lambda X) = d = \Phi(Y'_{h(\lambda)}Y) - \Phi(X'_\lambda X)$.

Hence, for the lemma to fail, the only possibility is that $\Psi(Y) - \Psi(X)$ is always larger (or always smaller) than d whenever $\Phi(Y) - \Phi(X) = d$. Below we assume $\Psi(Y) - \Psi(X) > d$, but the opposite case can be symmetrically handled. Choose any positive integer $k > d$, and let $Y^* = k_\lambda 0$ be the unique binary random variable supported on $\{0, k\}$ such that $\Phi(Y^*) = d$. Then $\Psi(Y^*) > d$ by assumption, and we can assume $\Psi(Y^*) = d + \eta$ for some $\eta > 0$. This Y^* and η will be fixed in the subsequent analysis.

Now take any positive integer $m > k$. We can define a continuum of random variables $X_\varepsilon \in L^\infty_{\mathbb{N}}$ for $\varepsilon \in [0,1]$. Specifically, for $\varepsilon \in [0, \lambda]$ we define X_ε to be equal to k with probability ε and equal to 0 with probability $1-\varepsilon$. And for $\varepsilon \in [\lambda, 1]$, we define X_ε to be equal to m with probability $\frac{\varepsilon-\lambda}{1-\lambda}$, equal to k with probability $\frac{\lambda(1-\varepsilon)}{1-\lambda}$ and equal to 0 with probability $1-\varepsilon$. The important thing here is that as ε increases, X_ε increases in first-order stochastic dominance in a continuous way. Thus $\Phi(X_\varepsilon)$ and $\Psi(X_\varepsilon)$ increase continuously with ε. In addition, note that $X_0 = 0$, $X_\lambda = Y^*$ and $X_1 = m$.

Let $n \leq \frac{m}{d}$ be any positive integer. Then we can define $\varepsilon_0, \varepsilon_1, \cdots, \varepsilon_n$ by the equations $\Phi(X_{\varepsilon_j}) = j \cdot d$ for every $0 \leq j \leq n$. It is easy to see

$$0 = \varepsilon_0 < \lambda = \varepsilon_1 < \cdots < \varepsilon_n \leq 1.$$

For $1 \leq j \leq n$, we have $\Phi(X_{\varepsilon_j}) - \Phi(X_{\varepsilon_{j-1}}) = d$. So by assumption $\Psi(X_{\varepsilon_j}) - \Psi(X_{\varepsilon_{j-1}}) > d$. Moreover when $j = 1$ we in fact have $\Psi(X_{\varepsilon_j}) - \Psi(X_{\varepsilon_{j-1}}) = \Psi(Y^*) - \Psi(0) = d + \eta$. Summing across j, we thus obtain

$$m = \Psi(X_1) - \Psi(X_0) \geq \Psi(X_{\varepsilon_n}) - \Psi(X_{\varepsilon_0}) = \sum_{j=1}^{n} \Big(\Psi(X_{\varepsilon_j}) - \Psi(X_{\varepsilon_{j-1}})\Big) \geq nd + \eta.$$

But we now have a contradiction because the inequality $m \geq nd + \eta$ cannot hold for all sufficiently large integers m and n that satisfy $m \geq nd$. To see this, observe that when d is a rational number, we can choose m, n so that $m = nd$. In that case the inequality $m \geq nd + \eta$ clearly fails. If instead d is an irrational number, then it is well known that the fractional part of nd can be arbitrarily close to one (as implied by the "equidistribution" property). Again we can find large integers m and n such that $nd + \eta > m \geq nd$. Thus a contradiction obtains either way, completing the proof. □

D.3 Proof of Proposition 1

Clearly, a preference with the representation $f(x,T) = u(x) \cdot \mathrm{e}^{-r\Phi(T)}$ of Theorem 4 satisfies betweenness if and only if

$$\Phi(T) = \Phi(S) \;\text{ implies }\; \Phi(T_\lambda S) = \Phi(S) \;\text{ for all }\; \lambda \in (0,1).$$

In this case, we say that Φ satisfies betweenness. Thus, to prove the current proposition it suffices to show that any Φ satisfying betweenness has one of the following forms:

1. $\Phi(T) = K_a(T)$ for $a \in \bar{\mathbb{R}}$.
2. $\Phi(T) = \beta \min[T] + (1-\beta)\max[T]$ for some $\beta \in (0,1)$
3. $\Phi(T) = \frac{-a_1}{a_2-a_1} K_{a_1}(T) + \frac{a_2}{a_2-a_1} K_{a_2}(T) = \frac{\log \mathbb{E}\left[\mathrm{e}^{a_2 T}\right] - \log \mathbb{E}\left[\mathrm{e}^{a_1 T}\right]}{a_2-a_1}$ for some $a_1 < 0 < a_2$.

We first show the "if" direction. Specifically, when $\Phi(T) = K_a(T)$ for some fixed $a \in \mathbb{R}_{\neq 0}$, then $\Phi(T) = \Phi(S)$ implies $\mathbb{E}\left[\mathrm{e}^{aT}\right] = \mathbb{E}\left[\mathrm{e}^{aS}\right]$. It follows that $\mathbb{E}\left[\mathrm{e}^{aT_\lambda S}\right] = \lambda \mathbb{E}\left[\mathrm{e}^{aT}\right] + (1-\lambda)\mathbb{E}\left[\mathrm{e}^{aS}\right] = \mathbb{E}\left[\mathrm{e}^{aS}\right]$, and so $\Phi(T_\lambda S) = \Phi(S)$. It is straightforward to check that the same is true when $a = 0$ or $\pm\infty$. So every $K_a(T)$ satisfies betweenness.

We next show $\Phi(T) = \beta \min[T] + (1-\beta)\max[T]$ also satisfies betweenness for any $\beta \in (0,1)$. Indeed, suppose

$$\beta \min[T] + (1-\beta)\max[T] = \beta \min[S] + (1-\beta)\max[S].$$

Then either $\min[T] \leq \min[S], \max[T] \geq \max[S]$ or the other way around. In the former case $T_\lambda S$ has the same minimum and maximum as T, whereas in the latter case it has the same minimum and maximum as S. Either way, $\Phi(T_\lambda S) = \Phi(T) = \Phi(S)$ holds.

We then consider $\Phi(T) = \left(\log \mathbb{E}\left[\mathrm{e}^{a_2 T}\right] - \log \mathbb{E}\left[\mathrm{e}^{a_1 T}\right]\right)/(a_2 - a_1)$ for some $a_1 < 0 < a_2$. If $\Phi(T) = \Phi(S)$, then $\log \mathbb{E}\left[\mathrm{e}^{a_2 T}\right] - \log \mathbb{E}\left[\mathrm{e}^{a_1 T}\right] = \log \mathbb{E}\left[\mathrm{e}^{a_2 S}\right] - \log \mathbb{E}\left[\mathrm{e}^{a_1 S}\right]$, which is equivalent to

$$\frac{\mathbb{E}\left[\mathrm{e}^{a_2 T}\right]}{\mathbb{E}\left[\mathrm{e}^{a_2 S}\right]} = \frac{\mathbb{E}\left[\mathrm{e}^{a_1 T}\right]}{\mathbb{E}\left[\mathrm{e}^{a_1 S}\right]}.$$

Since $\mathbb{E}\left[\mathrm{e}^{aT_\lambda S}\right] = \lambda \mathbb{E}\left[\mathrm{e}^{aT}\right] + (1-\lambda)\mathbb{E}\left[\mathrm{e}^{aS}\right]$ for every a, it is not difficult to see that the above ratio equality continues to hold with T replaced by $T_\lambda S$. Hence $\Phi(T_\lambda S) = \Phi(S)$, and betweenness is satisfied.

Turning to the "only if" direction. We will characterize any monotone additive statistic Φ that satisfies the weak form of betweenness, i.e., $\Phi(T) = c$ implies $\Phi(T_\lambda c) = c$ when c is a constant. The following lemma is key to the argument:

Lemma 16. *Suppose $\Phi(T) = \int_{\mathbb{R}} K_a(T)\,\mathrm{d}\mu(a)$ has the property that $\Phi(T) = c$ implies the inequality $\Phi(T_\lambda c) \leq c$. Then the measure μ restricted to $[0,\infty]$ is either the zero measure, or it is supported on a single point.*

Proof. It suffices to show that if μ puts any mass on $(0,\infty]$, then that mass is supported on a single point and $\mu(\{0\}) = 0$. For this let $N > 0$ denote the essential maximum of the support of μ; that is, $N = \min\{x: \ \mu((x,\infty]) = 0\}$. We allow $N = \infty$ when the support of μ is unbounded from above, or when μ has a non-zero mass at ∞. For any positive real number $b < N$, consider the same random variable $X_{n,b}$ as in the proof of Lemma 5, given by

$$\begin{aligned}\mathbb{P}\left[X_{n,b} = n\right] &= \mathrm{e}^{-bn}\\ \mathbb{P}\left[X_{n,b} = 0\right] &= 1 - \mathrm{e}^{-bn}.\end{aligned}$$

As shown in the proof of Lemma 5, $\frac{1}{n}K_a(X_{n,b})$ is uniformly bounded in $[0,1]$, and

$$\lim_{n\to\infty} \frac{1}{n}K_a(X_{n,b}) = \frac{(a-b)^+}{a}.$$

Thus if we let $c_n = \Phi(X_{n,b})$, then by the Dominated Convergence Theorem,

$$\lim_{n\to\infty}\frac{c_n}{n} = \lim_{n\to\infty}\frac{1}{n}\Phi(X_{n,b}) = \lim_{n\to\infty}\int_{\bar{\mathbb{R}}}\frac{1}{n}K_a(X_{n,b})\,\mathrm{d}\mu(a) = \int_{(b,\infty]}\frac{a-b}{a}\,\mathrm{d}\mu(a).$$

Denote $\gamma = \int_{(b,\infty]}\frac{a-b}{a}\,\mathrm{d}\mu(a)$. This number γ is strictly positive because $b < N$ implies $\mu((b,\infty]) > 0$. We can also assume $\gamma < 1$, since otherwise μ must be the point mass at ∞.

