The Topology of Change:
Foundations of Probability with Black Swans

Graciela Chichilnisky

Abstract Classic probability theory treats rare events as 'outliers' that are often disregarded and underestimated. Yet in a moment of change rare events can become frequent and frequent events rare. We therefore postulate new axioms for probability theory that require a balanced treatment for rare and frequent events, based on what we call "the topology of change". The axioms extend the foundation of probability to integrate rare but potentially catastrophic events or *black swans:* natural hazards, market crashes, catastrophic climate change and major episodes of species extinction. The new results presented in this article include a characterization of a family of purely finitely additive measures that are - somewhat surprisingly - absolutely continuous with respect to the Lebesgue measure. This is a new development from an earlier characterization of all the probabilities measures implied by the new axioms as a combination of purely finitely additive and countably additive measures that was first established in Chichilnisky (2000, 2002, 2008, 2009), and the results are contrasted here to the work of Kolmogorov (1932), De Groot (1970), Arrow (1971)), Dubins and Savage (1963), Savage (1954), Von Neumann Morgernstern (1954), and Hernstein and Milnor (1972).

1 Introduction

Classic probability theory treats rare events as 'outliers' and often disregards them. This is an unavoidable shortcoming of classic theory that has been known for some time and conflicts with observations about the distribution of rare events in natural and human systems, such as earthquakes and financial markets. It is now known that the shortcoming originates from the axioms created by Kolmogorov (1932) to provide a foundation for probability theory, Chichilnisky (2000, 2002). It turns out that the same phenomenon that underestimates rare events leads classic probability

Graciela Chichilnisky
Columbia University, New York 10027 USA. e-mail: chichilnisky@columbia.edu

theory to underestimate the likelihood of change. In a situation of change, events that are rare become frequent and events that are frequent become rare. Therefore by ignoring rare events we tend to underestimate the possibility of change. In a slight abuse of language it could be said that classic probability theory leads us to 'ignore' change. The change we refer to includes rare events of great importance that should not be underestimated, for example *black swans* such as catastrophic climate change and major episodes of species extinction.

Sensitivity to change - or lack thereof - is a topological issue at its core. It measures how likelihoods change with changes in measurements or observations. If we are sensitive to change our responses change in harmony with the signals. To disregard change means that our response "needle" is either insensitive to, or at odds with, the signals. In mathematical terms this is all about continuity of the response and as such it is defined and measured by topology. In a recent discovery it was found that an important continuity axiom of classic probability theory is responsible for the insensitivity to rare events. De Groot (1972) calls this axiom SP4, Arrow (1982) called it "monotone continuity" (Arrow (1972), and similar continuity axioms appear in Hernstein and Milnor (1972), see Chichilnisky (2000, 2002)). The continuity that these axioms provide is coarse, and it was shown to be responsible for the insensitivity to rare events (Chichilnisky (2009, 2010)). In that sense the classic axioms lead to insensitivity about the likelihood of change. The fact is that a single continuity axiom explains why classic probability theory is insensitive to rare events and why it ignores change.

To overcome this limitation, new axioms for probability theory were created that balance the treatment of rare and frequent events, based on a more sensitive notion of continuity or a 'finer' topology – and new types of probability distributions have been identified as emerging from the new axioms (Chichilnisky 2000, 2002). In order to be sensitive to rare events, the new axioms have to use a different continuity criterion, a topology finer than that implicit in axiom SP4 or in the monotone continuity axiom, both of which involve averages. This new topology is about extremes not averages, and it is appropriately called "the topology of change" because it is more sensitive to the measurement of rare events that are often at stake in a situation of change. This new topology is the *sup norm* topology on L_∞ that, while new in this area, has been used earlier by Debreu (1953) to formalize Adam Smith's theorem on the Invisible Hand, and was used in Chichilnisky (2000, 2002) to axiomatize choice under uncertainty. The sup norm provides a finer notion of continuity than "monotone continuity" and SP4. This sensitivity tallies with the experimental evidence of how people react to rare events (Le Doux (1996), Chichilnisky (2009)). Using the topology of change, the new axioms of probability theory extend the classic foundations of probability, treating rare and frequent events in a more balanced fashion and providing a more balanced view on the likelihood of change.

This article provides new results in this framework, as follows. We introduce the Swan Axiom, a new axiom that is based on continuity in the topology of change. We show how the old and the new topologies differ, namely how continuity in the sense of monotone continuity and SP4 does not imply continuity in the topology of change and how this changes the probability distributions from countably addi-

tive to a combination of countably additive and purely finitely additive measures. We also identify a new family of purely finitely additive measures that is continuous with respect to the "topology of change". Somewhat surprisingly, we show that the change in topology - from probability distributions that satisfy Monotone Continuity to those who satisfy the topology of change - does not necessarily give rise to discontinuity with respect to the Lebesgue measure on R, such as 'delta functions' or measures having or "atoms". Indeed the new results presented in this article show the opposite: each of the measures in the family we provide of purely finitely additive measures satisfying the new axioms is in fact *absolutely continuous* with respect to the Lebesgue measure. Therefore the new notion of continuity that derives from the new axioms does not imply atoms nor assigns positive weights to sets of Lebesgue measure zero. These new results tally with the earlier characterization of probabilities measures satisfying the new axioms as combinations of purely finitely additive and countably additive measures, Chichilnisky (2000, 2002, 2008, 2009). We contrasts the new measures with the those defined by Kolmogorov (1932), De Groot (1970), Arrow (1971)), Dubins and Savage (1963), Savage (1954), Von Neumann Morgernstern (1954), and Hernstein and Milnor (1972). Finally we show that the new results rather than contradicting classic theory can be seen as an extension of it. The new theory of probability offered here is an extension of the old: the probability distributions implied by the new axioms coincide with classic countably additive distributions when the sample is populated only by frequent events, even though they are quite different in general. As already stated the new probability measures consist of a convex combination of countable and finitely additive measures with strictly positive elements of both which, in practical terms, assign more weight to black swans than do normal distributions, and predict more realistically the emergence of change and generally the incidence of 'outliers'.[1] When applied to decision theory under uncertainty, this gives rise to a new type of rationality that changes and updates Bayesian updating rules and also Von Neumann Morgernstern foundations of game theory (Chichilnisky, 1996, 2000, 2009, 2010, 2011), appearing to coincide with the observations in Le Doux (1996) of how the brain makes decisions using both the amigdala and the cortex.

