Two theories of conditional probability and non-conglomerability

Teddy Seidenfeld **Mark J. Schervish** **Joseph B. Kadane**
Carnegie Mellon University
teddy@stat.cmu.edu mark@stat.cmu.edu kadane@stat.cmu.edu

Abstract

Conglomerability of conditional probabilities is suggested by some (e.g., Walley, 1991) as necessary for rational degrees of belief. Here we give sufficient conditions for non-conglomerability of conditional probabilities in the de Finetti/Dubins sense. These sufficient conditions cover familiar cases where P(·) is a continuous, countably additive probability. In this regard, we contrast the de Finetti/Dubins sense of conditional probability with the more familiar account of regular conditional distributions, in the fashion of Kolmogorov.

Keywords. Non-conglomerability, conditional probability, κ-additive probability, regular conditional distribution.

1 Introduction

Consider a finitely, but not necessarily countably additive probability P(·) defined on a sigma-field of sets $\mathcal{B}$, each set a subset of the sure-event Ω. In other terms, $<\Omega, \mathcal{B}, P>$ is a (finitely additive) measure space.

We begin by reviewing the theory of conditional probability that we associate with de Finetti (1974) and Dubins (1975).

Let B, C, D, E, F, G $\in \mathcal{B}$, with B ≠ ∅ and F ∩ G ≠ ∅.

Definition 1. A *conditional probability* P(· | B) satisfies the following three conditions:

(i) P(C ∪ D | B) = P(C | B) + P(D | B), whenever B ∩ C ∩ D = ∅;

(ii) P(B | B) = 1.

In order to regulate conditional probability given a non-empty *null* event, i.e., one that itself may be of unconditional or conditional probability 0, we require the following.

(iii) P(E ∩ F | G) = P(E | F ∩ G)P(F | G).

Throughout, we follow the usual identification of unconditional probability with conditional probability given the sure-event, P(·) = P(· | Ω).

This account of conditional probability is not the usual theory from contemporary Mathematical Probability, which we associate with Kolmogorov (1956). That theory, instead, defines conditional probability through regular conditional distributions, as follows.

Let $\mathcal{A}$ be a sub-σ-field of $\mathcal{B}$.

Definition 2. P(· | $\mathcal{A}$) is a *regular conditional distribution* [rcd] on $\mathcal{B}$, given $\mathcal{A}$ provided that:

1. For each ω ∈ Ω, P(·|$\mathcal{A}$)(ω) is a countably additive probability on $\mathcal{B}$.
2. For each B ∈ $\mathcal{B}$, P(B| $\mathcal{A}$)(·) is an $\mathcal{A}$-measurable function.
3. For each A ∈ $\mathcal{A}$,

$$P(A \cap B) = \int_A P(B \mid \mathcal{A})(\omega)\, dP(\omega).$$

That is, P(B| $\mathcal{A}$) is a version of the Radon-Nikodym derivative of P(· ∩ B) with respect to P(·).

Definition 3: An *$\mathcal{A}$-atom* is the intersection of all elements of $\mathcal{A}$ that contain a given point ω of Ω.

When P(A) > 0 and ω ∈ A ∈ $\mathcal{A}$ and A is an $\mathcal{A}$-atom, then

$$P(B \mid \mathcal{A})(\omega) = P(A \cap B) / P(A).$$

The theory of conditional probability that we use here differs from the received theory of Kolmogorovian regular conditional distributions in at least five ways:

(1) The theory of regular conditional distributions requires that probabilities and conditional probabilities are countably additive. The theory of conditional probability from Definition 1 requires only that probability is finitely additive. In this note we bypass this difference by exploring countably additive conditional probabilities.

(2) When P(A) = 0 and A is not empty, a regular conditional probability given A is relative also to a sub-sigma field $\mathcal{A} \subseteq \mathcal{B}$, where A ∈ $\mathcal{A}$. By contrast, in the theory of conditional probability, P(· | A), depends solely on the event A and not on any sub-field that embeds it. Example 2, below, illustrates this difference.

(3) Dubins (1975) establishes that for each set Ω there is a *full* conditional probability function, P(B|A), defined whenever A ≠ ∅ and B are elements of $\mathcal{B}$, the powerset of Ω. However, some countably additive probabilities do not admit regular conditional distributions relative to a particular sub-sigma field $\mathcal{A} \subseteq \mathcal{B}$, even when each sigma-field, $\mathcal{A}$ and $\mathcal{B}$, is countably generated. The canonical example of a measure space that admits no rcd's is obtained by extending the σ-field of Borel sets on [0,1] under Lebesgue measure, μ, with the addition of one non-measurable set.

Denote the initial measure space by $<[0,1], \mathcal{B}, \mu>$. A familiar maneuver allows an extension of $\mathcal{B}$ to a larger σ-field of sets, $\mathcal{B}'$, generated by adding one Lebesgue non-measurable set to $\mathcal{B}$, and an extension of μ to a countably additive probability μ′ over $\mathcal{B}'$. However, there is no rcd $\mu'(\cdot|\mathcal{A})(\omega)$ on $\mathcal{B}'$ given $\mathcal{A}$ when, e.g., $\mathcal{A} = \mathcal{B}$. (See Halmos, 1950, p. 211; Billingsley, 1986, Exercise 33.13; Breiman, 1968, p.81; Doob, 1953, p. 624; or Loeve, 1955, p. 370 for variations on this common theme.) Though, for each $B \in \mathcal{B}$, the extended measure space has Radon-Nikodym derivatives P(B |·) satisfying condition 3, above, these resist assembly of pointwise probabilities into a countably additive probability distribution over $\mathcal{B}'$ as required by condition 1.

In our (2001, Corollary 1) we show that, quite generally, a measure space admitting rcd's can be extended to another measure space admitting rcd's if and only if the latter lies within the measure completion of the former. In rejoinder to the existence problem, however, a sufficient condition for rcd's to exist on $\mathcal{B}$ (given any sub σ-field $\mathcal{A}$) is that $\mathcal{B}$ be isomorphic under a 1-1 measurable mapping to the σ-field of a random variable. (See, Billingsley 1986, T.33.3; or Breiman, 1968, T. 4.30.)