Now, as $\Phi(X_{n,b}) = c_n$ we know by assumption that $\Phi(Y_{n,b}) \le c_n$ for each n, where $Y_{n,b}$ is the mixture between $X_{n,b}$ and the constant c_n (in what follows λ is fixed as n varies):

$$\begin{aligned}\mathbb{P}\left[Y_{n,b} = n\right] &= \lambda\mathrm{e}^{-bn}\\ \mathbb{P}\left[Y_{n,b} = 0\right] &= \lambda(1 - \mathrm{e}^{-bn})\\ \mathbb{P}\left[Y_{n,b} = c_n\right] &= 1-\lambda.\end{aligned}$$

Using $\lim_{n\to\infty} c_n/n = \gamma$, we have

$$\begin{aligned}\lim_{n\to\infty}\frac{1}{n}K_a(Y_{n,b}) &= \lim_{n\to\infty}\frac{1}{n}\frac{1}{a}\log\left[\lambda\left(1-\mathrm{e}^{-bn}+\mathrm{e}^{(a-b)n}\right)+(1-\lambda)\mathrm{e}^{a\cdot c_n}\right]\\ &= \begin{cases} 0 & \text{if } a < 0\\ (1-\lambda)\gamma & \text{if } a = 0\\ \gamma & \text{if } 0 < a < \frac{b}{1-\gamma}\\ \frac{a-b}{a} & \text{if } a \ge \frac{b}{1-\gamma}.\end{cases}\end{aligned}$$

Note that the cutoff point $a = \frac{b}{1-\gamma}$ is where $a - b = a\gamma$. When a is smaller than this, the dominant term in the bracketed sum above is $(1-\lambda)\mathrm{e}^{a\cdot c_n}$. Whereas for larger a, the dominant term becomes $\lambda\mathrm{e}^{(a-b)\cdot n}$.

Crucially, $\lim_{n\to\infty} \frac{1}{n} K_a(Y_{n,b}) \geq \frac{(a-b)^+}{a}$ holds for every a, with strict inequality for $a \in [0, \frac{b}{1-\gamma})$. Thus again by the Dominated Convergence Theorem,

$$\lim_{n\to\infty} \frac{c_n}{n} \geq \lim_{n\to\infty} \frac{1}{n}\Phi(Y_{n,b}) = \lim_{n\to\infty} \int_{\bar{\mathbb{R}}} \frac{1}{n} K_a(Y_{n,b})\,\mathrm{d}\mu(a) \geq \int_{(b,\infty]} \frac{a-b}{a}\,\mathrm{d}\mu(a).$$

But we know that the far left is equal to the far right. So both inequalities hold equal, and in particular $\lim_{n\to\infty} \frac{1}{n} K_a(Y_{n,b}) = \frac{(a-b)^+}{a}$ holds μ-almost surely.

As discussed, $\lim_{n\to\infty} \frac{1}{n} K_a(Y_{n,b}) > \frac{(a-b)^+}{a}$ for any $a \in [0, \frac{b}{1-\gamma})$. So we can conclude that $\mu([0, \frac{b}{1-\gamma})) = 0$. This must hold for any $b \in (0, N)$ and corresponding γ. Letting b arbitrarily close to N thus yields $\mu([0, N)) = 0$ (since $\frac{b}{1-\gamma} > b$). It follows that when restricted to $[0, \infty]$ the measure μ is concentrated at the single point N, as we desire to show. □

From this lemma, we know that if Φ satisfies the weak form of betweenness, then its associated measure μ can only be supported on one point in all of $[0, \infty]$. By a symmetric argument, μ also has at most one point support in all of $[-\infty, 0]$. Thus either $\mu = \delta_a$ for some $a \in \bar{\mathbb{R}}$, or μ is supported on two points $\{a_1, a_2\}$ with $a_1 < 0 < a_2$. We study the latter case below.

So suppose $\Phi(T) = \beta K_{a_1}(T) + (1-\beta) K_{a_2}(T)$ for some $\beta \in (0,1)$. If $a_1 = -\infty$ while $a_2 < \infty$, then $\Phi(T) = \beta \min[T] + (1-\beta) K_{a_2}(T)$. Take any non-constant T and let c denote $\Phi(T)$. Note that since $K_{a_2}(T) > \min[T]$, $c = \beta \min[T] + (1-\beta) K_{a_2}(T)$ lies strictly between $\min[T]$ and $K_{a_2}(T)$. Consider the mixture $T_\lambda c$, then $\min[T_\lambda c] = \min[T]$, whereas

$$K_{a_2}(T_\lambda c) = \frac{1}{a_2} \log\left(\lambda \mathbb{E}\left[\mathrm{e}^{a_2 T}\right] + (1-\lambda)\mathrm{e}^{a_2 c}\right) < \frac{1}{a_2} \log \mathbb{E}\left[\mathrm{e}^{a_2 T}\right] = K_{a_2}(T),$$

where the inequality uses $c < K_{a_2}(T) = \frac{1}{a_2} \log \mathbb{E}\left[\mathrm{e}^{a_2 T}\right]$ and $a_2 > 0$. We thus deduce that

$$\Phi(T_\lambda c) = \beta \min[T_\lambda c] + (1-\beta) K_{a_2}(T_\lambda c) < \beta \min[T] + (1-\beta) K_{a_2}(T) = c,$$

contradicting the betweenness assumption. A symmetric argument rules out the possibility that $a_1 > -\infty$ while $a_2 = \infty$.

It remains to consider $a_1 \in (-\infty, 0)$ and $a_2 \in (0, \infty)$. Here we just need to show that $\beta = \frac{-a_1}{a_2 - a_1}$. Let us again take an arbitrary non-constant T, and let

$$c = \Phi(T) = \frac{\beta}{a_1} \log \mathbb{E}\left[\mathrm{e}^{a_1 T}\right] + \frac{1-\beta}{a_2} \log \mathbb{E}\left[\mathrm{e}^{a_2 T}\right].$$

For an arbitrary $\lambda \in [0,1]$, we must also have

$$c = \Phi(T_\lambda c) = \frac{\beta}{a_1} \log \mathbb{E}\left[\lambda \mathrm{e}^{a_1 T} + (1-\lambda)\mathrm{e}^{a_1 c}\right] + \frac{1-\beta}{a_2} \log \mathbb{E}\left[\lambda \mathrm{e}^{a_2 T} + (1-\lambda)\mathrm{e}^{a_2 c}\right]. \quad (14)$$

Since (14) holds for every λ, we can differentiate it with respect to λ to obtain

$$0=\frac{\beta(\mathbb{E}\left[\mathrm{e}^{a_1T}\right]-\mathrm{e}^{a_1c})}{a_1\mathbb{E}\left[\lambda\mathrm{e}^{a_1T}+(1-\lambda)\mathrm{e}^{a_1c}\right]}+\frac{(1-\beta)(\mathbb{E}\left[\mathrm{e}^{a_2T}\right]-\mathrm{e}^{a_2c})}{a_2\mathbb{E}\left[\lambda\mathrm{e}^{a_2T}+(1-\lambda)\mathrm{e}^{a_2c}\right]}.$$

Plugging in $\lambda=0$ and $\lambda=1$ gives, respectively,

$$\frac{\beta(\mathbb{E}\left[\mathrm{e}^{a_1T}\right]-\mathrm{e}^{a_1c})}{a_1\mathrm{e}^{a_1c}}=-\frac{(1-\beta)(\mathbb{E}\left[\mathrm{e}^{a_2T}\right]-\mathrm{e}^{a_2c})}{a_2\mathrm{e}^{a_2c}}. \tag{15}$$

$$\frac{\beta(\mathbb{E}\left[\mathrm{e}^{a_1T}\right]-\mathrm{e}^{a_1c})}{a_1\mathbb{E}\left[\mathrm{e}^{a_1T}\right]}=-\frac{(1-\beta)(\mathbb{E}\left[\mathrm{e}^{a_2T}\right]-\mathrm{e}^{a_2c})}{a_2\mathbb{E}\left[\mathrm{e}^{a_2T}\right]}. \tag{16}$$

Since $c=\beta K_{a_1}(T)+(1-\beta)K_{a_2}(T)$, the fact that $K_{a_2}(T)>K_{a_1}(T)$ implies c is strictly between $K_{a_1}(T)$ and $K_{a_2}(T)$. Thus, using $a_1<0<a_2$ we deduce $\mathrm{e}^{a_1c}<\mathbb{E}\left[\mathrm{e}^{a_1T}\right]$ and $\mathrm{e}^{a_2c}<\mathbb{E}\left[\mathrm{e}^{a_2T}\right]$.

We can therefore divide (15) by (16) to obtain

$$\frac{\mathbb{E}\left[\mathrm{e}^{a_1T}\right]}{\mathrm{e}^{a_1c}}=\frac{\mathbb{E}\left[\mathrm{e}^{a_2T}\right]}{\mathrm{e}^{a_2c}}.$$

Plugging this back to (15), we conclude $\frac{\beta}{a_1}=-\frac{1-\beta}{a_2}$, so $\beta=\frac{-a_1}{a_2-a_1}$ as we desire to show.