The article is organized as follows. First we show how the standard notion of continuity or topology that is used in classic probability theory – "monotone continuity" as defined by Arrow 1971, and in Hernstein and Milnor 1972, De Groot 1970 – implies *countably additive measures* that are by nature insensitive to rare events and hence to change: these probability measures assign a negligible weight to rare events, no matter how important these may be, treating such events as outliers, Chichilnisky (2009 a,b). On the other hand the *purely finitely additive measures* defined by Dubins and Savage (1972) assign no weight to frequent events, which is equally troubling, as illustrated in the Appendix. Our new axiomatization for probability theory is shown to balance the two approaches and to extend both, requiring sensitivity to rare as well as to frequent events. This as we saw requires a

---

[1] The theory presented here explains also Jump-Diffusion processes (Chichilnisky 2012), the existence of 'heavy tails' in power law distributions, and the lumpiness of most of the physical systems that we observe and measure.

notion of continuity that is defined with respect to a finer topology that is sensitive to rare as well as to frequent events, the topology of change. The results presented here highlight the classic issue of topology and continuity that have always been at the core of the axioms of probability theory (Villegas, 1964, Arrow 1971).

2 The Mathematics of Uncertainty

Uncertainty is described by a distinctive and exhaustive set of **events** represented by sets $\{U_\alpha\}$ whose union describes a universe $\mathscr{U}$. An event is identified with its **characteristic function** $\phi_U : U \to R$ [2]. The relative likelihood or probability of an event[3] is a real number $W(U)$ that measures how likely it is to occur. The probability of the universe is 1 and that of the empty set is zero. Classic axioms for subjective probability (respectively likelihoods) were introduced by Kolmogorov (1932), see Savage (1954) and De Groot (1970). The relative likelihood or probability of two disjoint events is the sum of their probabilities: $W(U_1 \cup U_2) = W(U_1) + W(U_2)$ when $U_1 \cap U_2 = \emptyset$. This corresponds to the definition of a probability as a *measure* on a family ($\sigma-$algebra) of measurable sets of $\mathscr{U}$ [4].

A measure is a continuous linear function that assigns to each event U a real number. The space of events can therefore be identified with the space of characteristic functions, which are measurable and essentially bounded functions. When $\mathscr{U} = R$, the characteristic functions are in $L_\infty(R)$, the space of Lebesgue measurable and essentially bounded real valued functions on R, which we endow with the *"topology of change"*, defined as the sup norm $f : R \to R$, namely $\| f \| = ess \sup_R | f(x) |$. Recall that the functions in L_∞ are defined a.e. with respect to the Lebesgue measure on R, and are each absolutely continuous with respect to the Lebesgue measure on R. Since measures are continuous real valued function on L_∞, they are by definition in the dual space of L_∞, denoted L_∞^*, namely in the space of all continuous real valued functions on L_∞. A measure μ therefore satisfies the usual conditions (1) $\mu(A \cup B) = \mu(a) + \mu(B)$ if A and B are disjoint, and $\mu(\varnothing) = 0$. A *countably additive measure* is an element of L_∞^* that satisfies also (2) $\mu(\sum A_i) = \sum_i \mu(A_i)$ $i = 1, ...\infty$, when the sets A_i are disjoint. A *purely finitely additive measure* is an element of L_∞^* that satisfies condition (1) but not condition (2); therefore for a purely finitely additive measure there are cases where the measure of an infinite sequence of disjoint sets is not the sum of the sequence of their measures. The *space of all purely finitely additive measures* is denoted PA.

It is well known that $L_\infty^* = L_1 + PA$ where L_1 is the space of integrable functions on R with respect to the Lebesgue measure; this is a classic representation theorem (Yosida (1953)). Indeed, each countably additive measure can be represented by an integrable continuous function on $L_\infty(R)$ namely a function $g : R \to R$ in $L_1(R)$,

---

[2] $\phi_U(x) = 1$ when $x \in U$ and $\phi_U(x) = 0$ when $x \notin U$

[3] In this article we make no difference between probabilities and relative likelihoods.

[4] This is Savage's (1954) definition of probability

where the representation takes the form $\mu(A) = \int_A g(x)dx$. This representation does not apply to purely finitely additive measures.[5] A *vanishing sequence of events* $\{E_\alpha\}_{=1,2...}$ is defined as one satisfying.$\forall \alpha$, $E_{\alpha+1} \subset E_\alpha$ and $\cap_{\alpha=1}^\infty E_\alpha = \emptyset$ a.e. The following two continuity axioms were introduced in Villegas (1964), see also Arrow (1971), Hernstein and Milnor (1972) and DeGroot (1970), in each case for the purpose of ensuring countable additivity:

Axiom 1 (Monotone Continuity Axiom (MC)) *For every **vanishing** sequence of events $\{E_\alpha\}_{=1,2...}$ the probability $W(E_i) \to 0$ as $i \to \infty$.*

In words, this axiom requires that the probability of the sets along a vanishing sequence goes to zero. For example consider the decreasing sequence made of infinite intervals of the form (n, ∞) for $n = 1, 2$. . This is a vanishing sequence. Monotone continuity implies that the likelihood of this sequence of events goes to zero even though all its sets are unbounded and essentially identical. A similar example can be constructed with a decreasing sequence of bounded sets, $(-1/n, 1/n)$ for $n = 1, 2...$,which is also a vanishing sequence as it is a decreasing sequence and their intersection is a single point $\{0\}$: observe that the set consisting of a single point $\{0\}$ is almost everywhere (a.e.) equal to the empty set on R, and that the events in this section are always defined a.e. with respect to the Lebesgue measure of R.[6]