(4) Blackwell (1955), Blackwell and Ryll-Nardzewski (1963) and Blackwell and Dubins (1975) introduce an additional constraint, *propriety* of an rcd, matching condition (ii) of de Finetti/Dubins' theory of conditional probabilities.

Definitions 4:

- An rcd $P(\cdot|\mathcal{A})(\omega)$ on $\mathcal{B}$ given $\mathcal{A}$, is *proper at* ω if $P(A|\mathcal{A})(\omega) = 1$ whenever $\omega \in A \in \mathcal{A}$.
- $P(\cdot|\mathcal{A})(\omega)$ is *improper at* ω, otherwise.
- $P(\cdot|\mathcal{A})$ is *proper* if $P(\cdot|\mathcal{A})(\omega)$ is proper for each $\omega \in \Omega$.

Definition 5: Say that a probability distribution is *extreme* if its range is the two point set {0,1}.

Theorem 1 (Blackwell and Dubins, 1975) When $\mathcal{B}$ is a countably generated σ-field, *no* rcd on $\mathcal{B}$ given $\mathcal{A}$ is proper if there exists *some* extreme probability on $\mathcal{A}$ supported by no $\mathcal{A}$-atom belonging to $\mathcal{A}$.

In other words, provided there exists even one extreme probability on $\mathcal{A}$ which is supported by none of its $\mathcal{A}$-atoms, then the sub-σ-field $\mathcal{A}$ is anomalous for all rcd's on $\mathcal{B}$ given $\mathcal{A}$ in that they are improper, each and every one! However, this result does not identify at how many points, ω, or how badly, the rcd is improper. The following result addresses that question.

Assume that $\mathcal{A}$ is an atomic sub-σ-field of $\mathcal{B}$, with $\mathcal{A}$-atoms *a*. Denote by $a(\omega)$ that $\mathcal{A}$-atom containing the point ω.

Theorem 2 (our 2001): Let P be an extreme probability on $\mathcal{A}$ that is not supported by any of its $\mathcal{A}$-atoms. If an rcd $P(\cdot|\mathcal{A})(\omega)$ on $\mathcal{B}$ given $\mathcal{A}$ exists, there is one where $P\{\omega: P(a(\omega)|\mathcal{A})(\omega) = 0\} = 1$. And, if $\mathcal{B}$ is countably generated, then this rcd is unique.

Theorem 2 asserts that when $\mathcal{B}$ is countably generated and the antecedent of Theorem 1 is satisfied, then almost surely with respect to P, the rcd's on $\mathcal{B}$ given $\mathcal{A}$ are maximally improper, in two senses simultaneously:

- The set of points where propriety fails has measure 1 under P.
- For P-almost all points ω, $P(a(\omega)|\mathcal{A})(\omega) = 0$ when propriety requires that $P(a(\omega)|\mathcal{A})(\omega) = 1$.

The following Corollary applies Theorem 2 when conditioning on the sub-sigma field associated with de Finetti's theorem on exchangeability.

Let $\Omega = \{0,1\}^{\aleph_0}$; let $\mathcal{B}$ = the Borel subsets of Ω; and let P be a symmetric probability, in the sense of Hewitt and Savage (1955) defined as follows. Let *T* be an arbitrary finite permutation of the positive integers, i.e., a permutation of the coordinates of Ω that leaves all but finitely many places fixed. For $B \in \mathcal{B}$, given *T*, define the set $T^{-1}B = \{\omega: T(\omega) \in B\}$.

Definitions 6:

- P is called a *symmetric probability* if $P(T^{-1}B) = P(B)$, for each $B \in \mathcal{B}$ and *T*.
- If $B = T^{-1}B$ for all (finite) permutations *T*, B is called *a symmetric event*.

Let $\mathcal{A}$ be the sub-σ-field of $\mathcal{B}$ generated by the class $\boldsymbol{T}$ of all finite permutations of the coordinates of Ω, i.e., $\mathcal{A}$ is the σ-field of the symmetric events. $\mathcal{A}$ is atomic, with $\mathcal{A}$-atoms comprised by a countable set of sequences, each pair of sequences in the same atom differing by some finite permutation of its coordinates. In all but two cases the $\mathcal{A}$-atoms are countably infinite sets. The two exceptions are the two constant sequences, which are singleton sets.

Corollary (see our 2001). Each rcd $P(\cdot|\mathcal{A})(\omega)$ on $\mathcal{B}$ given $\mathcal{A}$, for a symmetric probability P, satisfies

$$P\{\omega: P(a(\omega) \mid A)(\omega) = 0)\} = 1,$$

provided that $P(<0,0,0,\ldots>) = P(<1,1,1\ldots>) = 0$.

For additional related results see (Berti and Rigo, 2007)

(5) Our focus in this paper is on a fifth feature that distinguishes the de Finetti/Dubins theory of conditional probability and the Kolmogorovian theory of regular conditional probability. This aspect of the difference involves *conglomerability* of conditional probability functions.

Let $E \in \mathcal{B}$, let $\boldsymbol{N}$ be an index set and let $\pi = \{h_\nu: \nu \in \boldsymbol{N}\}$ be a partition of the sure event where the conditional probabilities, $P(E \mid h_\nu)$, are well defined for each $\nu \in \boldsymbol{N}$.

Definition 7: The conditional probabilities $P(E \mid h_\nu)$ are *conglomerable* in π provided that, for each event $E \in \mathcal{B}$ and arbitrary real constants k_1 and k_2,

if $k_1 \leq P(E \mid h_\nu) \leq k_2$ for each $\nu \in \mathbf{N}$, then $k_1 \leq P(E) \leq k_2$.

In our (1984) we show that if P is merely finitely additive (i.e., if P is finitely but not countably additive) with conditional probabilities that satisfy Definition 1, then P fails conglomerability in some countable partition. That is, for each merely finitely additive probability P there is an event E, an $\varepsilon > 0$, and a countable partition $\pi = \{h_n: n = 1, \ldots\}$, where $P(E) > P(E \mid h_n) + \varepsilon$ for each $h_n \in \pi$.