D.4 Proof of Proposition 2

The "if" direction is straightforward, so we focus on the "only if." Note that for the representation given by Theorem 4, the independence axiom requires $\Phi(T_\lambda R)=\Phi(S_\lambda R)$ whenever $\Phi(T)=\Phi(S)$. This is stronger than the betweenness axiom, so we know from Proposition 1 that $\succeq$ must be represented by $f(x,T)=u(x)\cdot\mathrm{e}^{-r\Phi(T)}$ where Φ takes one of the following three forms:

(i) $\Phi(T)=K_a(T)$ for some $a\in\bar{\mathbb{R}}$, or

(ii) $\Phi(T)=\beta\min[T]+(1-\beta)\max[T]$ for some $\beta\in(0,1)$, or

(iii) $\Phi(T)=\frac{-a_1}{a_2-a_1}K_{a_1}(T)+\frac{a_2}{a_2-a_1}K_{a_2}(T)=\frac{\log\mathbb{E}\left[\mathrm{e}^{a_2T}\right]-\log\mathbb{E}\left[\mathrm{e}^{a_1T}\right]}{a_2-a_1}$ for some $a_1<0<a_2$.

We just need to show that form (ii) and (iii) violate the independence axiom. Suppose Φ takes form (ii). Let $S=1-\beta$ be a constant, and let T be distributed uniformly on $\{0,1\}$. Then $\Phi(T)=\Phi(S)=1-\beta$, but for any $\lambda\in(0,1)$

$$\Phi(T_\lambda 1)=1-\beta<\beta(1-\beta)+1-\beta=\Phi(S_\lambda 1).$$

This contradicts the independence axiom.

Next suppose Φ takes form (iii). Denote $\beta = \frac{-a_1}{a_2-a_1} \in (0,1)$, so that

$$\Phi(T) = \beta K_{a_1}(T) + (1-\beta)K_{a_2}(T).$$

We choose S and T such that $\Phi(T) > \Phi(S)$ but $K_{a_1}(T) < K_{a_1}(S)$. For example, let $S = 1$, and let T be supported on $\{0, k\}$, with $\mathbb{P}[T = n] = 1/k$. Then

$$K_a(T) = \frac{1}{a} \log \mathbb{E}\left[1 - 1/k + \mathrm{e}^{ak}/k\right].$$

For k tending to infinity, $K_a(T)$ tends to zero if $a < 0$, and to infinity if $a > 0$. Hence, for k large enough, S and T will have the desired property.

Now, let $R = n$. Then

$$\begin{aligned} K_a(S_\lambda n) &= \frac{1}{a} \log \mathbb{E}\left[\lambda \mathbb{E}\left[\mathrm{e}^{aS}\right] + (1-\lambda)\mathrm{e}^{an}\right] \\ K_a(T_\lambda n) &= \frac{1}{a} \log \mathbb{E}\left[\lambda \mathbb{E}\left[\mathrm{e}^{aT}\right] + (1-\lambda)\mathrm{e}^{an}\right] \end{aligned}$$

and so

$$K_a(S_\lambda n) - K_a(T_\lambda n) = \frac{1}{a} \log \left(\frac{\lambda \mathbb{E}\left[\mathrm{e}^{aS}\right] + (1-\lambda)\mathrm{e}^{an}}{\lambda \mathbb{E}\left[\mathrm{e}^{aT}\right] + (1-\lambda)\mathrm{e}^{an}} \right).$$

It easily follows that for $a > 0$,

$$\lim_{n\to\infty} K_a(S_\lambda n) - K_a(T_\lambda n) = 0,$$

whereas for $a < 0$,

$$\lim_{n\to\infty} K_a(S_\lambda n) - K_a(T_\lambda n) = K_a(S) - K_a(T).$$

Thus, as n tends to infinity,

$$\begin{aligned} \lim_n \Phi(S_\lambda n) - \Phi(T_\lambda n) &= \lim_n \beta \left[K_{a_1}(S_\lambda n) - K_{a_1}(T_\lambda n)\right] + (1-\beta)\left[K_{a_2}(S_\lambda n) - K_{a_2}(T_\lambda n)\right] \\ &= \beta \left[K_{a_1}(S) - K_{a_1}(T)\right] > 0. \end{aligned}$$

Therefore, for n large enough, we have found S and T such that $\Phi(T) > \Phi(S)$ but $\Phi(T_\lambda n) < \Phi(S_\lambda n)$.

If we let $c = \Phi(T) - \Phi(S) > 0$ and define $S' = S + c$, then $\Phi(T) = \Phi(S')$ and $\Phi(T_\lambda n) < \Phi(S'_\lambda n)$, where the latter follows from monotonicity of Φ. This contradicts the independence axiom and completes the proof of Proposition 2.

D.5 Proof of Proposition 3

We show that the result follows from Proposition 4, which we prove in the next section. Indeed, to characterize risk-averse preferences, it is equivalent to characterize those measures μ that are "more risk-averse than" the measure ν that is a point mass at zero (since this ν corresponds to $\Phi_\nu(T) \equiv T$). When $\nu = \delta_0$, the condition (i) in Proposition 4 is trivially satisfied because $\int_{[b,\infty]} \frac{a-b}{a} \,\mathrm{d}\nu(a) = 0$ whereas $\int_{[b,\infty]} \frac{a-b}{a} \,\mathrm{d}\mu(a) \geq 0$. On the other hand, condition (ii) requires $\int_{[-\infty,b]} \frac{a-b}{a} \,\mathrm{d}\mu(a) \leq 0$ for every $b < 0$. Since the integrand $\frac{a-b}{a}$ is strictly positive for $a \in [-\infty, b)$, this implies $\mu([-\infty, b)) = 0$, which further implies $\mu([-\infty, 0)) = 0$ since $b < 0$ is arbitrary. Hence μ is supported on $[0, \infty]$ as we desire to show.

Symmetrically, a measure ν is risk-seeking if and only if the measure $\mu = \delta_0$ is more risk-averse than ν. In this case condition (ii) in Proposition 4 is trivial whereas condition (i) reduces to ν being supported on $[-\infty, 0]$. This completes the proof.

D.6 Proof of Proposition 4

We first show that conditions (i) and (ii) are necessary for $\int_{\bar{\mathbb{R}}} K_a(T) \,\mathrm{d}\mu(a) \geq \int_{\bar{\mathbb{R}}} K_a(T) \,\mathrm{d}\nu(a)$ to hold for every T. This part of the argument closely follows the proof of Lemma 5 above. Specifically, by considering the same random variables $X_{n,b}$ as defined there, we have the key equation (9). Since the limit on the left-hand side is larger for μ than for ν, we conclude that for every $b > 0$, $\int_{[b,\infty]} \frac{a-b}{a} \,\mathrm{d}\mu(a)$ on the right-hand side must be larger than the corresponding integral for ν. Thus condition (i) holds, and an analogous argument shows condition (ii) also holds.

To complete the proof, it remains to show that when conditions (i) and (ii) are satisfied,

$$\int_{\bar{\mathbb{R}}} K_a(T) \,\mathrm{d}\mu(a) \geq \int_{\bar{\mathbb{R}}} K_a(T) \,\mathrm{d}\nu(a)$$

holds for every T. Since μ and ν are both probability measures, we can subtract $\mathbb{E}[T]$ from both sides and arrive at the equivalent inequality

$$\int_{\bar{\mathbb{R}}_{\neq 0}} (K_a(T) - \mathbb{E}[T]) \,\mathrm{d}\mu(a) \geq \int_{\bar{\mathbb{R}}_{\neq 0}} (K_a(T) - \mathbb{E}[T]) \,\mathrm{d}\nu(a). \tag{17}$$

Note that we can exclude $a = 0$ from the range of integration because $K_a(T) = \mathbb{E}[T]$ there. Below we show that condition (i) implies

$$\int_{(0,\infty]} (K_a(T) - \mathbb{E}[T]) \,\mathrm{d}\mu(a) \geq \int_{(0,\infty]} (K_a(T) - \mathbb{E}[T]) \,\mathrm{d}\nu(a). \tag{18}$$

Similarly, condition (ii) gives the same inequality when the range of integration is instead $[-\infty, 0)$. Adding these two inequalities would yield the desired comparison in (17).

