October 4, 2000

IMPROPER REGULAR CONDITIONAL DISTRIBUTIONS[1]

By Teddy Seidenfeld, Mark J. Schervish and Joseph B. Kadane
Carnegie Mellon University

Improper regular conditional distributions (rcd's) given a σ-field $\mathcal{A}$ have the following anomalous property. For sets $A \in \mathcal{A}$, $\Pr(A \mid \mathcal{A})$ is not always equal to the indicator of A. Such a property makes the conditional probability puzzling as a representation of uncertainty. When rcd's exist and the σ-field $\mathcal{A}$ is countably generated, then almost surely the rcd is proper. We give sufficient conditions for an rcd to be improper in a maximal sense, and show that these conditions apply to the tail σ-field and the σ-field of symmetric events.

1. Introduction. The theory of regular conditional distributions (rcd's) is a standard part of the received view of mathematical probability. Nonetheless, there are some anomalous cases of conditional probability distributions where, in the terminology of Blackwell, Dubins, and Ryll-Nardzewski, the rcd is not everywhere *proper*, given the conditioning sub-σ-field, $\mathcal{A}$. That is, let $P(\cdot \mid \mathcal{A})(\omega)$ denote the rcd for the measure space $(\Omega, \mathcal{B}, P)$ given the conditioning sub-σ-field, $\mathcal{A}$. That the rcd is *proper* at ω means that whenever $\omega \in A \in \mathcal{A}$, $P(A \mid \mathcal{A})(\omega) = 1$. The rcd is *improper* if it is not everywhere proper. Here, we explore the extent of such *impropriety*, focusing on atomic sub-σ-fields, $\mathcal{A}$, with atoms $a(\omega)$, where the impropriety of the rcd is maximal in two senses, local and global, at once. The failure of propriety at the point ω is locally maximal as $P(a(\omega) \mid \mathcal{A})(\omega) = 0$. The failure of propriety is globally maximal as the rcd is improper at P-almost all points. Also, we consider a connection between the impropriety of rcd's for symmetric measures,

[1]Research partially supported by NSF grant DMS 9801401.
AMS 1991 *subject classifications.* 60A10.
Key words and phrases. Completion of σ-field, countably generated σ-field, nonmeasurable set, symmetric σ-field, tail σ-field.

given the sub-σ-field of symmetric events, and Vitali-styled nonmeasurable sets. This connection leads us to a conjecture about the possibility of using certain finitely additive extensions of P as a way around the impropriety of the countably additive rcd in these cases.

2. Regular conditional distributions. Let $(\Omega, \mathcal{B}, P)$ be a measure space. Denote by ω points in Ω. In what follows all probability distributions are countably additive unless otherwise stated.

When $P(A) > 0,\ A \in \mathcal{B}$, then the conditional probability distribution over $\mathcal{B}$ given A is well defined by $P(\cdot \mid A) = P(\cdot \cap A)/P(A)$. Of course, this approach does not provide guidance for conditional probability given P-null (measure 0) events. For that, the received view comes from Kolmogorov's seminal 1933 work (1950).

DEFINITION 1. In the usual terminology, with $\mathcal{A}$ a sub-σ-field of $\mathcal{B}$, $P(\cdot \mid \mathcal{A})$ is a *regular conditional distribution* [rcd] on $\mathcal{B}$, given $\mathcal{A}$ provided that:

1. For each $\omega \in \Omega$, $P(\cdot \mid \mathcal{A})(\omega)$ is a probability on $\mathcal{B}$.
2. For each $B \in \mathcal{B},\ P(B \mid \mathcal{A})(\cdot)$ is an $\mathcal{A}$-measurable function.
3. For each $A \in \mathcal{A},\ B \in \mathcal{B}\ \int_A P(B \mid \mathcal{A})(\omega)dP(\omega) = P(A \cap B)$. That is, $P(B \mid \mathcal{A})$ is a version of the Radon-Nikodym derivative of $P(\cdot \cap B)$ with respect to P.

DEFINITION 2. An $\mathcal{A}$-*atom* is the intersection of all the elements of $\mathcal{A}$ that contain a given point ω of Ω.

Thus, condition 2 for rcd's requires that $P(B \mid \mathcal{A})(\cdot)$ is constant on the $\mathcal{A}$-atoms.

Two limitations in this approach are well documented in the literature.