The following illustrates a failure of conglomerability for a merely finitely additive probability P in a countable partition $\pi = \{h_n: n \in \{1, 2, \ldots\}\}$, where each element of the partition is not null, i.e., $P(h_n) > 0$, n = 1, 2, … .

Then, apart from the requirement of countable additivity, both theories agree on the relevant conditional probabilities: $P(E \mid h_n) = P(E\cap h_n)/P(h_n)$ is well defined. Thus, the failure of conglomerability in this example is due to the failure of countable additivity, rather than to a difference in how conditional probability is defined.

***Example* 1** (Dubins, 1975): Let $\boldsymbol{\Omega} = \{(i, n): i \in \{1, 2\}$ and $n \in \{1, 2, \ldots\}\}$ and let $\mathcal{B}$ be the powerset of $\boldsymbol{\Omega}$.
Let event $E = \{\{1, n\}: n \in \{1, 2, \ldots\}\}$ and events $h_n = \{\{1,n\}, \{2,n\}\}$, n = 1, ... Observe that the h_n form a partition: $\pi = \{h_n: n \in \{1, 2, \ldots\}\}$.
Partially define the (unconditional) probability P by

(a) $P(\{(i, n)\}) = 1/2^{n+1}$ if i = 1, n = 1, 2, ...
(b) $P(\{(i, n)\}) = 0$ if i = 2, n = 1, 2, ...
(c) $P(E) = 0.5$.

So P is countably additive given E, and strongly finitely additive given E^c. (A finitely additive probability is strongly finitely additive if there is a countable partition of the sure event each of whose elements is null.)
Clearly, $P(h_n) = P(\{(1,n)\}) + P(\{2,n)\}) = 1/2^{n+1} > 0$ for each $n \in \{1, 2, \ldots\}$.
But P is not conglomerable in π, as:
$P(E \mid h_n) = P(E \cap h_n)/P(h_n) = 1$, for each $n \in \{1, 2, \ldots\}$, whereas $P(E) = 0.5$. Example 1
In our (1996), we discuss this example in connection with the value of information.

The non-conglomerability of Example 1 extends to a non-trivial IP class, $\mathcal{P}$. Let $\mathcal{P}$ be the set of all finitely additive conditional probabilities whose unconditional probabilities $P(\cdot) = P(\cdot \mid \Omega)$ satisfy conditions (*a*), (*b*) and (*c*) of Example 1. The class $\mathcal{P}$ is convex in the usual sense, applied to unconditional probabilities. That is, assume $\mathcal{P}$ contains two finitely additive conditional probabilities $P_1(\cdot \mid \cdot)$ and $P_2(\cdot \mid \cdot)$ with unconditional probabilities, respectively, $P_1(\cdot)$ and $P_2(\cdot)$. Let $0 \leq x \leq 1$. Then there is a finitely additive conditional probability $P_3(\cdot \mid \cdot)$ in $\mathcal{P}$ whose unconditional probability $P_3(\cdot)$ satisfies $P_3(\cdot) = xP_1(\cdot) + (1-x)P_2(\cdot)$. The cardinality of $\mathcal{P}$ is $2^{|\Re|}$, where $|\Re|$ is the cardinality of the continuum. This follows as $\mathcal{P}$ includes all finitely additive probabilities P where $P(\cdot \mid E^c)$ is a non-principal ultrafilter probability on the positive integers and there are $2^{|\Re|}$-many such non-principal ultrafilters.) Each $P \in \mathcal{P}$ fails conglomerability in π exactly as in Example 1. Hence, with respect to lower and upper unconditional and conditional probabilities, the IP-set $\mathcal{P}$ fails to be conglomerable in the partition π. In Section 5 we give sufficient conditions for an IP set of countably additive probabilities to experience non-conglomerability in an uncountable partition.

2 Non-conglomerable σ-additive probability

The focus of this note is non-conglomerability for countably additive probabilities. In the appendix to our (1986) we show that for a continuous, countably additive probability defined on the continuum, and assuming conditional probabilities that satisfy Definition 1 rather than regular conditional distributions, then non-conglomerability results by considering continuum-many different partitions of the continuum. These alternative partitions are generated by sets of equivalent (non-linearly transformed) random variables. Conglomerability cannot be satisfied in all the partitions.

Here we generalize that result to a large class of countably additive probabilities, P, that are not κ-additive for some uncountable cardinal κ, by identifying for each such P specific partitions where P fails to be conglomerable.

In the following presentation, let α, β, and γ be ordinals and λ and κ cardinals.

Definitions 8:

- A probability P is *κ-additive* if, for each increasing γ-sequence of measurable events, $\{E_\alpha: \alpha < \gamma \leq \kappa\}$, where $E_\alpha \subseteq E_\beta$ whenever $\alpha < \beta < \gamma$, then $P(\cup_{\alpha<\gamma} E_\alpha) = \sup_{\alpha<\gamma} P(E_\alpha)$.

That is, with $\gamma \leq \kappa$, P is κ-additive provided that probability is continuous from below over γ-long sequences that approximate events from below. This definition agrees with the usual definition of countable additivity; let $\kappa = \aleph_0$.

- Say that P is *not κ-additive* when, for some event E and increasing γ-sequence that approximates E from below, $P(\cup_{\alpha<\gamma} E_\alpha) > \sup_{\alpha<\gamma} P(E_\alpha)$.
- If P is κ-additive for each cardinal κ, then call P *perfectly additive*.

Consider a countably additive probability P that is not κ-additive for some cardinal κ. Since the cardinals below a given cardinal form a well-ordered set, we consider the least cardinal κ for which P is not κ-additive. And since we assume that P is countably additive, then κ is some uncountable cardinal – unless P is perfectly additive. Thus, assume that for an uncountable cardinal κ, P is not κ-additive but is λ-additive for each cardinal $\lambda < \kappa$.