Axiom 2 (De Groot's Axiom SP_4)[7] *If $A_1 \supset A_2 \supset$ ...is a decreasing sequence of events and B is some fixed event that is less likely than A_i for all i, then the probability or likelihood of the intersection $\cap_i^\infty A_i$ is larger than the probability or likelihood of the event B.*

The following proposition establishes that the two axioms presented above, Monotone Continuity and SP_4, are equivalent and that both imply countable additivity:

Proposition 1. *A relative likelihood (or probability measure) satisfies the **Monotone Continuity** Axiom if and only if it satisfies Axiom SP_4 and each of the two axioms implies countable additivity of the corresponding relative likelihood.*

Proof. Assume that Axiom SP_4 is satisfied. When the intersection of a decreasing (nested) vanishing sequence of events $\{A_i\}$ is empty namely $\cap_i A_i = \emptyset$ and the set B is less likely to occur than every set A_i, then the subset B must be as likely as the empty set, namely its probability must be zero. In other words, if B is more likely than the

---

[5] Savage's probabilities can be either purely finitely additive or countably additive. In that sense they include all the probabilities in this article. However this article will exclude probabilities that are either purely finitely additive, or those that are countably additive, and therefore our characterization of a probability is strictly finer than that Savage's (1954), and different from the view of a measure as a countably additive set function in De Groot, 1970.

[6] An equivalent definition of Monotone Continuity is that for every two events E_1 and E_2 in $\{E_\alpha\}_{=1,2...}$ with $W(E_1) > W(E_2)$, there exists N such that altering arbitrarily the events E_1 and E_2 on a subset subset E^i, where $i > N$, does not alter the probability or relative likelihood ranking of the events, namely $W(E_1') > W(E_2')$ where E_1' and E_2' are the altered events.

[7] See De Groot (1970, 2004).

empty set, then regardless of how small is the set B, it is impossible for every set A_i to be as likely as B. Equivalently, the probability of the sets that are far away in the vanishing sequence $\{A_i\}$ must go to zero. Therefore SP_4 implies Monotone Continuity (MC). Reciprocally, assume MC is satisfied. Consider a decreasing sequence of events A_i and define a new sequence by subtracting from each set the intersection of the family, namely $A_1 - \cap_i^\infty A_i$, $A_2 - \cap_i^\infty A_i$, ... Let B be a set that is more likely than the empty set but less likely than every A_i. Observe that the intersection of the new sequence is empty, $\cap_i A_i - \cap_i^\infty A_i = \emptyset$ and since $A_i \supset A_{i+1}$ the new sequence is, by definition, a vanishing sequence. Therefore by MC $\lim_i W(A_i - \cap_i^\infty A_i) = 0$. Since $W(B) > 0$, B must be more likely than $A_i - \cap_i^\infty A_i$ for some i onwards. Furthermore, $A_i = (A_i - \cap_i^\infty A_i) \cup (\cap_i^\infty A_i)$ and $(A_i - \cap_i^\infty A_i) \cap (\cap_i^\infty A_i) = \emptyset$, so that $W(A_i) > W(B)$ is equivalent to $W(A_i - \cap_i^\infty A_i) + W((\cap_i^\infty A_i) > W(B)$. Observe that $W(\cap_i^\infty A_i) < W(B)$ would contradict the inequality $W(A_i) = W(A_i - \cap_i^\infty A_i) + W((\cap_i^\infty A_i) > W(B)$, since as we saw above, by MC, $\lim_i W(A_i - \cap_i^\infty A_i) = 0$, and $W(A_i - \cap_i^\infty A_i) + W((\cap_i^\infty A_i) > W(B)$. It follows that $W(\cap_i^\infty A_i) > W(B)$, which establishes De Groots's Axiom SP_4. Therefore Monotone Continuity is equivalent to De Groot's Axiom SP_4. A proof that each of the two axioms implies countable additivity is in Villegas (1964), Arrow (1971) and De Groot (1970).

The next section shows that the two classic axioms, Monotone Continuity and SP_4, are biased against or neglect rare events, no matter how important these may be.

3 Rare Events and Change

The axioms presented in this article originated from Chichilnisky (1996, 2000, 2002), except for one new axiom - **the Swan Axiom** - that is introduced here and represents the essence of the new probability theory. Below we explain how the Swan Axiom relates to standard theory and its connection with Godel's incompleteness theorem and the Axiom of Choice that are at the foundation of Mathematics.

To explain how the new theory intersects with standard probability or relative likelihood, we compare the results presented here with Savage's (1954) axiomatization of probability measures as finitely additive measures, as well as with Villegas' (1964) and Arrow's (1971) classic work that is based instead on countably additive measures. Savage (1954) axiomatizes subjective probabilities as finitely additive measures representing the decision makers' beliefs, an approach that can ignore frequent events as shown in the Appendix. To overcome this, Villegas (1964) and Arrow (1971) introduced their additional continuity axiom ('Monotone Continuity') that ensures as we saw above the countably additivity of the measures. However this requirement of monotone continuity has unusual implications when the subject is confronted with rare events. A practical example it discussed below: it predicts that in exchange for a couple of cents, one should be willing to accept a small risk of death, a possibility that Arrow himself described as 'outrageous' (1971, p. 48 and 49). The issue of course is the "smallness" of the risk and here is where topology

enters. Monotone continuity has a low bar for smallness while the sup norm has a higher bar as we shall see below. This article defines a realistic solution, and it implies that for some very large payoffs and in certain special situations, one may be willing to accept a small risk of death - but not in others. This means that Monotone Continuity holds in some cases but not in others, a possibility that leads to the axiomatization proposed in this article, which is the logical negation of Monotone Continuity - one that is consistent with recent experimental observations reported in Chanel and Chichilnisky (2009 a,b).

This section explains in what sense standard probability theory is biased against - or disregards - rare events. The next section defines new axioms for relative likelihood, and compares them with the classic axioms. In this section the definitions and results are given for a general measure space of events; the definitions are refined below when the events are Borel measurable sets in the real line R.