2.1. *The "Borel paradox".* One controversial aspect of this theory of conditional probability was pointed out by Kolmogorov (1950, pp. 50–51). He calls it the "Borel paradox." (See, e.g., Billingsley, 1995, p. 441, problem 33.1.) Put simply, the Borel paradox shows that $P(\cdot \mid \mathcal{A})(\omega)$ is *not* a probability distribution on $\mathcal{B}$ given *events* in $\mathcal{A}$ but, rather, it is a probability distribution given a σ-field. Specifically, with $\mathcal{B}_Z$ a σ-field generated by the random variable Z, let sub-σ-fields $\mathcal{A}_X$ and $\mathcal{A}_Y$ be generated by the random variables X and Y, respectively. Suppose that $X = x^*$ is the same event (in $\mathcal{B}$) as $Y = y^*$. Nonetheless, if $X(\omega) = x^*$, $P(\cdot \mid \mathcal{A}_X)(\omega)$ and $P(\cdot \mid \mathcal{A}_Y)(\omega)$ may be different distributions, with sup norm distance arbitrarily close to 1. In rebuttal to this objection, Kolmogorov points out that between any two conditioning sub-σ-fields, this "paradox" can occur only on a P-null set of points. That is, it is a measure-0 failure, at worst. However, if sufficiently many sub-σ-fields are considered simultaneously, as might arise through a family of continuous transformations of a bivariate conditioning sub-σ-field, the Borel paradox may become a problem of full measure. (See the appendix to Kadane, Schervish and Seidenfeld, 1986.)

2.2. *Rcd's may not exist.* The canonical example of a measure space and conditioning sub-σ-field that admits no rcd is obtained by letting $\mathcal{B}$ be an extension of the Borel sets on [0,1] under Lebesgue measure with the addition of one non-measurable set, and letting $\mathcal{A}$ be the sub-σ-field of Borel sets themselves. (See, e.g., Halmos, 1950, p. 211.) The same example is duplicated with only minor variations in Billingsley (1995, Exercise 33.13, p. 443); Breiman (1968, p. 81); Doob (1953, p. 624); and Loeve (1955, p. 370 #1). Though, for each $B \in \mathcal{B}$, the extended measure space has Radon-Nikodym derivatives $P(B \mid \mathcal{A})$ satisfying condition 3, above, the derivatives resist assembly of these pointwise probabilities into a full probability distribution on $\mathcal{B}$, measurable with respect to $\mathcal{A}$, as required by conditions 1 and 2. In the counterexample, exceptional null sets pile up to create a failure. That these texts use a common couterexample involving a non-measurable set to preclude existence of an rcd is not accidental, as Corollary 1 establishes.

Definition 3. For an event A, let $I_A(\omega) = 1$ if $\omega \in A$ and $I_A(\omega) = 0$ if $\omega \in A^{\mathbf{c}}$, i.e. $I_A(\cdot)$ is the indicator for A.

Definition 4. Sub-σ-field $\mathcal{A}$ is *atomic* if it contains each of its $\mathcal{A}$-atoms.

Theorem 1. *Let $\mathcal{A}$ be a countably generated sub-σ-field of $\mathcal{B}$. Let $P(\cdot \mid \mathcal{A})$ be a regular conditional distribution on $\mathcal{B}$, given $\mathcal{A}$. Then, there exists a set $C^* \in \mathcal{A}$, with $P(C^*) = 1$ such that for each $A \in \mathcal{A}$ and $\omega \in C^*$, $P(A \mid \mathcal{A})(\omega) = \mathbf{I}_A(\omega)$.*

The proof of Theorem 1 is established with the aid of Lemma 1.

LEMMA 1. (Billingsley, 1995, p. 431, Example 33.3) *Assume that* $P(\cdot \mid \mathcal{A})$ *is a regular conditional distribution on* $\mathcal{B}$*, given* $\mathcal{A}$*. Let* $A \in \mathcal{A}$*. Then there exists a set* $C \in \mathcal{A}$ *with* $P(C) = 1$ *such that for each* $\omega \in C$*,* $P(A \mid \mathcal{A})(\omega) = \mathbf{I}_A(\omega)$*.*

PROOF OF THEOREM 1. Apply Lemma 1 to each element $\{A_n : n = 1, \ldots\}$ of a countable set of generators for $\mathcal{A}$. Let $\{C_n : n = 1, \ldots\}$ be the resulting sequence of almost sure events. Define set $C^* = \cap\{C_n\}$. Then C^* satisfies the conclusion to the theorem, as it does so for each generator $A_n (n = 1, \ldots)$. □