We make the following two structural assumptions on the measurable sets $\mathcal{B}$.

- We take the measure completion of P. Each subset of a P-null event is measurable.

That is, if $E \in \mathcal{B}$ with $P(E) = 0$ and $F \subseteq E$ then $F \in \mathcal{B}$.

We require also that $\mathcal{B}$ includes sufficiently many events.

- If E is not P-null and $|E| = \kappa$, then E can be partitioned into two measurable sets of the same cardinality

That is, if $P(E) > 0$ then there exits $E_1, E_2 \in \mathcal{B}$, $E_1 \cap E_2 = \varnothing$, $E_1 \cup E_2 = E$, with $|E_1| = |E_2|$.

Note that, given the first assumption, the second structural assumption can be satisfied in a variety of ways. For example, assume that when E is a κ-sized non-null event, $P(E) > 0$, then there is a κ-sized, null sub-event: There exists $E_1 \subset E$, $|E_1| = \kappa$, and $P(E_1) = 0$.

These two assumptions provide for a rich space of measurable events while stopping short of requiring P to be defined on a power set, which otherwise would require κ to be greater than a weakly inaccessible cardinal, by Ulam's [1930] result for real-valued measurable cardinals.

Here we identify a simple condition involving *tiers* of points that ensures P fails to be conglomerable in a partition of cardinality κ.

Definition 9: A *tier* τ is a (measurable) set of points such that for each pair of points $\{\omega_i, \omega_j\} \subset \tau$ $(i \neq j)$

$$0 < P(\{\omega_i\} \mid \{\omega_i, \omega_j\}) < 1.$$

Proposition: Let P be σ-additive but not κ-additive ($\kappa \geq \aleph_1$), having conditional probabilities defined relative to non-empty sets in $\mathcal{B}$, $P(\cdot \mid \mathcal{B})$, and which satisfies the two structural assumptions on $\mathcal{B}$ identified above. If there is an uncountable tier τ of points, $|\tau| \geq \kappa$ with $P(\tau) > 0$, then P fails to be conglomerable in a partition π with $|\pi| = \kappa$.

Thus, rather than thinking that non-conglomerability is an anomalous feature of finite but not countably additive probabilities, and arises solely with finitely but not countably additive probabilities in countable partitions, here we argue for a different conclusion: Let $P(\cdot \mid \cdot)$ be a conditional probability according to Definition 1. Non-conglomerability of P's conditional probabilities occurs in a partition whose cardinality $|\pi| = \kappa$ matches the κ-non-additivity of P.

We summarize: Let P be defined on a measurable space $<\boldsymbol{\Omega}, \mathcal{B}>$, where $\mathcal{B}$ includes each of the points of the space, $\boldsymbol{\Omega} = \{\omega_\alpha: \alpha < \kappa\}$, with α ranging over all ordinals less than κ. That is, without loss of generality, assume $\boldsymbol{\Omega}$ has cardinality κ and where, if a measurable event E is null, i.e., whenever $P(E) = 0$, then $\mathcal{B}$ includes each subset of E, and where κ-sized non-null events can be split into two measurable κ-sized events. Then if some tier of points is not null, P fails to be conglomerable in a partition of cardinality κ.

Since P is not perfectly additive, it follows that κ is a regular cardinal: it has cofinality κ. Otherwise, κ is singular with cofinality(κ) = $\lambda < \kappa$. Then, using this λ-sequence which is cofinal in κ, as P is λ-additive for each $\lambda < \kappa$, P would be κ-additive as well.

3 Proof of the Proposition

Suppose there exists a tier of points τ, $|\tau| = \kappa$, with $P(\tau) > 0$. Then $P(\{\omega\}) = 0$ for each $\omega \in \tau$, because $P(\tau) > 0$ and P is λ-additive for each cardinal $\lambda < \kappa$. Partition τ into two disjoint sets, $T_0 \cap T_1 = \varnothing$ with $T_0 \cup T_1 = \tau$; each with cardinality κ, $|T_0| = |T_1| = \kappa$; and label them so that $P(T_0) \leq P(T_1) = d > 0$.

We identify a partition of cardinality κ where P fails to be conglomerable, which we write as $\pi = \{h_\alpha: \alpha < \kappa\} \cup \{h'_\beta: \beta < \gamma \leq \kappa\}$, where $\{h_\alpha: \alpha < \kappa\} \cap \{h'_\beta: \beta < \gamma \leq \kappa\} =$

$\varnothing$, and where $P(T_1 \mid h) < d/2$ for each $h \in \pi$. Possibly the second set, $\{h'_\beta: \beta < \gamma \le \kappa\}$, is empty, as we explain below. Each element $h \in \pi$ is a finite set. Each element h_α contains exactly one point from T_1, and some positive finite number of points from T_0, selected to insure that $P(T_1 \mid h) < d/2$. If the second set, $\{h'_\beta: \beta < \gamma \le \kappa\}$, is not empty, each $h'_\beta = \{\omega_\beta\}$ is a singleton with $\omega_\beta \in \Omega - T_1$. So, if $\{h'_\beta: \beta < \gamma \le \kappa\}$ is not empty, then $P(T_1 \mid h'_\beta) = 0$ for each h'_β. Next we establish the existence of such a measurable partition π.

By the Axiom of Choice, consider a κ-long well ordering of T_1, $\{\omega_1, \omega_2, \ldots, \omega_\beta, \ldots\}$ with ordinal indices $0 < \beta < \kappa$. Define π by induction. As each of T_0, T_1 is a subset of the tier τ, consider the countable partition of T_0 into sets
$\rho_{1,n} = \{\omega \in T_0: (n-1)/n \le P(\{\omega_1\} \mid \{\omega_1, \omega\}) < n/(n+1)\}$
for n = 1, 2 … .