Definition 1. A probability W is said to be **biased against rare events** or **insensitive to rare events** when it neglects events that are 'vanishing' according to the definition provided in Section 3 above. Formally, a probability is insensitive to rare events when given two events A and B and any vanishing sequence of events (E_j), $\exists$ $N = N(f,g) > 0$, such that $W(A) > W(B) \Leftrightarrow W(A') > W(B') \ \forall \ A', B'$ satisfying $A' = A$ and $B' = B$ a.e. on $E_j^c \subset R$ when $j > N$.[8] As already discussed this implies a bias against the likelihood of change.

Proposition 2. *A probability satisfies Monotone Continuity if and only if it is biased against rare events and underestimates the likelihood of change.*

Proof. Chichilnisky (2000).

Corollary 1. *Countably additive probabilities are biased against rare events and underestimate change.*

Proof. It follows from Propositions 1 and 2 and Chichilnisky (2000).

Proposition 3. *Purely finitely additive probabilties are biased against frequent events.*

Proof. See Appendix.

The following example illustrates the role of Monotone Continuity and SP_4 in introducing a bias against rare events. The best way to explain the role of Monotone Continuity is by means of the example provided by Kenneth Arrow, Arrow (1971), p. 48 and 49. He explains that if a is an action that involves receiving one cent, b is another that involves receiving zero cents, and c is a third action involving receiving one cent and facing a small probability of death, then *Monotone Continuity* requires that the third action involving death and one cent should be preferred to the action with zero cents when the probability of death is small enough. One accepts a small chance of death in exchange for one cent. Even Arrow says 'this may sound outrageous at first blush...' Arrow (1971) p. 48 and 49. Outrageous or not, Monotone Continuity (MC) leads to neglect rare events that involve change with major consequences, like death. It can be said that death is a black swan: this is the content of Proposition 2 above.

---

[8] Here E^c denotes the complement of the set E, $A' = A$ a.e. on $E_j^c \Leftrightarrow A' \cap E_j^c = A \cap E_j^c$ a.e.

4 New Axioms for Probability Theory: the Topology of Change

This section presents the new axiomatic foundation for probability theory that is neither biased against rare nor against frequent events (Chichilnisky, 2000, 2002).

The new axioms for probability - or relative likelihoods - are as follows:

Axiom 3 *Probabilities are additive and continuous in the topology of change.*

Axiom 4 *Probabilities are unbiased against rare events.*

Axiom 5 *Probabilities are unbiased against frequent events.*

Additivity is a natural condition and **continuity** captures the notion that 'nearby' events are thought as being similarly likely to occur; this property is important to ensure that 'sufficient statistics' exist and it is basaed on a finer topology than Monotone continuity - the sup norm of L_∞ that we called the "topology of change" . Axiom 3 defines continuity with respect to a finer topology Axioms 4 and 5 together are equivalent to the Swan Axiom defined in the previous section, which is required to avoid a bias against rare and frequent events as shown in Section 3. The concept of continuity bears further elaboration. Topology provides the notion of what is meant by 'nearby'; different topologies define different notions of 'nearby' and therefore different notions of what is meant by 'continuity.' For example, 'nearby' was defined in Villegas (1964) and Arrow (1971) as follows: two events are **close** or **nearby** when they differ on a **small set -** thus reducing the problem to determine what is a small set. As stated in Arrow (1971) p. 48: "An event that is far out on a *vanishing sequence* is 'small' by any reasonable standards" Arrow (1971) p. 48. As the sets (n, ∞) are all similar, there is no reason why they become "small" for large enough n, according to Villegas and Arrow.

To overcome the bias against rare events, we introduce a new axiom that is the logical negation of MC: this means that sometimes MC holds and other times it does not. We call this the **Swan Axiom,** and is stated formally below:

Axiom 6 (Swan Axiom) *There exist vanishing sequences of sets $\{U_i\}$ $-$namely, $\forall i, U_{i+1} \subset U_i$ and $\cap U_i = \varnothing-$ where the limit of the measures $\mu(U_i)$ as $i \to \infty$ is not zero.*

Observe that in some cases the measures of the sets in a vanishing family may converge to zero and in other cases they do not. In words, this axiom is the logical negation of Monotone Continuity and can be equivalently described as follows: "There exist events A and B with $W(A) > W(B)$, and for every vanishing sequence of events $\{E_\alpha\}_{=1,2...}$ an $N > 0$ such that altering arbitrarily the events A and B on the set E^i, where $i > N$, alters the probability ranking of the events, namely $W(B') > W(A')$, where B' and A' are the altered events."

Proposition 4. *A probability that satisfies the Swan Axiom is neither biased against rare events, nor biased against frequent events.*

Proof. This is immediate from the definition.

Example: To illustrate how this axiom works in practice consider an experiment where the subjects are offered a certain amount of money to choose and eat a pill at random from a pile that contains one pill that causes death (Chanel and Chichilnisky 2009 a,b). Experimentally, it is observed that in some cases people accept a sum of money and choose a pill provided the pile is large enough - namely when the probability of death is small enough – thus satisfying the monotone continuity axiom and in the process determining a statistical value for their lives. But there are also cases where the subjects will not accept to choose any pill, no matter how large is the pile. Some people refuse the payment if it involves a small probability of death, no matter how small the probability may be (Chanel and Chichilnisky, 2009, a,b). This conflicts with the Monotone Continuity axiom, as explicitly presented by Arrow (1971). Our Axiom provides a reasonable resolution to this dilemma that is realistic and consistent with the experimental evidence. It implies that there exist catastrophic outcomes such as the risk of death, so terrible that one is unwilling to face a small probability of death to obtain one cent versus half a cent, no matter how small the probability may be. According to our Swan Axiom, no probability of death may be acceptable when only one cent and half a cent are involved. Our Axiom also implies that in other cases there may be a small enough probability that the lottery involving death may be acceptable, or that the payoff is large enough to justify the small risk. This is a possibility discussed by Arrow (1971), where he explains that for large payoffs (for example, one billion US dollars), one may be willing to accept a small probability of death. In other words: sometimes one is willing to take a risk of death with a small enough probability of a catastrophe, and in other cases one is not. This is the content of the Swan Axiom.