COROLLARY 1. *Let* $\mathcal{A}$ *be an atomic, countably generated sub-*σ*-field of* $\mathcal{B}$*, where the* $\mathcal{A}$*-atoms are the singletons. Let* $P(\cdot \mid \mathcal{A})$ *be a regular conditional distribution on* $\mathcal{B}$*, given* $\mathcal{A}$*. Then* $\mathcal{B}$ *is a sub-*σ*-field of the measure completion of* P *on* $\mathcal{A}$*.*

PROOF. This results from Theorem 1 (see also Loeve, 1955, p. 356), as follows: Let $C^* \in \mathcal{A}$ be the P-measure 1 set guaranteed to exist by the Theorem 1. Then, as each singleton $\{\omega\}$ is an element of $\mathcal{A}$ by assumption, for each $\omega \in C^*$, $P(\{\omega\} \mid \mathcal{A})(\omega) = 1$. Let $E \in \mathcal{B}$. Then, for $\omega \in C^* \cap E$, $P(E \mid \mathcal{A})(\omega) = 1$. For $\omega \in C^* \cap E^{\mathbf{c}}, P(E^{\mathbf{c}} \mid \mathcal{A})(\omega) = 1$ and thus $P(E \mid \mathcal{A})(\omega) = 0$. Hence, $P(E \mid \mathcal{A})(\omega) = \mathbf{I}_E(\omega)$, almost surely with respect to P. But since $\{\omega : P(E \mid \mathcal{A})(\omega) = 1\}$ is $\mathcal{A}$-measurable and likewise for $\{\omega : P(E \mid \mathcal{A})(\omega) = 0\}$, the set E differs from some set in $\mathcal{A}$ by a P-null event. That is, E must be in the measure completion of $\mathcal{A}$. □

There is a familiar and helpful sufficient condition for existence of an rcd on $\mathcal{B}$ given each of its sub-σ-fields $\mathcal{A}$. That is, that $\mathcal{B}$ is isomorphic (under a 1-1 measurable mapping) to the σ-field of a random variable. See, e.g., Billingsley (1995, Theorem 33.3, p. 439); or, Breiman (1968, Theorem 4.30, p.78). If this condition holds, we shall call $(\Omega, \mathcal{B})$ a *Borel space.* If $(\Omega, \mathcal{B})$ is a Borel space, then $\mathcal{B}$ is countably generated. When $\mathcal{B}$ is countably generated, regardless whether $(\Omega, \mathcal{B})$ is a Borel space, if an rcd exists given a sub-σ-field $\mathcal{A}$, it is almost surely unique.

LEMMA 2. *Let* $P_i(\cdot \mid \mathcal{A})(\omega)(i = 1, 2)$ *be two rcd's for* P *on* $\mathcal{B}$ *given* $\mathcal{A}$*, and assume that* $\mathcal{B}$ *is countably generated by a set that forms a* π*-system, i.e., the countably many generators are closed under finite intersections. (Alternatively, let* $\mathcal{B}$ *be separable* σ*-field, i.e., one with a countable dense set.) Then,* $P\{\omega : P_1(\cdot \mid \mathcal{A})(\omega) = P_2(\cdot \mid \mathcal{A})(\omega)\} = 1$.

PROOF. Let $B_i(1 = 1, \ldots)$ be a π-system (or countable dense set) for $\mathcal{B}$. Let

$$\begin{aligned}
W_{1i} &= \{\omega : P_1(B_i \mid \mathcal{A})(\omega) > P_2(B_i \mid \mathcal{A})(\omega)\}, \\
W_{2i} &= \{\omega : P_1(B_i \mid \mathcal{A})(\omega) < P_2(B_i \mid \mathcal{A})(\omega)\}, \\
W_{3i} &= \{\omega : P_1(B_i \mid \mathcal{A})(\omega) = P_2(B_i \mid \mathcal{A})(\omega)\}.
\end{aligned}$$