Observe that $\cup_n \rho_{1,n} = T_0$. Since $|T_0| = \kappa \ge \aleph_1$, by the pigeon-hole principle, consider the least n* such that ρ_{1,n^*} is infinite. Let $U_1 = \{\omega_{1,1}, \ldots, \omega_{1,m}\}$ be m-many points chosen from ρ_{1,n^*}. Note that $P(\{\omega_1\} \mid U_1 \cup \{\omega_1\}) \le n^*/(m+n^*)$. Choose m sufficiently large so that $n^*/(m+n^*) < d/2$. Let $h_1 = U_1 \cup \{\omega_1\}$.

For ordinals $1 < \beta < \kappa$, define h_β, by induction, as follows. Denoting $T_{0,1} = T_0$, and let $T_{0,\beta} = T_0 - (\cup_{0<\alpha<\beta} h_\alpha)$. Since, for each α, $0 < \alpha < \beta$, by hypothesis of induction h_α is a finite set, then $|\cup_{0<\alpha<\beta} h_\alpha| < \kappa$. So, $|T_{0,\beta}| = \kappa$. Since $T_{0,\beta}$ is a subset of τ, just as above, consider the countable partition of $T_{0,\beta}$ into sets
$\rho_{\beta,n} = \{w \in T_{0,\beta}: (n-1)/n \le P(\{\omega_\beta\} \mid \{\omega_\beta, \omega\}) < n/(n+1)\}$ for n = 1, 2, … . Again, by the pigeon-hole principle, consider the least integer n* such that ρ_{β,n^*} is infinite. Let $U_\beta = \{\omega_{\beta,1}, \ldots, \omega_{\beta,m}\}$ be m-many points chosen from ρ_{β,n^*}. Note that

$$P(\{\omega_\beta\} \mid U_\beta \cup \{\omega_\beta\}) \le n^*/(m+n^*).$$

Choose m sufficiently large that $n^*/(m+n^*) < d/2$. Let $h_\beta = U_\beta \cup \{\omega_\beta\}$. Observe that $T_1 \subset \cup_{0<\beta<\kappa} h_\beta$ and that for each $0 < \beta < \kappa$, $P(T_1 \mid h_\beta) < d/2$. In order to complete the partition π, consider a catch-all set with all the remaining points $\omega_\beta \in \Omega - \cup_{0<\beta<\kappa} h_\beta$. Note that each such ω_β is not a member of T_1, if any such points exist. Add each such point $\{\omega_\beta\} = h'_\beta$ as a separate partition element of π. Thus, if there are any such points, $P(T_1 \mid h'_\beta) = 0 < d/2$.

Hence, P is not conglomerable in π as $P(T_1) = d > 0$, yet for each $h \in \pi$, $P(T_1 \mid h) < d/2$.◊ Proposition

4 An Example of the Proposition

Next, we illustrate the Proposition and with it also the difference (2) between the theory of conditional probability according to Definition 1 and the theory of regular conditional distributions.

Example 2: Let $<\Omega, \mathcal{B}>$ be the measurable space of Lebesgue measurable subsets of the half-open unit interval of real numbers: $\Omega = [0,1)$ and $\mathcal{B}$ is its algebra of Lebesgue measurable subsets. Let P be the uniform, countably additive probability with constant density function $f(\omega) = 1$ for each real number $0 \le \omega < 1$, and $f(\omega) = 0$ otherwise. So $P(\{\omega\}) = 0$ for each $\omega \in \Omega$. Evidently P is not κ-additive, because $\kappa = |\Omega| = |\Re|$.

Consider the uniform density function f to identify conditional probability given finite sets as uniform over those finite sets, as well. That is, when $F = \{\omega_1, \ldots, \omega_k\}$ is a finite subset of Ω with k-many points, let $P(\cdot \mid F)$ be the perfectly additive probability that is uniform on these k-many points. These conditional probabilities create a single tier, $\tau = \Omega$, because $P(\{\omega_1\} \mid \{\omega_1, \omega_2\}) = 0.5$ for each pair of points in Ω.

Consider the two events $E = \{\omega: 0 \le \omega < 0.9\}$ and its complement with respect to Ω, $E^c = \{\omega: 0.9 \le \omega < 1\}$, where $P(E) = 0.9$. Let g be the 1-1 (continuous) map between E and E^c defined by $g(\omega) = 0.9 + \omega/9$, for $\omega \in E$. Consider the κ-size partition of Ω by pair-sets, $\pi = \{\{\omega, g(\omega)\}: \omega \in E\}$. By assumption, $P(\{\omega\} \mid \{\omega, g(\omega)\}) = 1/2$ for each pair $\{\omega, g(\omega)\} \in \pi$. But then P is not conglomerable in π.◊Example 2

The theory of regular conditional distributions treats the example differently. We continue Example 2 from that point of view.

Example 2 (continued) Consider the measure space $<\Omega, \mathcal{B}, P>$ as above. Let the random variable $X(\omega) = \omega$, so that $X \sim U[0,1)$, X has the uniform distribution on Ω. In order to consider conditional probability given the pair of points $\{\omega, g(\omega)\}$, let

$$g(X) = (X/9) + 0.9 \qquad \text{if } 0 \le X < 0.9$$

$g(X) = 9(X - 0.9)$ if $0.9 \le X < 1$.

Define the random variable

$Y(\omega) = X(\omega) + g(X(\omega)) - 0.9$.

Observe that $Y \sim U[0, 1.0)$. Also, note that Y is 2-to-1 between Ω and [0.0, 1.0). That is $Y = y$ entails that either $\omega = 0.9y$ or $\omega = 0.1(y + 9)$.

Let the sub-sigma field $\mathcal{A}$ be generated by the random variable Y. The regular conditional distribution relative to this sub-sigma field, $P(\mathcal{B} \mid \mathcal{A})(\omega)$, is a real-valued function defined on Ω that is $\mathcal{A}$-measurable and satisfies the integral equation

$$\int_A P(B \mid \mathcal{A})(w)\, dP(\omega) = P(A \cap B)$$

whenever $A \in \mathcal{A}$ and $B \in \mathcal{B}$.