We saw in Proposition 2 that the notion of continuity defined by Villegas and Arrow - namely Monotone Continuity – conflicts with the Swan Axiom and neglects rare events. Indeed Proposition 1 shows that countably additive measures are biased against rare events. On the other hand, Proposition 3 and the Example in the Appendix show that purely finitely additive measures can also be biased, in this case against frequent events. A natural question is whether it is possible to eliminate simultaneously both biases. The following theorem addresses this issue:

Theorem 1. *A probability that satisfies the Swan Axiom is neither biased against frequent nor against rare events. The resulting measures are neither purely finitely additive nor countably additive. They are a strict convex combinations of both.*

Proof. The next Section contains a proof of Theorem 1 and provides examples when the events are Borel sets in R or within an interval $(a,b) \subset R$.

Theorem 1 establishes that neither Savage's approach, nor Villegas' and Arrow's approaches, satisfy the three new axioms stated above. These three axioms require more than the additive probabilities of Savage, since purely finitely additive probabilities are finitely additive and yet they must be excluded here; at the same time the axioms require less than the countably additivity of Villegas and Arrow, since countably additive probabilities are biased against rare events. Theorem 1 above shows that a strict combination of both does the job.

Theorem 1 shows how the Swan Axiom resolves the bias problem against frequent and rare events, but it does not by itself prove the existence of likelihoods that satisfy all three axioms. What is missing is an appropriate definition of 'nearby', namely of topology and continuity, that does not conflict with the Swan Axiom. The following shows that this can be achieved.

We now specialize the space of measurable sets so they are Borel measurable subsets of the real line R, and consider the Lebesgue measure on R. In this context a probability or likelihood function $W : L_\infty \to R$ is called **biased against rare events**, or **insensitive to rare events** when it neglects events that are small according to a probability measure μ on R that is absolutely continuous with respect to the Lebesgue measure. Formally:

Definition 2. A probability is insensitive to rare events when given two events f and $g \; \exists \; \varepsilon = \varepsilon(f, g) > 0$, such that $W(f) > W(g) \Leftrightarrow W(f') > W(g') \; \forall \; f', g'$ satisfying $f' = f$ and $g' = g$ a.e. on $A \subset R$ and $\mu(A^c) < \varepsilon$. Here A^c denotes the complement of the set A.

Definition 3. A probability or likelihood function $W : L \to R$ is said to be **insensitive to frequent events** when given any two events $f, g \; \exists \; \varepsilon(f, g) > 0$ that $W(f) > W(g)$ $\Leftrightarrow W(f') > W(g') \; \forall \; f', g'$ satisfying $f' = f$ and $g' = g$ a.e. on $A \subset R$ and $\mu(A^c) > 1 - \varepsilon$.

Definition 4. W is called **sensitive** to rare (or frequent) events when it is **not insensitive** to rare (or frequent) events.

Below we identify an event with its characteristic function, so that events are contained in the space of bounded real valued functions on the universe space $\mathscr{U}$, $L_\infty(\mathscr{R})$, and endow this space with the sup norm rather than with the notion of smallness and continuity defined by Arrow and Villegas as described above. In this case the probability or likelihood $W : L_\infty(\mathscr{U}) \to R$ is taken to be continuous with respect to the sup norm. Events are elements of the Borel measurable sets of the real line R or an interval (a, b), they are identified with the characteristic functions, denoted f, g etc, and 'continuity' is based on a topology used earlier in Debreu (1953) and in Chichilnisky (2000, 2002, 2009 a,b), the sup norm $\| f \| = \underset{x \in R}{\operatorname{ess\,sup}} | f(x) |$. This is a sharper and more stringent definition of closeness than the one used by Villegas and Arrow, since an event can be small under the Villegas-Arrow definition but not under ours, see the Appendix for examples. The difference in the use of topologies as shown below achieves sensitivity to rare events. To simply notation, a probability that satisfies the classic axioms in De Groot (1970) is from now on called a **standard probability,** and is therefore countably additive. As already mentioned, a classic representation result is that for any event $f \in L_\infty$ a standard (countably additive) probability has the form $W(f) = \int_R f(x).\phi(x)d\mu$, where $\phi \in L_1(R)$ is an integrable function in R..

The next step is to show existence and characterize all the likelihoods or probability distributions that satisfy the 3 new axioms. The following three axioms are identical to the three axioms above, specialized to the case at hand, Borel sets of R,

and measures in L_∞ with the topology defined by the *sup norm* on $L_\infty(R)$, which we called "the topology of change"

Axiom 7 *$W : L_\infty \to R$ is linear and continuous with the sup norm or "topology of change".*

Axiom 8 *$W : L_\infty \to R$ is sensitive to frequent events.*

Axiom 9 *$W : L_\infty \to R$ is sensitive to rare events.*

The first and the second axiom agree with classic theory and standard likelihoods satisfy them. The third axiom is new.

Lemma 1. *A standard probability satisfies Axioms 7 and 8, but it is biased against rare events and therefore does not satisfy Axiom 9.*

Proof. Consider $W(f) = \int_R f(x)\phi(x)dx$, $\int_R \phi(x)dx = K < \infty$. Then

$$W(f)+W(g) = \int_R f(x)\phi(x)dx + \int_R g(x)\phi(x)dx = \int_R f(x)+g(x).\phi(x)dx = W(f+g),$$

and therefore W is linear. It is **continuous** with respect to the L_1 norm $\| f \|_1 = \int_R | f(x) | \phi(x)d\mu$ because $\| f \|_\infty < \varepsilon$ implies

$$W(f) = \int_R f(x).\phi(x)dx \leq \int_R | f(x) | .\phi(x)dx \leq \varepsilon \int \phi(x)dx = \varepsilon K.$$

Since the sup norm is finer than the L_1 norm, continuity in L_1 implies continuity with respect to the sup norm (Dunford and Schwartz, 1958). Thus a standard probability satisfies Axiom 1. It is obvious that for every two events f, g, with $W(f) > W(g)$, the inequality is reversed namely $W(g') > W(f')$ when f' and g' are appropriate variations of f and g that differ from f and g on sets of sufficiently large Lebesgue measure. Therefore Axiom 2 is satisfied. A standard probability is however not sensitive to rare events, as shown in Chichilnisky (2000, 2002, 2006, 2008, 2009 a,b).