To prove (18), we let $K_T(a) = a \cdot K_a(T) = \log \mathbb{E}\left[\mathrm{e}^{aT}\right]$ be the cumulant generating function of T. It is well known that $K_T(a)$ is convex in a, with $K_T'(0) = \mathbb{E}[T]$ and

$\lim_{a\to\infty} K_T'(a) = \max[T]$. Then the integral on the left-hand side of (18) can be calculated as follows:

$$\int_{(0,\infty]} (K_a(T) - \mathbb{E}[T])\,\mathrm{d}\mu(a) = \int_{(0,\infty)} (K_a(T) - \mathbb{E}[T])\,\mathrm{d}\mu(a) + (\max[T] - \mathbb{E}[T]) \cdot \mu(\{\infty\})$$
$$= \int_{(0,\infty)} (K_T(a) - a\mathbb{E}[T])\,\mathrm{d}\frac{\mu(a)}{a} + (\max[T] - \mathbb{E}[T]) \cdot \mu(\{\infty\})$$

Note that since the function $g(a) = K_T(a) - a\mathbb{E}[T]$ satisfies $g(0) = g'(0) = 0$, it can be written as

$$g(a) = \int_0^a g'(t)\,\mathrm{d}t = \int_0^a \int_0^t g''(b)\,\mathrm{d}b\,\mathrm{d}t = \int_0^a g''(b) \cdot (a-b)\,\mathrm{d}b.$$

Plugging back to the previous identity, we obtain

$$\int_{(0,\infty]} (K_a(T) - \mathbb{E}[T])\,\mathrm{d}\mu(a) = \int_{(0,\infty)} \int_0^a K_T''(b) \cdot (a-b)\,\mathrm{d}b\,\mathrm{d}\frac{\mu(a)}{a} + (\max[T] - \mathbb{E}[T]) \cdot \mu(\{\infty\})$$
$$= \int_0^\infty K_T''(b) \int_{[b,\infty)} (a-b)\,\mathrm{d}\frac{\mu(a)}{a}\,\mathrm{d}b + (K_T'(\infty) - K_T'(0)) \cdot \mu(\{\infty\})$$
$$= \int_0^\infty K_T''(b) \int_{[b,\infty)} \frac{a-b}{a}\,\mathrm{d}\mu(a)\,\mathrm{d}b + \int_0^\infty K_T''(b) \cdot \mu(\{\infty\})\,\mathrm{d}b$$
$$= \int_0^\infty K_T''(b) \int_{[b,\infty]} \frac{a-b}{a}\,\mathrm{d}\mu(a)\,\mathrm{d}b,$$

where the last step uses $\frac{a-b}{a} = 1$ when $a = \infty > b$.

The above identity also holds when μ is replaced by ν. It is then immediate to see that (18) follows from condition (i) and the fact that $K_T''(b) \geq 0$ for all b. This completes the proof.

D.7 Proof of Theorem 5

When $\succeq$ is represented by a monotone additive statistic Φ, we can easily check that the axioms are satisfied. For the Rabin and Weizsäcker axiom, note that $X_1 \succ Y_1$ and $X_2 \succ Y_2$ imply $\Phi(X_1) > \Phi(Y_1)$ and $\Phi(X_2) > \Phi(Y_2)$. By additivity of Φ, we thus have $\Phi(X_1 + X_2) > \Phi(Y_1 + Y_2)$. It follows from monotonicity of Φ that $X_1 + X_2$ cannot be first-order stochastically dominated by $Y_1 + Y_2$. The archimedeanity and responsiveness axioms are straightforward.

Turning to the "only if" direction, we suppose $\succeq$ satisfies the axioms. We first show that for any gamble X and any $\varepsilon > 0$,

$$\max[X] + \varepsilon \succ X \succ \min[X] - \varepsilon.$$

To see why, suppose for contradiction that X is weakly preferred to $\max[X] + \varepsilon$ (the other case can be handled similarly). Then we obtain a contradiction to the Rabin and Weizsäcker axiom by observing that $X \succ \max[X] + \frac{\varepsilon}{2}$, $\frac{\varepsilon}{4} \succ 0$ but $X + \frac{\varepsilon}{4} <_1 \max[X] + \frac{\varepsilon}{2} + 0$.

Given these upper and lower bounds for X, we can define $\Phi(X) = \sup\{c \in \mathbb{R} : c \preceq X\}$, which is well-defined and finite. By definition of the supremum and responsiveness, for any $\varepsilon > 0$ it holds that $\Phi(X) - \varepsilon \prec X \prec \Phi(X) + \varepsilon$. Thus by archimideanity, $\Phi(X) \sim X$ is the (unique) certainty equivalent of X. Clearly, Φ is a statistic.

It remains to show that $\Phi(X)$ is monotone and additive. For this we first show $\Phi(X + c) = \Phi(X) + c$ for any constant c. Suppose not, and $\Phi(X + c) = \Phi(X) + c'$ for some $c' > c$ (the case of $c' < c$ is similar). Let $\varepsilon \in (0, \frac{c'-c}{2})$ be a small positive number. Then by responsiveness, $X_1 = X + c + \varepsilon$ is strictly preferred to $X + c$ and thus preferred to the constant $Y_1 = \Phi(X) + c'$. On the other hand, $X_2 = \Phi(X) + \varepsilon$ is strictly preferred to $\Phi(X)$ and thus preferred to $Y_2 = X$. But $X_1 + X_2 = X + \Phi(X) + c + 2\varepsilon$ is stochastically dominated by $Y_1 + Y_2 = X + \Phi(X) + c'$, contradicting the Rabin and Weizsäcker axiom.

Using this result, and the archimedeanity axiom, we next show the following continuity property: whenever $X \succ Y$, there exists $\varepsilon > 0$ such that $X \succ Y + \varepsilon$ also holds. Indeed, suppose for contradiction that $Y + \varepsilon \succeq X$ for every $\varepsilon > 0$. Then by responsiveness, we in fact have the strict preference $Y + \varepsilon \succ X$. Thus $Y + \varepsilon \succ X \succ Y \succ Y - \varepsilon$. Since $Y \pm \varepsilon \sim \Phi(Y) \pm \varepsilon$, we deduce $\Phi(Y) + \varepsilon \succ X \succ \Phi(Y) - \varepsilon$ for every $\varepsilon > 0$. This implies $X \sim \Phi(Y) \sim Y$ by archimedeanity, which is a contradiction.

We now show $X \sim Y$ implies $X + Z \sim Y + Z$ for any independent Z. Suppose for contradiction that $X + Z \succ Y + Z$. Then we can find $\varepsilon > 0$ such that $X + Z \succ Y + Z + 2\varepsilon$. By responsiveness, it also holds that $Y + \varepsilon \succ Y \sim X$. But the sum $X + Z + Y + \varepsilon$ is stochastically dominated by $Y + Z + 2\varepsilon + X$, contradicting the Rabin and Weizsäcker axiom.

Therefore, from $X \sim \Phi(X)$ and $Y \sim \Phi(Y)$ we can apply the preceding result twice to obtain $X + Y \sim \Phi(X) + Y \sim \Phi(X) + \Phi(Y)$, so that $\Phi(X + Y) = \Phi(X) + \Phi(Y)$. Finally, we show $\Phi(\cdot)$ is monotone with respect to first-order stochastic dominance. Consider any $Y \geq_1 X$, and suppose for contradiction that $X \succ Y$. Then there exists $\varepsilon > 0$ such that $X \succ Y + 2\varepsilon$. This leads to a contradiction since $X \succ Y + 2\varepsilon$, $\varepsilon \succ 0$, but $X + \varepsilon$ is stochastically dominated by $Y + 2\varepsilon + 0$.

This completes the proof that the certainty equivalent $\Phi(X)$ is a monotone additive statistic. Hence the theorem.

E Sub- and Super-additive Statistics

In cases where the additivity assumption may seem too strong, we can weaken it to sub- or super-additivity as we describe in this section. Say a statistic Φ is *sub-additive* if $\Phi(X + Y) \leq \Phi(X) + \Phi(Y)$ whenever X, Y are independent bounded random variables, and *super-additive* if the reverse inequality holds. Say Φ is *homogeneous* if the equality $\Phi(X_1 + X_2) = \Phi(X_1) + \Phi(X_2)$ holds when X_1 and X_2 are independent and furthermore *identically distributed*. These properties are all implied by additivity.

The following result characterizes homogeneous and sub-additive (or super-additive) statistics on L^∞:[16]

Theorem 6. *$\Phi\colon L^\infty \to \mathbb{R}$ is monotone, homogeneous and sub-additive if and only if there exists a nonempty closed convex set C of Borel probability measures on $\bar{\mathbb{R}}$, such that for every $X \in L^\infty$ it holds that*

$$\Phi(X) = \max_{\mu \in C} \int_{\bar{\mathbb{R}}} K_a(X)\,\mathrm{d}\mu(a).$$

Likewise, Φ is monotone, homogeneous and super-additive if and only if

$$\Phi(X) = \min_{\mu \in C} \int_{\bar{\mathbb{R}}} K_a(X)\,\mathrm{d}\mu(a).$$

We use a few examples to illustrate that homogeneity and sub-additivity (or super-additivity) are both important for such representations. An example of a monotone statistic that is super-additive but not homogeneous is

$$\Phi(X) = \log\left(\frac{1}{2}\left[\mathrm{e}^{\min[X]} + \mathrm{e}^{\max[X]}\right]\right).$$

The super-additivity condition $\Phi(X+Y) \geq \Phi(X) + \Phi(Y)$ is equivalent to

$$2\left[\mathrm{e}^{\min[X]+\min[Y]} + \mathrm{e}^{\max[X]+\max[Y]}\right] \geq \left[\mathrm{e}^{\min[X]} + \mathrm{e}^{\max[X]}\right] \cdot \left[\mathrm{e}^{\min[Y]} + \mathrm{e}^{\max[Y]}\right],$$

which reduces to $(\mathrm{e}^{\max[X]} - \mathrm{e}^{\min[X]}) \cdot (\mathrm{e}^{\max[Y]} - \mathrm{e}^{\min[Y]}) \geq 0$. The same argument shows that $\min[X]$ and $\max[X]$ can be substituted with any pair of monotone additive statistics $\Psi(X)$ and $\Psi'(X)$ satisfying $\Psi \leq \Psi'$ (see Proposition 4). The resulting Φ would also be monotone, super-additive but not homogeneous.

Note that if $\Phi(X)$ is monotone and super-additive, then $-\Phi(-X)$ is monotone and sub-additive. In this way we also have an example of a monotone statistic that is sub-additive but not homogeneous.