In our case, then $P[B \mid \mathcal{A}](\omega)$ almost surely satisfies:

$P(X = 0.9Y \mid Y)(\omega) = 0.9$

and $P(X = 0.1(Y + 9.0) \mid Y)(\omega) = 0.1$.

Thus, relative to the random variable Y, this regular conditional distribution assigns conditional probabilities as if $P(\{\omega\} \mid \{\omega, g(\omega)\}) = 0.9$ for almost all pairs $\{\omega, g(\omega)\}$ with $0 \le \omega < 0.9$. However, just as in the Borel "paradox" (Kolmogorov, 1956), for a particular pair $\{\omega, g(\omega)\}$, the evaluation of $P(\{\omega\} \mid \{\omega, g(\omega)\})$ is not determinate and is defined only relative to which sub-sigma field $\mathcal{A}$ embeds it.

For an illustration of this last feature of the received theory of regular conditional distributions, consider a different pair of complementary events with respect to Ω. Let $F = \{\omega: 0 \le \omega < 0.5\}$ and $F^c = \{\omega: 0.5 \le \omega < 1\}$. So, $P(F) = 0.5$.

Let $f(X) = 1.0 - X$ if $0 < X < 1$.
$= 0$ if $X = 0$.

Analogous to the construction above, let

$$Z(\omega) = |X(\omega) - f(X(\omega))|.$$

So Z is uniformly distributed, $Z \sim U[0, 1)$, and is 2-to-1 from Ω onto [0, 1). Consider the sub-sigma field $\mathcal{A}'$ generated by the random variable Z. Then the regular conditional distribution $P(\mathcal{B} \mid \mathcal{A}')(\omega)$, almost surely satisfies:

$P(X = 0.5 - Z/2 \mid Z \ne 0)(\omega) = 0.5$

and $P(X = 0.5 + Z/2 \mid Z \ne 0)(\omega) = 0.5$

and for convenience,

$P(X = 0 \mid Z = 0) = P(X = 0.5 \mid Z = 0) = 0.5$.

However, $g(.09) = .91 = f(.09)$ and $g(.91) = .09 = f(.91)$. That is, $Y = 0.1$ if and only if $Z = 0.82$. So in the received theory, it is permissible to have

$P(\omega = .09 \mid Y = 0.1) = 0.9$

as evaluated with respect to the sub-sigma field generated by Y, and also to have

$P(\omega = .09 \mid Z = 0.82\}) = 0.5$

as evaluated with respect to the sub-sigma field generated by Z, even though the conditioning events are the same event.◊ Example 2 (continued)

5 Non-conglomerability for an IP Bounded Density Ratio model

Our focus in this note is on non-conglomerability for a single, σ-additive but non-κ-additive probability P that has conditional probabilities according to Definition 1, and where some non-null tier τ (i.e., $P(\tau) > 0$) is composed of null points from Ω. We highlight this case as we think it typifies how conditional probabilities given finite set of points are associated with familiar continuous statistical models. Thus, we have demonstrated non-conglomerability in a particular partition for what we judge is the usual interpretation of conditional probabilities from a single continuous, countably additive probability distribution.

The *Proposition* applies to each element of an IP model, when that model uses conditional probabilities from a countably additive, continuous probability that satisfy Definition 1. This puts pressure, we think on those who (e.g., Walley, 1991) appear to require conglomerability in arbitrary partitions as a condition for coherent IP degrees of belief. Here is a Corollary to the Proposition illustrating the point.

Let $\mathcal{P}$ be a set of countably additive, but not κ-additive probabilities. Assume each $P \in \mathcal{P}$ is defined on a common measurable space $\{\Omega, \mathcal{B}\}$, where the points of Ω are the atoms of $\mathcal{B}$, and where each P has conditional probabilities $P(\cdot \mid \cdot)$ satisfying Defintion 1. Assume that $\mathcal{P}$ satisfies the following *Bounded Density Ratio* [BDR] condition, which is a weakened variant of DeRobertis and Hartigan's (1981) *Density Ratio* model:

- **BDR** There exist a set $T \subseteq \Omega$ where,

(1) T can be partitioned into two sets T_0, T_1 with

$|T_0| = |T_1| = \kappa$ and $Inf_{P \in \mathcal{P}}[P(T_1)] = d > 0$.

(2) For each pair, $\omega_\alpha \ne \omega_\beta \in T$,

$Sup_{P \in \mathcal{P}} [\, P(\{\omega_\alpha\} \mid \{\omega_\alpha, \omega_\beta\})\,] < 1$

Note that the BDR condition requires only that the probability distributions that belong to $\mathcal{P}$ have bounded relative densities with respect to pairs of atoms from $\mathcal{B}$. As a consequence of the BDR condition, with respect to each $P \in \mathcal{P}$, the distinguished P-non-null set T belongs to one P-non-null tier.

*Corollary***:** When $\mathcal{P}$ is an IP BDR model, then $\mathcal{P}$ fails to be conglomerable. Specifically, there exits a κ-sized partition by finite sets, $\pi = \{h_\alpha: |h_\alpha| < \aleph_0, \alpha < \kappa\}$ where

$Sup_{\, h \in \pi, P \in \mathcal{P}}[\, P(T_1 \mid h)\,] \;<\; d = Inf_{\, P \in \mathcal{P}}[\, P(T_1)\,]$.

*Proof***:** The proof of the Corollary parallels the proof of the Proposition, with one change. That difference is in the sets $\rho_{\beta,n}$. For the Corollary, denoting these by $\rho'_{\beta,n}$, we define them inductively as follows.

Let $\rho'_{1,n} = \{\omega \in T_0$:

$(n\text{-}1)/n \;\leq\; Sup_{\, P \in \mathcal{P}}[P(\{\omega_1\} \mid \{\omega_1, \omega\}) \;<\; n/(n+1)\}$

for n = 1, 2, … . By BDR(2), the sets $\{\rho'_{1,n}: n = 1, 2, \ldots\}$ partition T_0.