5 Existence and Representation Theorems

Theorem 2. *There exists a probability distribution or likelihood function $W : L_\infty \to R$ satisfying the new Axioms 7, 8, and 9. A probability distribution satisfies Axioms 7, 8, and 9 if and only if there exist two continuous linear functions on L_∞ denoted ϕ_1 and ϕ_2, and a real number λ, $0 < \lambda < 1$, such that for any observable event $f \in L_\infty$ the likelihood*

$$W(f) = \lambda \int_{x \varepsilon R} f(x)\phi_1(x)dx + (1-\lambda)\phi_2(f) \tag{1}$$

where $\phi_1 \in L_1(R,\mu)$ defines a countably additive measure on R and where ϕ_2 is a purely finitely additive measure.

Proof. This result follows from the representation Theorem in Chichilnisky (2000, 2002).

Corollary 2. *Absent rare events, a probability that satisfies Axioms 7, 8, and 9 is consistent with classic axioms and yields a countably additive measure.*

Proof. Axiom 9 is an empty requirement when there are no rare events while, as shown above, Axioms 7 and 8 are consistent with standard relative likelihood.

6 Heavy Tails and Families of Purely Finitely Additive Measures

This section presents new results adding to the introduction of the Swan Axiom 6 defined in Section 4 above: the different notions of continuity, how heavy tails originate from the new axioms and defines a family of purely finitely additive measures that are each absolutely continuous with respect to the Lebesgue measure on *R*.

A main difference introduced by the new axioms is the use of a finer topology - the "topology of change", which is the sup norm on L_∞, and defines the continuity properties of probability distributions. In the classic axioms a probability distribution is continuous if it satisfies Monotone Continuity or equivalently SP4. Here the continuity required is with respect to the topology of change, which is a finer topology. The following example explains the difference that this makes on the concept of continuity of probability distributions:

6.1 Contrasting Monotone Continuity and the Topology of Change

Different topologies define different approaches to 'continuity'. Consider the family $\{E^i\}$ where $E^i = [i, \infty)$, $i = 1, 2, ...$ This is a vanishing family because $\forall i \, E^i \supset E^{i+1}$ and $\bigcap_{i=1}^{\infty} E^i = \varnothing$. Consider now the events $f^i(t) = K > 0$ when $t \in E^i$ and $f^i(t) = 0$ otherwise, and $g^i(t) = 2K$ when $t \in E^i$ and $g^i(t) = 0$ otherwise. Then for all i, $\sup_{E^i} | f^i(t) - g^i(t) | = K$. In the sup norm topology this implies that f^i and g^i are **not** 'close' to each other, as the difference $f^i - g^i$ does not converge to zero. No matter how far along we are along the vanishing sequence E^i the two events f^i, g^i differ by at least the number K. Yet since the events f^i, g^i differ from $f \equiv 0$ and $g \equiv 0$ respectively only in the set E^i, and $\{E^i\}$ is a vanishing sequence, for large enough i they are as 'close' as desired according to Villegas-Arrow's definition of 'nearby' events.

6.2 *Heavy Tails*

The following illustrates the additional weight that the new axioms assign to rare events; in this example in the form of 'heavy tails' (e.g. Chichilnisky (2000). The finitely additive measure ϕ_2 appearing in the second term in 1 can be illustrated as follows. On the subspace of events with limiting values at infinity, $L'_\infty = \{f \varepsilon L_\infty : \lim_{x \to \infty}(x) < \infty\}$, define $\phi_2(f) = \lim_{x \to \infty} f(x)$ and extend this to a function on all of L_∞ using Hahn Banach's theorem. The difference between a standard probability and the likelihood defined in (1) is the second term ϕ_2, which focuses all the weight at infinity. This can be interpreted as a 'heavy tail' namely a part of the distribution that is not part of the standard density function ϕ_1 and gives more weight to the sets that contain *terminal* events namely sets of the form (x, ∞).

6.3 *The Family* PA *of Purely Finitely Additive Measures on* R

This section provides a new family of purely finitely additive measures that are absolutely continuous with respect to the Lebesgue measure, and studies its properties.

Definition 5. An *open neighborhood* of a real number $x \in R$ has the standard meaning under the usual topology of the line R. An *'open neighborhood of* ∞' is defined to be a set of the form $\{x \in R : x. > r \text{ for } r \in R\}$. As already stated, the word "essentially" means a.e. with respect to the Lebesgue measure on R that has been used to define the space L_∞.

We now define a property on measures in the space L_∞^* :

Definition 6 (Property (P)). A measure in L_∞^* is said to satisfy Property (P) at x if it assigns measure zero to any set that is essentially contained in the complement of an open neighborhood of x. A measure in L_∞^* is said to satisfy Property (P) at ∞, if it assigns measure 0 to any measurable set that is essentially contained in the complement of an open neighborhood of ∞ as defined above. A measure is said to satisfy Property (P) if it satisfies Property (P) either at ∞ or at any $x \in R$.

Lemma 2. *A measure satisfying property* (P) *is always purely finitely additive.*

Proof. Consider first the case where the measure has property (P) at ∞. Define a countable family of disjoint sets $F = \{A_1, A_2 ...\}$ recursively as follows: $A_1 = \{x : -1 < x < 1\}$ and for all n, $A_n = \{x : -n < x < n\} - A_{n-1}$. Observe that each set A_n has measure zero, since by assumption μ satisfies property (P), and that each of the sets in the family F is bounded. The sets in the family F are also disjoint by construction. If μ was countably additive, then we should have $\mu(\cup F) = \mu(\cup_{n=1}^\infty A_n) = \sum_{n=1}^\infty \mu(A_n) = 0$. Yet the measure of the union of the countable family F is not 0, because $\cup F = R$, the entire real line, so that $\mu(\cup F) = 1$. Therefore μ fails to be countably additive on the countable and disjoint family F.