Each of these is an $\mathcal{A}$-measurable set as $P_j(B_i \mid \mathcal{A})(\cdot)$ is an $\mathcal{A}$-measurable function, for each $i = 1, 2, \ldots$ and $j = 1, 2$. It is sufficient to show that $P(W_{3i}) = 1$ for all i. If, to the contrary, for some i $P(W_{3i}) < 1$, argue for a contraction as follows. Suppose then that $P(W_{1i}) > 0$. Then,

$$\begin{aligned}
P(B_i \cap W_{1i}) &= \int_{W_{1i}} P_1(B_i \mid A)(\omega) dP(\omega) > \int_{W_{li}} P_2(B_i \mid \mathcal{A})(\omega) dP(\omega) \\
&= P(B_i \cap W_{1i}),
\end{aligned}$$

which is a contradiction. □

When the sufficient condition for existence of rcd's fails because the measure space is not countably generated, rcd's may nonetheless exist though they can form mutually singular families of distribution when evaluated at each point ω.

EXAMPLE 1. Let $\mathcal{B} = \mathcal{A}$ be the σ-field of all countable and co-countable sets in [0,1]. Let P be a probability that assigns 0 to each point (real number) in [0,1]. Each of the following is readily seen to be an rcd for P on $\mathcal{B}$, given $\mathcal{A}$.

1. Let $P_1(\cdot \mid \mathcal{A})(\omega)$ be the "indicator" rcd that concentrates all its mass at ω, i.e. for $B \in \mathcal{B}, P(B \mid \mathcal{A})(\omega) = I_B(\omega)$. It is a simple fact that there always is such an obvious rcd on a space $\mathcal{B}$ given $\mathcal{B}$, regardless the algebraic structure of $\mathcal{B}$.

2. Let $P_2(\cdot \mid \mathcal{A})(\omega)$ be defined so that $P_2(\cdot \mid \mathcal{A})(\omega) = P(\cdot)$, for each point ω. It is straightforward to verify that this function is an rcd for $\mathcal{B}$ given $\mathcal{A}$.

Note that, for each $\omega, P_1(\{\omega\} \mid \mathcal{A})(\omega) = 1$ and $P_2(\{\omega\} \mid \mathcal{A})(\omega) = 0$, so these are mutually singular distributions, as evaluated at each point, ω. The second of the two rcd's in Example 1 displays an anomaly that is the focus of the balance of this paper.

3. Proper rcd's. For our investigation of the received theory of conditional probability, the central concept comes from important works by Blackwell and Ryll-Nardzewski (1963) and Blackwell and Dubins (1975).

DEFINITION 5. An rcd $P(\cdot \mid \mathcal{A})$ on $\mathcal{B}$ given $\mathcal{A}$ is *proper at the point* ω if $P(A \mid \mathcal{A})(\omega) = 1$ whenever $\omega \in A \in \mathcal{A}$. Say that $P(\cdot \mid \mathcal{A})$ on $\mathcal{B}$ given $\mathcal{A}$, is *improper at* ω otherwise. An rcd $P(\cdot \mid \mathcal{A})$ on $\mathcal{B}$ given $\mathcal{A}$ is *proper* if it is proper at each point ω.

The extent of impropriety for rcd's is the principal subject of this paper. Where an rcd is improper at ω, its conditional probability function evaluated at ω cannot be used as a *coherent* degree of belief, at least, in the sense of *coherence* intended by deFinetti (1974) or Savage (1954). That is, we understand coherence of degrees of belief to include the requirement that a conditional probability function is supported by its conditioning event. Conditioning on a σ-field does not entail conditioning on the events in the σ-field. However, if conditioning on a σ-field is to represent coherent degrees of belief, then the rcd should be proper.

We begin our discussion of the extent of impropriety of rcd's with an important and, we find, surprising result due to Blackwell and Dubins (1975).

DEFINITION 6. A probability distribution is *extreme* if its range is the two point set $\{0, 1\}$.

THEOREM 2. (Blackwell and Dubins, 1975) *If $\mathcal{B}$ is a countably generated σ-field and if there exists* some *extreme probability on $\mathcal{A}$ supported by no $\mathcal{A}$-atom belonging to $\mathcal{A}$, then $\mathcal{A}$ is not countably generated, which entails that* no *probability admits a proper rcd on $\mathcal{B}$ given $\mathcal{A}$.*

Thus, this result gives a sufficient condition for when an rcd cannot be proper.

We index the extent of impropriety of an rcd at a point ω with the Definition 7.