Consider the least n* such that ρ'_{1,n^*} is infinite. Let $U_1 = \{\omega_{1,1}, \ldots, \omega_{1,m}\}$ be m-many points chosen from ρ'_{1,n^*}. Note that for each $P \in \wp$ $P(\{\omega_1\} \mid U_1 \cup \{\omega_1\}) \leq n^*/(m+n^*)$. Choose m sufficiently large so that $n^*/(m+n^*) < d/2$. Let $h_1 = U_1 \cup \{\omega_1\}$. So,

$$Sup_{\, P \in \mathcal{P}}[\, P(T_1 \mid h_1)\,] \;\leq\; d/2$$

Define h_β, by induction, just as in the proof of the Proposition. For $\beta < \kappa$, define $T_{0,\beta} = T_0 - (\cup_{0<\alpha<\beta} h_\alpha)$.

Consider the countable partition of the set $T_{0,\beta}$ into sets

$\rho'_{\beta,n} = \{\omega \in T_{0,\beta}$:

$P(n\text{-}1)/n \;\leq\; Sup_{\, P \in \mathcal{P}}[\, P(\{\omega_\beta\} \mid \{\omega_\beta, \omega\})\,] \;<\; n/(n+1)\}$

for n = 1, 2, … . The proof of the Corollary then follows the proof of the Proposition, resulting in the required partition π. $\lozenge_{\text{Corollary}}$

6 Concluding Remarks

In a different paper (2012), we investigate the question of non-conglomerability for a single countably additive but κ-non-additive probability where no set of P-null points forms a P-non-null tier. Though the mathematics for analyzing this case is rather different from the reasoning used in the Proposition presented here, we point the reader to some interesting features about tiers that we use to address this other case.

Definition 10: Consider the relation, ~, of relative-non-nullity on pairs of points in Ω. That is, for two different points, $\omega_1 \neq \omega_2$ they bear the relation $\omega_1 \sim \omega_2$ provided that

$0 < P(\{\omega_1\} \mid \{\omega_1, \omega_2\}) < 1$.

We make ~ into an equivalence relation by stipulating that, for each point ω, $\omega \sim \omega$.

Next we state and prove an elementary fact.

Fact: ~ is an equivalence relation.

Proof: Only transitivity requires verification. Assume $\omega_1 \sim \omega_2 \sim \omega_3$. That is, assume

$0 < P(\{\omega_1\} \mid \{\omega_1, \omega_2\}), P(\{\omega_2\} \mid \{\omega_2, \omega_3\}) < 1$.

Then by (iii) of Definition 1 for conditional probability:

$P(\{\omega_1\} \mid \{\omega_1, \omega_2, \omega_3\}) =$

$\quad P(\{\omega_1\} \mid \{\omega_1, \omega_2\})\, P(\{\omega_1, \omega_2\} \mid \{\omega_1, \omega_2, \omega_3\})$.

Also, $P(\{\omega_3\} \mid \{\omega_1, \omega_2, \omega_3\}) =$

$\quad P(\{\omega_3\} \mid \{\omega_2, \omega_3\})\, P(\{\omega_2, \omega_3\} \mid \{\omega_1, \omega_2, \omega_3\})$.

Now argue indirectly by cases.

If $P(\{\omega_1\} \mid \{\omega_1, \omega_3\}) = 0$,
then $P(\{\omega_1\} \mid \{\omega_1, \omega_2, \omega_3\}) = 0$
and $P(\{\omega_1, \omega_2\} \mid \{\omega_1, \omega_2, \omega_3\}) = 0$,
since, by assumption. $P(\{\omega_1\} \mid \{\omega_1, \omega_2\}) > 0$.
Then $P(\{\omega_2\} \mid \{\omega_1, \omega_2, \omega_3\}) = 0 = P(\{\omega_2\} \mid \{\omega_2, \omega_3\})$,
which contradicts $\omega_2 \sim \omega_3$.
If $P(\{\omega_1\} \mid \{\omega_1, \omega_3\}) = 1$,
then $0 = P(\{\omega_3\} \mid \{\omega_1, \omega_3\}) = P(\{\omega_3\} \mid \{\omega_1, \omega_2, \omega_3\})$.
Then $0 = P(\{\omega_2, \omega_3\} \mid \{\omega_1, \omega_2, \omega_3\})$,
since $0 < P(\{\omega_3\} \mid \{\omega_2, \omega_3\})$.
So, $0 = P(\{\omega_2\} \mid \{\omega_1, \omega_2, \omega_3\}) = P(\{\omega_2\} \mid \{\omega_1, \omega_2\})$,
which contradicts $\omega_1 \sim \omega_2$.

Hence $0 < P(\{\omega_1\} \mid \{\omega_1, \omega_3\}) < 1$, as required. $\lozenge_{\text{Fact}}$

Thus, the equivalence relation ~ partitions Ω into disjoint *tiers* τ of relative non-null pairs of points. For each pair of points $\{\omega_1, \omega_2\}$ that belong to different tiers, $\omega_i \in \tau_i$ (i = 1, 2), when $\tau_1 \neq \tau_2$, then $P(\{\omega_1\} \mid \{\omega_1, \omega_2\}) \in \{0,1\}$. If $P(\{\omega_2\} \mid \{\omega_1, \omega_2\}) = P(\{\omega_3\} \mid \{\omega_2, \omega_3\}) = 1$, then $P(\{\omega_3\} \mid \{\omega_1, \omega_3\}) = 1$. Thus, the tiers are linearly ordered by the relations ↑, ↓ defined as:

Definitions 11:

- $\tau_1 \uparrow \tau_2$ if for each pair $\{\omega_1, \omega_2\}$, $\omega_i \in \tau_i$ (i = 1, 2), $P(\{\omega_2\} \mid \{\omega_1, \omega_2\}) = 1$.

The reverse ordering also is linear. We express this as

- $\tau_2 \downarrow \tau_1$ if for each pair $\{\omega_1, \omega_2\}$, $\omega_i \in \tau_i$ (i = 1, 2), $P(\{\omega_2\} \mid \{\omega_1, \omega_2\}) = 1$.

That is, $\tau_2 \downarrow \tau_1$ if and only if $\tau_1 \uparrow \tau_2$.