Since by definition μ is a measure, and it fails to be countably additive, it must be a purely finitely additive measure.

A similar argument can be given for the case where the measure has property (P) at a finite number $x \in R$. Define now $F = \{A_n\}_{n=1,2,...}$ recursively as follows: $A_1 = [x-1, x+1]^c$ where the super-index c denotes the complement of a set, and for all $n \geq 1$, $A_n = [x-1/n, x+1/n]^c - A_{n-1}$. Observe that each set in the family F has measure 0. The union of the family is not the whole space as before - since the point $\{x\}$ is not in the union; yet the entire space minus $\{x\}$ should have the same measure than the space as a whole, because by definition a measure is a continuous linear function on L_∞, the space of measurable and essentially bounded functions with the Lebesgue measure on R, which means that the measure must provide the same value to functions in L_∞ that are essentially equal, in the sense of differing only in a set of Lebesgue measure 0. The characteristic functions of two measurable sets differing in a set of measure zero, must therefore be are assigned the same value by a measure, so the union of the family F must be assigned the same measure as the entire space, namely $\mu(\cup F) = 1$. Therefore the measure μ fails to be countably additive, and since it is a measure it must be purely finitely additive.

Observe that in Lemma 1 the same argument applies for a measure that has property (P) at x for a finite $x \in R$, or one that has property (P) at $\{\infty\}$. The "test" family F is defined similarly in both cases, where for a finite x $A_1 = \{x : -\varepsilon < x < \varepsilon\}$, and $A_n = \{x : -n < x < n\} - A_{n-1}$. The only difference in the argument arises from the fact that, for a finite $\{x\}$, the union of the family $\cup F$ is not all of R, but rather $R - \{x\}$. But this is essentially the same as R in the Lebesgue measure used to define $L\infty$.

Lemma 3. *Using Hahn-Banach's theorem it is possible to define purely finitely additive measures on R.*

Proof. Lemma 1 started from assuming the existence of a measure in L_∞^* that satisfies property (P) at ∞. Using Hahn Banach's theorem we now define the desired measure, namely a continuous linear function h from L_∞ to R, and show that it satisfies (P) at ∞. Therefore by Lemma 1, the function h is a purely finitely additive measure, as we wished to prove.

Consider the subspace CL_∞ of all functions f in L_∞ that are continuous and have an essential limit at ∞. CL_∞ is a closed linear subspace of the Banach space $L\infty$. On the subspace CL_∞ define the function $h(f) = ess\lim_{x \to \infty} f(x)$. By construction the function h is well defined on CL_∞; this function is continuous, linear and has norm 1. The function h can therefore be extended by using Hahn-Banach's theorem to all of L_∞, as a continuous, linear function that preserves the norm of h. Since h has norm 1 the extension is not the zero function. Call this extension h as well; by construction, $h \in L_\infty^*$. Therefore by definition, the extended function h defines a measure. Now observe that $h : L_\infty \to R$ satisfies Property (P) since when applied to characteristics functions of bounded sets, it assigns to them measure zero. A similar argument can be replicated to show the existence of purely finite measures that satisfy property (P) at any $x \in R$.

We have mentioned that it is not possible to construct a purely finitely additive measure on R the same way as one constructs a countably additive measure on R. This is not surprising since the Hanh-Banach Theorem that is used to define a purely finitely additive measure in Lemma 2 is itself not constructible. The next and last section show the connection between the new axioms for probability (or relative likelihoods) presented here and the Axioms of Choice and Godel's (1940) work.

7 The Axiom of Choice and Godel's Incompleteness theorem

There is a connection between Axioms 3, 4, and 5 presented here and the Axiom of Choice that is at the foundation of mathematics (Godel, 1940). The Axiom of Choice postulates that there exists a universal and consistent fashion to select an element from every set.

The best way to describe the situation is by means of an example, see also Dunford and Schwartz (1958), Yosida (1952, 1974), Chichilnisky and Heal (1997) and Kadane and O'Hagan (1995).

Example: Representing a purely finitely additive measure.

Define a measure ρ as follows: for every Borel measurable set $A \subset R$, $\rho(A) = 1$ if $A \supset \{x : x > a,$ for some $a \in R\}$, and otherwise $\rho(A) = 0$. Then ρ is not countably additive, because the family of countably many disjoint sets $\{V_i\}_{i=0,1,\dots}$ defined as $V_i = (i, i+1] \bigcup (-i-1, -i]$, satisfy $V_i \bigcap V_i = \varnothing$ when $i \neq j$, and $\bigcup_{i=0}^{\infty} V_i = \bigcup_{i=0}^{\infty} (i, i+1] \bigcup (-i-1, -i] = R$, so that $\rho(\bigcup_{i=0}^{\infty} V_i) = 1$, while $\sum_{i=0}^{\infty} \rho(V_i) = 0$, which contradicts countable additivity. Since the contradiction arises from assuming that ρ is countably additive, ρ must be purely finitely additive. Observe that ρ assigns zero measure to any bounded set, and a positive measure only to unbounded sets that contain a 'terminal set' of the form

$$\{x \in R : x > a \text{ for some } a \in R\}.$$

One can define a function on L_∞ that represents this purely finitely additive measure ρ if we restrict our attention to the closed subspace L'_∞ of L_∞ consisting of those functions $f(x)$ in L_∞ that have a limit when $x \to \infty$, by the formula $\rho(f) = \lim_{x \to \infty} f(x)$, as in Lemma 3 of the previous section. The function $\rho(.)$ can be seen as a limit of a sequence of delta functions whose support increases without bound. The problem is now to extend the function ρ to another defined on the entire space L_∞. This could be achieved in various ways but as we will see, each of them requires the Axiom of Choice.