As for an example of a monotone statistic that is homogeneous but not sub-additive or super-additive, we consider

$$\Phi(X) = \begin{cases} \min[X] & \text{if } \min[X] + \max[X] \leq 0 \\ \max[X] & \text{otherwise.} \end{cases}$$

[16] The representation in the super-additive case is reminiscent of the "cautious expected utility" representation of Cerreia-Vioglio et al. (2015), which evaluates a gamble X according to its minimum certainty equivalent across a family of utility functions. The difference is that our agent potentially takes the minimum across averages of certainty equivalents. In fact, since CARA certainty equivalents are increasing in the level of risk seeking, taking the minimum across these certainty equivalents (and not their averages) in our setting would reduce to a single CARA certainty equivalent.

This statistic is homogeneous because $\min[X]$ and $\max[X]$ are homogeneous. To see it is monotone, we will show $\Phi(Y) \geq \Phi(X)$ whenever $Y \geq_1 X$. If $\Phi(X) = \min[X]$ then $\Phi(Y) \geq \min[Y] \geq \min[X]$ holds. Otherwise $\Phi(X) = \max[X]$ and $\min[X] + \max[X] > 0$. Since $\max[Y] \geq \max[X]$ and $\min[Y] \geq \min[X]$, we also have $\min[Y] + \max[Y] > 0$. Hence $\Phi(Y) = \max[Y] \geq \Phi(X)$ also holds.

In addition, this statistic Φ is neither sub-additive nor super-additive. To see this, note that if X, Y are non-constant random variables, and if $\min[X] + \max[X] \leq 0 < \min[Y] + \max[Y]$, then whether $\Phi(X+Y) = \min[X+Y]$ or $\max[X+Y]$ depends on the sign of $\min[X]+\max[X]+\min[Y]+\max[Y]$. In the former case $\Phi(X+Y) = \min[X+Y] \leq \min[X] + \max[Y] = \Phi(X) + \Phi(Y)$, whereas in the latter case $\Phi(X+Y) \geq \Phi(X) + \Phi(Y)$. Both situations can occur.

Interestingly, the results in Theorem 6 need to be modified when we consider the smaller domain L^∞_+ of non-negative bounded random variables. This is elaborated below:

Proposition 7. $\Phi \colon L^\infty_+ \to \mathbb{R}$ *is monotone, homogeneous and sub-additive if and only if there exists a nonempty closed convex set* C *of Borel sub-probability measures on* $\bar{\mathbb{R}}$ *satisfying* $\max_{\mu \in C} |\mu| = 1$*, such that for every* $X \in L^\infty_+$ *it holds that*

$$\Phi(X) = \max_{\mu \in C} \int_{\bar{\mathbb{R}}} K_a(X) \, \mathrm{d}\mu(a).$$

The key distinction from Theorem 6 is that the maximum here can now be taken over a set of *sub-probability measures*. This possibility is ruled out in the case of all bounded random variables, since in that case we require $\Phi(c) = c$ also for negative constants c.

One might suspect that the analogue of Proposition 7 holds for monotone, homogeneous and super-additive statistics on L^∞_+, with minimization over super-probability measures. This is not quite true, as the following example suggests: for every $X \in L^\infty_+$, $\Phi(X) = 0$ if $\min[X] = 0$ and $\Phi(X) = \max[X]$ if $\min[X] > 0$. This statistic is readily checked to be monotone, homogeneous and super-additive.

However, it cannot be written as the form $\inf_{\mu \in C} \int_{\bar{\mathbb{R}}} K_a(X) \, \mathrm{d}\mu(a)$, and the key issue is a failure of *upper-semicontinuity (henceforce usc)*. Specifically, for any finite measure μ, the integral $\int_{\bar{\mathbb{R}}} K_a(X) \, \mathrm{d}\mu(a)$ is usc with respect to X in the sense that

$$\int_{\bar{\mathbb{R}}} K_a(X) \, \mathrm{d}\mu(a) \geq \lim_{\varepsilon \to 0_+} \int_{\bar{\mathbb{R}}} K_a(X + \varepsilon) \, \mathrm{d}\mu(a).$$

In fact we have equality since the reverse inequality always holds. It is well known that the infimum of a family of usc functions is also usc. But for the statistic Φ defined above, if X is the Bernoulli random variable that equals 0 and 1 with equal probabilities. Then $\Phi(X) = 0$ whereas $\Phi(X + \varepsilon) = 1 + \varepsilon$. So $\Phi(X) < \lim_{\varepsilon \to 0_+} \Phi(X + \varepsilon)$ and this Φ is not usc.

In what follows, we define a statistic Φ to be usc if for every X in the domain,

$$\Phi(X) = \lim_{\varepsilon \to 0_+} \Phi(X + \varepsilon).$$

Note that given monotonicity, it would be equivalent to write $\Phi(X) \geq \lim_{\varepsilon \to 0_+} \Phi(X+\varepsilon)$. Note also that usc was automatically satisfied when the domain was L^∞, in which case super-additivity gives

$$\Phi(X) \geq \Phi(X+\varepsilon) + \Phi(-\varepsilon) = \Phi(X+\varepsilon) - \varepsilon.$$

It was also satisfied under sub-additivity, since in that case

$$\Phi(X) \geq \Phi(X+\varepsilon) - \Phi(\varepsilon) = \Phi(X+\varepsilon) - \varepsilon.$$

So the combination of super-additivity and the smaller domain L^∞_+ is where we need to additionally assume usc.

The next result shows usc is exactly what we need to restore the representation:

Proposition 8. $\Phi\colon L^\infty_+ \to \mathbb{R}$ *is monotone, homogeneous, super-additive and upper-semicontinuous if and only if there exists a nonempty closed convex set C of finite Borel super-probability measures on $\bar{\mathbb{R}}$ satisfying $\min_{\mu \in C} |\mu| = 1$, such that for every $X \in L^\infty_+$ it holds that*

$$\Phi(X) = \inf_{\mu \in C} \int_{\bar{\mathbb{R}}} K_a(X) \, \mathrm{d}\mu(a).$$

We point out that usc is not sufficient to ensure the inf above is achieved as min. The reason is that the set of super-probability measures may contain measures with arbitrarily large total mass, so (sequential) compactness can be lost. To get a sharper result we need a stronger continuity notion, which we discuss in §E.4.

In the following sections we present the proofs for the above results.

E.1 Proof of Theorem 6

When the domain is all bounded random variables, it is sufficient to focus on the case of sub-additivity. This is because if Φ is monotone, homogeneous and super-additive, then $\Psi(X) = -\Phi(-X)$ is monotone, homogeneous and sub-additive. So the result for super-additivity can be immediately deduced from the result for sub-additivity. We will also omit the proof for the "if" direction of the theorem, which is straightforward.

Below we suppose Φ is sub-additive. For each random variable X and positive integer n, denote by X^{*n} the random variable that is the sum of n i.i.d. copies of X. Repeatedly applying sub-additivity, we have $\Phi(X^{*n}) \leq n\Phi(X)$ for each n, and equality holds when n is a power of two by homogeneity. Thus, for each n, if we choose any m with $2^m > n$, then by sub-additivity again

$$2^m \Phi(X) = \Phi(X^{*2^m}) \leq \Phi(X^{*n}) + \Phi(X^{*(2^m - n)}) \leq n\Phi(X) + (2^m - n)\Phi(X).$$

Thus the above inequalities hold equal, and we conclude that

$$\Phi(X^{*n}) = n\Phi(X), \quad \forall n \in \mathbb{N}_+.$$

This stronger property explains why we call $\Phi(X^{*2}) = 2\Phi(X)$ homogeneity.

The following lemma generalizes the key Lemma 1:

Lemma 17. *Let Φ be a monotone, homogeneous and sub-additive statistic defined on L^∞ or L^∞_+. If $K_a(X) \geq K_a(Y)$ for all $a \in \bar{\mathbb{R}}$ then $\Phi(X) \geq \Phi(Y)$.*

Proof. It suffices to show $\Phi(X+2\varepsilon) \geq \Phi(Y)$ for any $\varepsilon > 0$, which would imply $\Phi(X)+2\varepsilon \geq \Phi(Y)$ by sub-additivity. Denoting $\tilde{X} = X + \varepsilon$, then $K_a(\tilde{X}) = K_a(X) + \varepsilon > K_a(Y)$ for every $a \in \bar{\mathbb{R}}$. Thus by Theorem 3, there exists a bounded random variable Z such that

$$\tilde{X} + Z \geq_1 Y + Z.$$

Since first-order stochastic dominance is preserved under adding an independent random variable, we have

$$\tilde{X}_1 + \tilde{X}_2 + Z \geq_1 \tilde{X}_1 + Y_2 + Z \geq_1 Y_1 + Y_2 + Z,$$

where $\tilde{X}_1, \tilde{X}_2$ are i.i.d. copies of $\tilde{X}$ and similarly for Y_1, Y_2.