Next, consider the possibly empty set of P-non-null points. Let $\tau^* = \{\omega: P(\omega) > 0)$. Evidently, when $\varnothing \neq \tau^* \neq \tau$, then $\tau^* \downarrow \tau$, and τ* is the top element in the linear order of tiers.

We note that this linear order of tiers plays an important role in Dubins (1975) proof of the existence of fully defined finitely additive conditional probabilities, i.e., where $\mathcal{B}$ is the powerset of Ω and P(B | A) is well-defined whenever $\varnothing \neq$A, B are elements of $\mathcal{B}$. Also, it appears in both Levi's (1980, §5.5) and Regazzini's (1985) strengthened version of de Finetti's criterion of coherence for conditional previsions. Levi and Regazzini strengthen de Finetti's coherence criterion for a called off gamble given a null event in order to have coherent conditional previsions that satisfy Definition 1.

Under additional structural assumptions about $\mathcal{B}$, including measurability of the intervals of tiers formed under $\downarrow$, in our (2012) we extend the Proposition to include non-conglomerability for such cases as well. This permits us to conclude that the anomalous phenomenon of non-conglomerability is a result of adopting the de Finetti/Dubins theory of conditional probability instead of the rival Kolmogorovian theory of regular conditional distributions. Non-conglomerability is not a result primarily of the associated debate over whether probability is allowed to be merely finitely additive rather than satisfying countable additivity.

Restated, our conclusion is the observation that (subject to structural assumptions on the algebra $\mathcal{B}$) even when P is λ-additive for each $\lambda < \kappa$, if P is not κ-additive and has conditional probabilities that satisfy Definition 1, then P will experience non-conglomerability in a κ-sized partition. And then such conditional probabilities will not satisfy condition (3) of the theory of regular conditional distributions.

On the other hand, regular conditional distributions avoid non-conglomerability by allowing conditional probability to depend upon a sub-sigma field, rather than being defined given an event. And, occasionally, they avoid non-conglomerability by abandoning the requirement of *Propriety*, which is clause (ii) of Definition 1 of the de Finetti/Dubins theory of conditional probabilities.

Evidently, some countably additive continuous IP models that use the theory of conditional probabilities associated with Definition 1 require non-conglomerability in specific, uncountable partitions. We think this is a better alternative than using IP models with conditional probabilities based on the theory of regular conditional distributions. In future work on IP models with conditional probabilities, we hope to address the following question:

- With respect to a given IP model that use conditional probabilities, in the sense of Definition 1, in which partitions is non-conglomerability mandated?

Acknowledgments

In writing this paper we have benefitted by helpful comments from Jeremy Avigad, Jessi Cisewski, Paul Pedersen, Rafael Stern, Wilfried Sieg and three anonymous referees.

References

Berti, P. and Rigo, P. (2007) “0-1 Laws for Regular Conditional Distributions,” *Ann. Prob*. **35**: 649-662.

Billingsley, P. (1986) *Probability and Measure*, 2nd ed. N.Y.: John Wiley.

Blackwell, D. (1955) "On a class of probability spaces," *Proc. Third Berkeley Symp. Math. Statist. Prob*., pp 1-6. University of California Press.

Blackwell, D. and Ryll-Nardzewski, C. (1963) "Non-existence of everywhere proper conditional distributions," *Annals of Mathematical Statistics* **34**, 223-225.

Blackwell, D. and Dubins, L.E. (1975), "On existence and non-existence of proper, regular, conditional distributions," *Annals of Probability* **3**, 741-752.

Breiman, L. (1968) *Probability*. Mass.: Addison-Wesley.

deFinetti, B. (1974) *Theory of Probability*. N.Y: Wiley.

DeRobertis, L, and Hartigan, J. (1981) “Bayesian inference using intervals of measures,” Ann. Stat. **9**: 235-244.

Dubins, L.E. (1975) "Finitely Additive conditional probabilities, conglomerability, and disintegrations," *Ann. Prob*. **3**, 89-99.

Doob, J.L.(1953) *Stochastic Processes*. N.Y.: Wiley.

Halmos, P. (1950) *Measure Theory*. NY: van Nostrand.

Hewitt, E. and Savage, L.J. (1955) "Symmetric measures on Cartesian products," *A.M.S. Trans.* **80**, 470-501.

Kadane, J.B., Schervish, M.J., and Seidenfeld, T. (1986) "Statistical implications of finitely additive probability," in *Bayesian Inference and Decision Techniques with Applications*. Eds. P.K Goel and A.Zellner, New York: Elsevier, pp. 59-76.

Kadane, J.B., Schervish, M.J., and Seidenfeld, T. (1996) Reasoning to a Foregone Conclusion. *J. American Statistical Association* **91**: 1228-1235.

Kolmogorov, A.N. (1956) *Foundations of the Theory of Probability*. New York: Chelsea.

Levi, I. (1980) *The Enterprise of Knowledge*. Cambridge, MA: MIT Press.

Loeve, M. (1955) *Probability Theory*. NY: van Nostrand.

Regazzini, E. “Finitely additive conditional probabilities,” *Milan J. Mathematics*, **55** (1):69–89, 1985.

Schervish, M.J., Seidenfeld, T, and Kadane, J.B. (1984), "The extent of non-conglomerability of finitely additive probability," *Z.War*. **66**, 205-226.

Seidenfeld, T., Schervish, M.J. and Kadane, J.B. (2001) “Improper Regular Conditional Distributions,” *Annals of Probability* **29**:1612-1624.

Seidenfeld, T., Schervish, M.J., and Kadane, J.B. (2012) “Non-conglomerability for countably additive measures that are not κ-additive,” Technical Report available at http://www.hss.cmu.edu/philosophy/faculty-seidenfeld.php

Ulam, S., (1930) Zur Masstheorie in der allgemeinen Mengenlehre, *Fund. Math*. **16**: 140-150.

Walley, P. (1991) *Statistical Reasoning with Imprecise Probabilities*. London: Chapman and Hall.