One can use Hahn - Banach's theorem (Dunford Schwartz, 1958) to extend the function ρ from the closed subspace $L'_\infty \subset L_\infty$ to the entire space L_∞ preserving its norm. However, in its general form Hahn - Banach's theorem requires the Axiom

of Choice (Dunford Schwartz, 1958). Alternatively, one can extend the notion of a *limit* to encompass all functions in L_∞ including those with no standard limit. This can be achieved by using the notion of convergence along a *free ultrafilter* arising from compactifying the real line R as in Chichilnisky and Heal (1997). However the existence of a *free ultrafilter* also requires the Axiom of Choice.

This illustrates why the attempts to construct *purely finitely additive measures* that are representable as functions on L_∞, require the Axiom of Choice. Since our criteria require purely finitely additive measures, this provides a connection between the Axiom of Choice and our axioms for relative likelihood. It is somewhat surprising that the consideration of rare events that are neglected in standard statistical theory conjures up the Axiom of Choice, which is independent from the rest of mathematics (Godel, 1940).

Appendix

Example: A probability that is biased against frequent events.

Consider $W(f) = \liminf_{x \varepsilon R}(f(x))$. This is insensitive to frequent events of arbitrarily large Lebesgue measure (Dunford and Schwartz, 1958) and therefore does not satisfy Axiom 2. In addition it is not linear, failing Axiom 1.

Example: The dual space L_∞^* consists of countably additive and finitely additive measures.

The space of continuous linear functions on L_∞ is the 'dual' of L_∞, and is denoted L_∞^*. It has been characterized e.g. in Yosida (1952, 1974). L_∞^* consists of the sum of two subspaces (i) L_1 functions g that define countably additive measures ν on R by the rule $\nu(A) = \int_A g(x)dx$ where $\int_R | g(x) | dx < \infty$ so that υ is *absolutely continuous* with respect to the Lebesgue measure, and (ii) a subspace consisting of purely finitely additive measure. A countable measure can be identified with an L_1 function, called its 'density,' but purely finitely additive measures cannot be identified by such functions.

Example: A finitely additive measure that is not countably additive.

See Example in Section 7.

References

1. Anscombe FJ and Aumann RJ (1963) A Definition of Subjective Probability Annals of Mathematical Statistics Vol 43, 199-295.
2. Arrow K (1971) Essays in the Theory of Risk Bearing. (North Holland, Amsterdam)
3. Chanel O and Chichilnisky G (2009a) The Influence of Fear in Decisions: Experimental Evidence *Journal of Risk and Uncertainty,* December 2009, Vol 39, No. 3.
4. Chanel O and Chichilnisky G (2009b) The Value of Life: Theory and Experiments Working Paper, GREQE, Universite de Marseille and Columbia University, New York.
5. Chichilnisky G (1996) An Axiomatic Approach to Sustainable Development. Soc Choice Welf 13:321-257.
6. Chichilnisky G (1996) Updating Von Neumann Morgenstern Axioms for Choice under Uncertainty. Proceedings of a Conference on Catastrophic Risks, The Fields Institute for Mathematical Sciences, Toronto Canada.
7. Chichilnisky G (2000) An Axiomatic Approach to Choice under Uncertainty with Catastrophic Risks. Resource Energ Econ 22:221-231.
8. Chichilnisky G (2002) Catastrophic Risk. In Encyclopedia of Environmetrics, eds El-Shaarawi AH and Piegorsch WW (John Wiley & Sons, Chichester). Vol 1 pp 274-279
9. Chichilnisky G (2009) The Limits of Econometrics: Non Parametric Estimation in Hilbert Spaces. Econ Theor.
10. Chichilnisky G (2009) The Topology of Fear. J Math Econ, December 2009.
11. Chichilnisky G (2009) Choice under Uncertainty: The Work and Legacy of Kenneth Arrow Encyclopedia of Quantitative Finance, ed Cont R. (John Wiley and Sons, Chichester), in press.
12. Chichilnisky G and Heal G M (1997) Social Choice with Infinite Populations. Soc Choice Welf 14:303-319.

13. Chichilnisky G and Wu H-M (2006) General Equilibrium with Endogenous Uncertainty and Default. J Math Econ 42: 499–524.
14. Debreu G (1953) Valuation Equilibrium and Pareto Optimum. Proceedings of the National Academy of Sciences 40:588-592.
15. De Groot MH (1970 and 2004) Optimal Statistical Decisions. (John Wiley and Sons, Hoboken, New Jersey)
16. Dubins L (1975) Finitely Additive Conditional Probabilities, Conglomerability and Disintegration. The Annals of Probability 3:89-99.
17. Dubins L and Savage L (1965) How to Gamble if you Must (McGraw Hill, New York)
18. Dunford N and Schwartz JT (1958) Linear Operators, Part I (Interscience, New York)
19. Godel K (1940) The Consistency of the Continuum Hypothesis. Annals of Mathematical Studies 3, Princeton University Press, Princeton NJ USA.
20. Kadane JB and O'Hagan A (1995) Using finitely additive probability: uniform distributions of the natural numbers. J Am Stat Assoc 90:525-631.
21. Kolmogorov, A.N. (1933/1950) *Foundations of the Theory of Probability* New York: Chelsea Publishing Co.
22. Le Doux J (1996) The Emotional Brain (Simon and Schuster, New York)
23. Purves RW and Sudderth (1976) Some Finitely Additive Probability. The Annals of Probability 4:259-276.
24. Savage L The Foundations of Statistics, John Wiley and Sons New York 1054, revised edition 1972.
25. Villegas C (1964) On Quantitative Probability $\sigma-$Algebras. Ann Math Stat 35:1789 - 1800.
26. Von Neumann J (1932/1955) Mathematical Foundations of Quantuum Theory New Jersey: Princeton University Press.
27. Von Neumann J and Morgernstern O (1944) Theory of Games and Economic Behavior New Jersey:Princeton University Press.
28. Yosida K (1974) Functional Analysis, 4th edition (Springer Verlag, New York and Heidelberg)
29. Yosida K and Hewitt E (1952) Finitely Level Independent Measures. Trans Amer Math Soc 72:46-66.