Iterating this procedure, we obtain that for each positive integer n,

$$\tilde{X}^{*n} + Z \geq_1 Y^{*n} + Z.$$

Since $N \geq_1 Z \geq_1 -N$, we further have

$$\tilde{X}^{*n} + N \geq_1 Y^{*n} + (-N),$$

or equivalently

$$(X + \varepsilon)^{*n} + 2N \geq_1 Y^{*n}.$$

Now, if we choose n so large that $\varepsilon n \geq 2N$, then the above implies

$$(X + 2\varepsilon)^{*n} \geq_1 (X + \varepsilon)^{*n} + 2N \geq_1 Y^{*n}.$$

Thus $\Phi(X + 2\varepsilon) \geq \Phi(Y)$ follows from the monotonicity and homogeneity of Φ. $\square$

Given Lemma 17, we can follow the proof of Theorem 1 and view $\Phi(X)$ as a functional $F(K_X)$, which has the following five properties:

1. constants: $F(c) = c$ for every constant function c;

2. monotonicity: $K_X \geq K_Y$ implies $F(K_X) \geq F(K_Y)$;

3. homogeneity: $F(nK_X) = nF(K_X)$, $\forall n \in \mathbb{N}_+$;

4. sub-additivity: $F(K_X + K_Y) \leq F(K_X) + F(K_Y)$;

5. Lipschitz: $|F(K_X) - F(K_Y)| \leq \|K_X - K_Y\|$.

The proof of Lipschitz continuity is essentially the same as Lemma 3, except that we instead have

$$F(K_Y) - F(K_X) \leq F(K_{X+\varepsilon}) - F(K_X) \leq F(K_\varepsilon) = \Phi(\varepsilon) = \varepsilon.$$

The second inequality here uses sub-additivity.

This functional F is initially defined on $\mathcal{L} = \{K_X : X \in L^\infty\}$. We now extend it to all of $\mathcal{C}(\bar{\mathbb{R}})$:

Lemma 18. *Any functional F on $\mathcal{L}$ satisfying the above five properties can be extended to a functional on $\mathcal{C}(\bar{\mathbb{R}})$ maintaining these properties, with homogeneity strengthened to allow for scalar multiplication with any positive real number (instead of n).*

Proof. As in the proof of Lemma 4, we can extend F by homogeneity to the rational cone spanned by $\mathcal{L}$, and then extend by continuity to the entire cone. We thus have a functional H defined on $\mathrm{Cone}(\mathcal{L})$ that satisfies monotonicity, homogeneity (over $\mathbb{R}_+$), sub-additivity and Lipschitz continuity.

To further extend H to all continuous functions, we define for each $g \in \mathcal{C}(\bar{\mathbb{R}})$

$$I(g) = \inf_{f \geq g,\ f \in \mathrm{Cone}(\mathcal{L})} H(f). \tag{19}$$

Note first that $I(g)$ is well-defined and finite. This is because each $\in \mathcal{C}(\bar{\mathbb{R}})$ is bounded, so the constant function $f = \max[g] \in \mathrm{Cone}(\mathcal{L})$ is point-wise greater than g. Moreover, *any* function $f \in \mathrm{Cone}(\mathcal{L})$ that is point-wise greater than g must be point-wise greater than the constant function $\min[g]$. So by monotonicity, $H(f) \geq \min[g]$ for any such f.

Secondly, when $g \in \mathrm{Cone}(\mathcal{L})$ we have $I(g) = H(g)$ by monotonicity of H. So I extends H. It is also easy to see $I(g)$ maintains monotonicity and homogeneity.

Thirdly, we check I is sub-additive. Fix any g_1, g_2 and choose any $\varepsilon > 0$. Then by definition of the infimum, there exists $f_1, f_2 \in \mathrm{Cone}(\mathcal{L})$ such that $f_i \geq g_i$ and $H(f_i) < I(g_i) + \varepsilon$ for $i = 1, 2$. Thus the function $f_1 + f_2 \in \mathrm{Cone}(\mathcal{L})$ and is bigger than $g_1 + g_2$. This implies

$$I(g_1 + g_2) \leq H(f_1 + f_2) \leq H(f_1) + H(f_2) < I(g_1) + I(g_2) + 2\varepsilon,$$

where the second inequality uses the sub-additivity of H. Since ε is arbitrary, I is indeed sub-additive.

Finally, we check I is Lipschitz. Suppose $g_1 \leq g_2 + \varepsilon$ for some $\varepsilon > 0$, then for any $f_2 \in \mathrm{Cone}(\mathcal{L})$ that is greater than g_2, we have $f_2 + \varepsilon \in \mathrm{Cone}(\mathcal{L})$ being greater than g_1. So by sub-additivity of H and $H(\varepsilon) = \varepsilon$,

$$I(g_1) \leq H(f_2 + \varepsilon) \leq H(f_2) + \varepsilon.$$

Letting $H(f_2)$ approach $I(g_2)$ thus yields the desired result $I(g_1) \leq I(g_2) + \varepsilon$.

Hence this functional I is the desired extension of F to all of $\mathcal{C}(\bar{\mathbb{R}})$. $\square$

Given this extension I satisfying $I(K_X) = \Phi(X)$, the "only if" direction of Theorem 6 will follow from the next result characterizing such functionals I:

Lemma 19. *Let $I : \mathcal{C}(\bar{\mathbb{R}}) \to \mathbb{R}$ be a functional that is monotone, homogeneous, sub-additive and Lipschitz, and maps any constant function to this constant. Then there exists a non-empty closed convex set C of Borel probability measures on $\bar{\mathbb{R}}$, such that for every $g \in \mathcal{C}(\bar{\mathbb{R}})$*

$$I(g) = \max_{\mu \in C} \int_{\bar{\mathbb{R}}} g(a) \, \mathrm{d}\mu(a).$$

Proof. Homogeneity and sub-additivity implies I is *convex*, in the sense that $I(\lambda g_1 + (1 - \lambda) g_2) \leq \lambda I(g_1) + (1 - \lambda) I(g_2)$ for all $g_1, g_2 \in \mathcal{C}(\bar{\mathbb{R}})$ and $\lambda \in (0, 1)$. Thus I is a convex and continuous functional on the normed function space $\mathcal{C}(\bar{\mathbb{R}})$. By Theorem 7.6 in Aliprantis and Border (2006), the functional I coincides with its *convex envelope*, meaning that

$$I(g) = \sup\{J(g) : \ J \leq I \ \text{ and } J \text{ is an affine and continuous functional}\}. \tag{20}$$

Using the Riesz-Markov-Kakutani Representation Theorem, any such functional J can be written as

$$J(g) = b + \int_{\bar{\mathbb{R}}} g(a) \, \mathrm{d}\mu(a)$$

for some $b \in \mathbb{R}$ and some possibly signed finite measure μ.

Now observe from (20) that $J(0) \leq I(0) = 0$, so $b \leq 0$. Moreover, since I is homogeneous, we deduce from $J(ng) \leq I(ng) = nI(g)$ that $\frac{b}{n} + \int_{\bar{\mathbb{R}}} g(a) \, \mathrm{d}\mu(a) \leq I(g)$ for every positive integer n, and thus $\hat{J}(g) = \int_{\bar{\mathbb{R}}} g(a) \, \mathrm{d}\mu(a)$ lies between $J(g)$ and $I(g)$. It follows that we can replace each affine J by the linear functional $\hat{J}$ without affecting (20). So we can rewrite

$$I(g) = \sup_{\mu \in C} \int_{\bar{\mathbb{R}}} g(a) \, \mathrm{d}\mu(a) \tag{21}$$

for some set C of possibly signed measures μ.

Choose $g \leq 0$. Then by monotonicity of I we have $I(g) \leq 0$. Thus (21) implies that $\int_{\bar{\mathbb{R}}} g(a) \, \mathrm{d}\mu(a) \leq I(g) \leq 0$. Since this holds for any continuous function $g \leq 0$, we conclude that each $\mu \in C$ is a non-negative measure. Moreover, plugging $g = 1$ into (21) yields $|\mu| \leq 1$, whereas plugging $g = -1$ implies $|\mu| \geq 1$. Thus C is a nonempty set of probability measures.

Finally, note that taking the closed convex hull of C does not affect the equality in (21). So we can assume C is closed and convex. In this case the supremum is achieved as maximum, because any sequence of probability measures on the compact metric space $\bar{\mathbb{R}}$ has a weakly convergent sub-sequence, by Prokhorov's Theorem. This proves Lemma 19 and thus Theorem 6. $\square$

E.2 Proof of Proposition 7

The proof is essentially the same as Theorem 6, so we only point out the differences. Lemma 17 holds without change, and we can still view $\Phi(X)$ as a functional $F(K_X)$. However, in the current setting F is only defined on $\mathcal{L}_+ = \{K_X : \ X \in L^\infty_+\}$, which only contains non-negative functions.

Using the same construction as in Lemma 18, we can extend F to a functional I on $\mathcal{C}(\bar{\mathbb{R}})$. But note that in applying Lemma 19, we need to weaken the assumption that I maps any constant function to this constant. In the current setting, this only holds for non-negative constants. Therefore, when following the proof of Lemma 19, we can no longer deduce $|\mu| \geq 1$ from (21) by plugging in $g = -1$. The consequence is that the conclusion of Lemma 19 is correspondingly weakened to

$$I(g) = \max_{\mu \in C} \int_{\bar{\mathbb{R}}} g(a) \,\mathrm{d}\mu(a)$$

for a non-empty closed convex set C of *sub-probability measures* satisfying $\max_{\mu \in C} |\mu| = 1$. Note that supremum is still achieved, since Prokhorov's Theorem also applies to sub-probability measures. This gives the result in Proposition 7.

E.3 Proof of Proposition 8

The necessity of upper-semicontinuity (usc) for the "inf-integral" representation has been discussed, so we again focus on the "only if" direction. We first derive the following analogue of Lemma 17:

Lemma 20. *Let Φ be an monotone, homogeneous and usc statistic defined on L^∞_+. If $K_a(X) \geq K_a(Y)$ for all $a \in \bar{\mathbb{R}}$ then $\Phi(X) \geq \Phi(Y)$.*

Proof. Recall that we showed before Lemma 17 that a homogeneous and sub-additive statistic satisfies the stronger form of homogeneity: $\Phi(X^{*n}) = n\Phi(X)$. An analogous argument applies to a homogeneous and super-additive statistic. Thus, following the proof of Lemma 17, we have

$$(X + 2\varepsilon)^{*n} \geq Y^{*n}$$

for every $\varepsilon > 0$ and n sufficiently large. Thus $\Phi((X + 2\varepsilon)^{*n}) \geq \Phi(Y^{*n})$ by monotonicity and $\Phi(X + 2\varepsilon) \geq \Phi(Y)$ by homogeneity.

Now since Φ is usc, $\lim_{\varepsilon \to 0_+} \Phi(X + 2\varepsilon) = \Phi(X)$. Hence $\Phi(X) \geq \Phi(Y)$ as desired. □

Given this lemma, we can now view $\Phi(X)$ as a functional $F(K_X)$ that has the following five properties (slightly different from the sub-additive case studied before):

1. constants: $F(c) = c$ for every non-negative constant function $c \geq 0$;

2. monotonicity: $K_X \geq K_Y$ implies $F(K_X) \geq F(K_Y)$;

3. homogeneity: $F(nK_X) = nF(K_X),\ \forall n \in \mathbb{N}_+$;

4. super-additivity: $F(K_X + K_Y) \geq F(K_X) + F(K_Y)$;

5. upper-semicontinuity: $\lim_{\varepsilon \to 0_+} F(K_X + \varepsilon) = F(K_X)$.

This functional F is defined on $\mathcal{L}_+ = \{K_X : \ X \in L_+^\infty\}$, but we will extend it to $\mathcal{C}_+(\bar{\mathbb{R}})$, the space of all *non-negative* continuous functions on $\bar{\mathbb{R}}$.

Lemma 21. *Any functional F on $\mathcal{L}_+$ satisfying the above five properties can be extended to a functional on $\mathcal{C}_+(\bar{\mathbb{R}})$ maintaining these properties, with homogeneity strengthened to be over $\mathbb{R}_+$.*

Proof. The proof of this lemma is somewhat different from the proof of Lemma 18 before, due to the fact that we have now have usc instead of Lipschitz continuity. Thus, in the current setting we first extend F to a functional G defined on the rational cone $\mathrm{Cone}_{\mathbb{Q}}(\mathcal{L}_+)$ that maintains the five properties and satisfies homogeneity over $\mathbb{Q}_+$ (usc of G follows from that of F and the definition $G(\frac{m}{n}K_X) = \frac{m}{n}F(K_X)$). But in the next step, instead of extending to the whole cone by Lipschitz continuity, we directly extend G to all of $\mathcal{C}_+(\bar{\mathbb{R}})$ by the following construction:

$$I(g) = \inf_{\varepsilon > 0} \left(\sup_{f \leq g + \varepsilon,\ f \in \mathrm{Cone}_{\mathbb{Q}}(\mathcal{L}_+)} G(f) \right). \tag{22}$$

For any $g \geq 0$ and $\varepsilon > 0$, the constant function $f = 0$ is in the rational cone and satisfies $f \leq g + \varepsilon$. So the inner supremum in (22) is non-negative. Moreover, any $f \leq g + \varepsilon$ is smaller than the constant function $\max[g] + \varepsilon$, so by monotonicity of G we know that the inner supremum is at most $\max[g] + \varepsilon$. This implies $I(g) \in [0, \max[g]]$ is well-defined.

We also note that if g is in the rational cone, then the inner supremum is achieved by the function $f = g + \varepsilon$ by monotonicity of G. Thus in this case $I(g) = \inf_{\varepsilon > 0} G(g + \varepsilon) = G(g)$, where the latter equality holds by usc of G. Thus the functional I extends G, and in particular I satisfies the first property above that $I(c) = c$ for every $c \geq 0$.

Secondly, we check I is monotone. This is clear because if $g_1 \geq g_2$, then for any $\varepsilon > 0$, the inner supremum in (22) is larger for g_1 than for g_2. So $I(g_1) \geq I(g_2)$.

The third property to check is homogeneity. We first show I is homogeneous over $\mathbb{Q}_+$, i.e., $I(\frac{m}{n}g) = \frac{m}{n}I(g)$ whenever m, n are positive integers. Indeed, by writing $\varepsilon = \frac{m}{n}\hat{\varepsilon}$ and

$\hat{f} = \frac{m}{n} f$ we have

$$\begin{aligned}
I(\frac{m}{n} g) &= \inf_{\varepsilon>0} \sup_{f \le \frac{m}{n} g+\varepsilon,\ f \in \mathrm{Cone}_{\mathbb{Q}}(\mathcal{L}_+)} G(f) \\
&= \inf_{\hat{\varepsilon}>0} \sup_{f \le \frac{m}{n}(g+\hat{\varepsilon}),\ f \in \mathrm{Cone}_{\mathbb{Q}}(\mathcal{L}_+)} G(f) \\
&= \inf_{\hat{\varepsilon}>0} \sup_{\hat{f} \le g+\hat{\varepsilon},\ \hat{f} \in \mathrm{Cone}_{\mathbb{Q}}(\mathcal{L}_+)} G(\frac{m}{n}\hat{f}) \\
&= \frac{m}{n} \inf_{\hat{\varepsilon}>0} \sup_{\hat{f} \le g+\hat{\varepsilon},\ \hat{f} \in \mathrm{Cone}_{\mathbb{Q}}(\mathcal{L}_+)} G(\hat{f}) \\
&= \frac{m}{n} I(g).
\end{aligned}$$

From the second line to the third line above, we used the observation that f is in the rational cone if and only if $\hat{f} = \frac{m}{n} f$ is in the rational cone. Now since I is homogeneous over $\mathbb{Q}_+$ and also monotone, an approximation argument shows that I is in fact homogeneous over $\mathbb{R}_+$ (note that we are dealing with non-negative functions here).

Next, we check I is super-additive. This follows from the observation that for each $\varepsilon > 0$,

$$\sup_{f_1 \le g_1+\varepsilon,\ f_1 \in \mathrm{Cone}_{\mathbb{Q}}(\mathcal{L}_+)} G(f_1) \ + \sup_{f_2 \le g_2+\varepsilon,\ f_2 \in \mathrm{Cone}_{\mathbb{Q}}(\mathcal{L}_+)} G(f_2) \le \sup_{f \le g_1+g_2+2\varepsilon,\ f \in \mathrm{Cone}_{\mathbb{Q}}(\mathcal{L}_+)} G(f).$$

The above inequality holds because for any f_1, f_2 showing up on the left-hand side, the function $f = f_1 + f_2$ is in the rational cone and satisfies $f \le g_1 + g_2 + 2\varepsilon$. So the right-hand side is at least $G(f_1 + f_2) \ge G(f_1) + G(f_2)$.

Finally, we check I is also usc. Choose any $g \ge 0$ and denote $b = I(g)$. Then we need to show that for any $\gamma > 0$, there exists $\delta > 0$ such that $I(g + \delta) \le b + \gamma$. To see this, note from the definition (22) that there exists some $\bar{\varepsilon} > 0$ such that

$$\sup_{f \le g+\bar{\varepsilon},\ f \in \mathrm{Cone}_{\mathbb{Q}}(\mathcal{L}_+)} G(f) \le b + \gamma.$$

Thus, for any $\delta < \bar{\varepsilon}$, we have

$$\begin{aligned}
I(g+\delta) &= \inf_{\varepsilon>0} \left(\sup_{f \le g+\delta+\varepsilon,\ f \in \mathrm{Cone}_{\mathbb{Q}}(\mathcal{L}_+)} G(f) \right) \\
&\le \left(\sup_{f \le g+\delta+\varepsilon,\ f \in \mathrm{Cone}_{\mathbb{Q}}(\mathcal{L}_+)} G(f) \right) \Bigg|_{\varepsilon=\bar{\varepsilon}-\delta} \\
&= \sup_{f \le g+\bar{\varepsilon},\ f \in \mathrm{Cone}_{\mathbb{Q}}(\mathcal{L}_+)} G(f) \\
&\le b + \gamma.
\end{aligned}$$

This completes the proof. $\square$

Proposition 8 now follows from the following analogue of Lemma 19:

Lemma 22. *Let $I : \mathcal{C}_+(\bar{\mathbb{R}}) \to \mathbb{R}$ be a functional that is monotone, homogeneous, super-additive and upper-semicontinuous, and maps any non-negative constant function to this constant. Then there exists a non-empty closed convex set C of finite Borel super-probability measures on $\bar{\mathbb{R}}$, such that for every $g \in \mathcal{C}_+(\bar{\mathbb{R}})$*

$$I(g) = \inf_{\mu \in C} \int_{\bar{\mathbb{R}}} g(a) \, \mathrm{d}\mu(a).$$

Proof. Note first that homogeneity and super-additivity imply I is a concave functional on non-negative continuous functions. Next, we extend I to all continuous functions by letting $I(g) = -\infty$ whenever $g(a) < 0$ at some $a \in \bar{\mathbb{R}}$. This is in fact consistent with (22), since for sufficiently small ε there exists no functions $f \in \mathrm{Cone}_{\mathbb{Q}}(\mathcal{L}_+)$ that satisfies $f \geq g + \varepsilon$.

Although the resulting functional (let us still call it I) sometimes take the value $-\infty$, it is a *proper extended concave function* according to Definition 7.1 in Aliprantis and Border (2006). Specifically, I is "proper" because it never assumes the value ∞ and does not always equal $-\infty$. It is "concave" because its *hypograph*

$$\text{hypo } I = \{(g, \alpha) \in \mathcal{C}(\bar{\mathbb{R}}) \times \mathbb{R} : \ \alpha \leq I(g)\}$$

is a convex set. This is because the requirement $\alpha \leq I(g)$ forces $g \geq 0$, so $I(g_1) \geq \alpha_1$ and $I(g_2) \geq \alpha_2$ imply $I(\lambda g_1 + (1-\lambda) g_2) \geq \lambda \alpha_1 + (1-\lambda)\alpha_2$ by the concavity of I for non-negative functions.

We next show this hypograph is a closed set, so that I is an *upper-semicontinuous proper concave functional* according to Section 7.2 in Aliprantis and Border (2006). Indeed, choose any sequence $\{g_n\} \subset \mathcal{C}(\bar{\mathbb{R}})$ and $\{\alpha_n\} \subset \mathbb{R}$ satisfying $I(g_n) \geq \alpha_n$ for each n, and suppose $g_n \to g$ in the sup norm and $\alpha_n \to \alpha$. Then we first have $g_n \geq 0$ and thus $g \geq 0$. Moreover, for each $\varepsilon > 0$ it holds that $g_n \leq g + \varepsilon$ for sufficiently large n. Thus by monotonicity of I,

$$I(g + \varepsilon) \geq I(g_n) \geq \alpha_n$$

for every large n. Taking n to infinity yields $I(g + \varepsilon) \geq \alpha$. But since ε is arbitrary, we have $I(g) = \lim_{\varepsilon \to 0_+} I(g + \varepsilon) \geq \alpha$ as well. So (g, α) also belongs to the hypograph, which is thus closed.

Now that we know I is an usc proper concave functional, we can apply the direct analogue of Theorem 7.6 in Aliprantis and Border (2006) to deduce that I coincides with its *concave envelope*:

$$I(g) = \inf\{J(g) : \ J \geq I \ \text{ and } J \text{ is an affine and continuous functional}\}.$$

As in the proof of Lemma 19, any such functional J can be written as

$$J(g) = b + \int_{\bar{\mathbb{R}}} g(a) \, \mathrm{d}\mu(a)$$

for some $b \in \mathbb{R}$ and some possibly signed finite measure μ. In fact, $b \geq 0$ by $I(0) \leq J(0)$. And since I is homogeneous, we can replace b by $\frac{b}{n}$. So in the end we can without loss assume $b = 0$.

Since for any continuous $g \geq 0$ it holds that $J(g) = \int_{\bar{\mathbb{R}}} g(a)\,\mathrm{d}\mu(a) \geq I(g) \geq 0$, any such μ is a non-negative measure. We also know from $J(1) \geq I(1) = 1$ that $|\mu| \geq 1$. This leads to the desired representation

$$I(g) = \inf_{\mu \in C} \int_{\bar{\mathbb{R}}} g(a)\,\mathrm{d}\mu(a)$$

for a set C of finite super-probability measures μ. As before, assuming C to be closed and convex is without loss, although in this case the infimum need not be achieved. $\square$

E.4 Strengthening Proposition 8

A feature in Proposition 8 is that the infimum is not necessarily achieved. To get a sharper result, we define Φ to be *Lipschitz usc* if there exists a constant $\ell > 0$ such that

$$\Phi(X+1) - \Phi(X) \leq \ell$$

holds for every X in the domain. Note that when Φ is homogeneous, this condition is equivalent to the stronger condition that

$$\Phi(X+\varepsilon) - \Phi(X) \leq \ell\varepsilon$$

for every X and very $\varepsilon > 0$.[17] We will use these conditions interchangeably.

Proposition 9. $\Phi\colon L^\infty_+ \to \mathbb{R}$ *is monotone, homogeneous, super-additive and Lipschitz upper-semicontinuous if and only if there exists a nonempty closed convex set C of uniformly bounded Borel super-probability measures on $\bar{\mathbb{R}}$ satisfying* $\min_{\mu \in C} |\mu| = 1$, *such that for every $X \in L^\infty_+$ it holds that*

$$\Phi(X) = \min_{\mu \in C} \int_{\bar{\mathbb{R}}} K_a(X)\,\mathrm{d}\mu(a).$$

Proof. For the "if" direction, we simply note that if all measures $\mu \in C$ have total mass no greater than ℓ, then

$$\int_{\bar{\mathbb{R}}} K_a(X+\varepsilon)\,\mathrm{d}\mu(a) \leq \ell\varepsilon + \int_{\bar{\mathbb{R}}} K_a(X)\,\mathrm{d}\mu(a)$$

for any $\mu \in C$ and any $\varepsilon > 0$. From this it follows that $\Phi(X+\varepsilon) - \Phi(X) \leq \ell\varepsilon$, and so the statistic Φ must be Lipschitz usc.

[17] To see this, note that $\Phi(X^{*n}+1) - \Phi(X^{*n}) \leq \ell$ implies $\Phi(X+\frac{1}{n}) - \Phi(X) \leq \frac{\ell}{n}$ by homogeneity. Thus $\Phi(X+\varepsilon) - \Phi(X) \leq \ell\varepsilon$ holds when ε is a positive rational number. By monotonicity of Φ, it also holds for any positive real number ε.

Turning to the "only if" direction, suppose Φ is Lipschitz usc. Then it is in particular usc, and we obtain from Proposition 8 that

$$\Phi(X) = \inf_{\mu \in C} \int_{\bar{\mathbb{R}}} K_a(X) \, \mathrm{d}\mu(a).$$

We now show that the Lipschitz property of Φ further implies it is without loss to assume the measures in C are uniformly bounded. To do this, let n be any positive integer and consider the following subset of C:

$$C_n = \{\mu \in C : \; |\mu| \leq n\}.$$

Define an alternative statistic

$$\hat{S}(X) = \inf_{\mu \in C_n} \int_{\bar{\mathbb{R}}} K_a(X) \, \mathrm{d}\mu(a) \geq \Phi(X).$$

If $\Phi(X) = \hat{S}(X)$ for every X then we are done. Otherwise there exists $X \in L^\infty_+$ and $\delta > 0$ such that $\hat{S}(X) > \Phi(X) + \delta$.

Now take any positive $\varepsilon < \frac{\delta}{n}$, we will show that $\Phi(X + \varepsilon) \geq \Phi(X) + n\varepsilon$. Indeed, for any measure $\mu \in C_n$, we have

$$\int_{\bar{\mathbb{R}}} K_a(X + \varepsilon) \, \mathrm{d}\mu(a) \geq \int_{\bar{\mathbb{R}}} K_a(X) \, \mathrm{d}\mu(a) \geq \hat{S}(X) > \Phi(X) + \delta > \Phi(X) + n\varepsilon.$$

On the other hand, if $\mu \in C \backslash C_n$, then it also holds that

$$\int_{\bar{\mathbb{R}}} K_a(X + \varepsilon) \, \mathrm{d}\mu(a) = \int_{\bar{\mathbb{R}}} (K_a(X) + \varepsilon) \, \mathrm{d}\mu(a) = \varepsilon|\mu| + \int_{\bar{\mathbb{R}}} K_a(X) \, \mathrm{d}\mu(a) > \Phi(X) + n\varepsilon.$$

Hence we do have $\Phi(X + \varepsilon) \geq \Phi(X) + n\varepsilon$ for ε sufficiently small.

But Φ is assumed to be Lipschitz usc, so the previous conclusion cannot hold for every n (and some X). It follows that for some n, we must have $\Phi(X) = \hat{S}(X)$. Therefore

$$\Phi(X) = \inf_{\mu \in C} \int_{\bar{\mathbb{R}}} K_a(X) \, \mathrm{d}\mu(a)$$

for a set C of super-probability measures that are uniformly bounded. Finally, we can take C to be closed and convex, and then the infimum is achieved by Prokhorov's Theorem. $\square$

The following is an example of a super-additive statistic that is usc but not Lipschitz usc. For each $s \geq 1$, let $\mu_s = s \cdot \delta_{-s^2}$ be the measure that puts mass s on $a = -s^2$. Consider

$$\Phi(X) = \inf_{s \geq 1} \int_{\bar{\mathbb{R}}} K_a(X) \, \mathrm{d}\mu_s(a) = \inf_{s \geq 1} -\frac{1}{s} \log \mathbb{E}\left[\mathrm{e}^{-s^2 X}\right].$$

If X equals 0 or 1 with equal probabilities, then $\mathbb{E}\left[\mathrm{e}^{-s^2 X}\right] = \frac{1}{2} + \frac{1}{2}\mathrm{e}^{-s^2} \in (\frac{1}{2}, \frac{3}{4})$. The above infimum thus evaluates to 0, which is the limit as $s \to \infty$ (not achieved at any finite s). But for any $\varepsilon > 0$, we have

$$\Phi(X + \varepsilon) = \inf_{s \geq 1} \int_{\bar{\mathbb{R}}} K_a(X + \varepsilon) \, \mathrm{d}\mu_s(a) = \inf_{s \geq 1} -\frac{1}{s} \log \mathbb{E}\left[\mathrm{e}^{-s^2 X}\right] + \varepsilon s.$$

Since $\mathbb{E}\left[\mathrm{e}^{-s^2 X}\right] < \frac{3}{4}$, it holds that for every $s \geq 1$,

$$-\frac{1}{s}\log \mathbb{E}\left[\mathrm{e}^{-s^2 X}\right] + \varepsilon s > \frac{\log \frac{4}{3}}{s} + \varepsilon s \geq 2\sqrt{\log \frac{4}{3} \cdot \varepsilon}.$$

Thus $\Phi(X) = 0$ while $\Phi(X + \varepsilon)$ is at least on the order of $\sqrt{\varepsilon}$, which violates Lipschitz upper-semicontinuity.