DEFINITION 7. Fix ω and consider those A such that $\omega \in A \in \mathcal{A}$. If for some $\omega \in A \in \mathcal{A}$, $P(A \mid \mathcal{A})(\omega) = 0$, say that $P(\cdot \mid \mathcal{A})$ is *maximally improper at* ω. Otherwise, if for each $\omega \in A \in \mathcal{A}$, $1 > P(A \mid \mathcal{A})(\omega) > 0$, say that the rcd is *modestly proper at* ω.

In order to characterize the extent of impropriety of an rcd globally, across different states, we consider the inner P-measure of the set of points where it is improper. Let $\underline{P}$ denote the inner P-measure of a set.

DEFINITION 8. Let $B = \{\omega : P(\cdot \mid \mathcal{A})(\omega) \text{ is improper at } \omega\}$. Call $\underline{P}(B)$ the *lower P-bound on the extent of impropriety of the rcd* $P(\cdot \mid \mathcal{A})$. If B is P-measurable, call $\underline{P}(B)$ the *extent of impropriety of the rcd* $P(\cdot \mid \mathcal{A})$. Finally, say that $P(\cdot \mid \mathcal{A})$ is *maximally improper* if, with lower P-bound 1, it is maximally improper. That is, an rcd is maximally improper if, with respect to its measure completion, it is almost surely maximally improper.

EXAMPLE 2. (Example 1 continued). Evidently, rcd $P_1(\cdot \mid \mathcal{A})(\omega)$ is everywhere proper. However, rcd $P_2(\cdot \mid \mathcal{A})(\omega)$ is maximally improper!

In light of Theorem 1, if an rcd $P(\cdot \mid \mathcal{A})$ exists, then when $\mathcal{A}$ is countably generated, almost surely the rcd is proper. That is, then the extent of its impropriety is 0 and impropriety is restricted to a P-null set, at most. Blackwell (1955, p. 6) asked whether this null set can be reduced to the empty set when $\mathcal{B}$ is a Lusin space. Blackwell and Ryll-Nardzewski (1963) establish that the answer is negative when $\mathcal{A}$ is the σ-field generated by a real-valued random variable whose range is not a Borel set. We discuss their result in the next section, where we relate it to non-measurable sets when the conditioning sub-σ-field is the tail field or field of symmetric events.

Now for our central theorem about the extent of impropriety of rcd's. Generally, when the sufficient condition of Theorem 2 is satisfied, rcd's are maximally improper.

THEOREM 3. *Let $\mathcal{A}$ be an atomic sub-σ-field of $\mathcal{B}$. Assume that P is an extreme probability on $\mathcal{A}$ that is not supported by any $\mathcal{A}$-atoms. An rcd $P(\cdot \mid \mathcal{A})$ for P on $\mathcal{B}$ given $\mathcal{A}$ exists and is maximally improper.*

Remark: By Lemma 2, this rcd is unique when $\mathcal{B}$ is countably generated.

PROOF. By assumption, P is extreme on $\mathcal{A}$. Therefore, as is evident, $P(\cdot \mid \mathcal{A}) = P(\cdot)$ is an rcd for P on $\mathcal{B}$ given $\mathcal{A}$. That is: (1) for each point $\omega, P(\cdot \mid \mathcal{A})(\omega)$ is a probability on $\mathcal{B}$. Equally evident, (2) for each $B \in \mathcal{B}, P(B \mid \mathcal{A})(\cdot)$ is an $\mathcal{A}$-measurable function, with pre-image either Ω or $\emptyset$. Moreover, it is constant at every point ω, and thus it is constant on the atoms of $\mathcal{A}$. Finally (3), if $P(A) = 1$, then $P(B \cap A) = P(B) = \int_\Omega P(B)dP(\omega) = \int_A P(B \mid \mathcal{A})(\omega)dP(\omega)$; and if $P(A) = 0$, then $P(B \cap A) =$

$0 = \int_A P(B)dP(\omega) = \int_A P(B \mid \mathcal{A})(\omega)dP$. But, as P is extreme on $\mathcal{A}$ and is not supported by any $\mathcal{A}$-atoms, $P(a) = 0$ for each $\mathcal{A}$-atom a. Hence (P-almost surely), this rcd $P(\cdot \mid \mathcal{A})(\omega)$ on $\mathcal{B}$ satisfies $P(a \mid \mathcal{A})(\omega) = 0$ for each $\mathcal{A}$-atom a. Denote by $a(\omega)$ that $\mathcal{A}$-atom containing the point ω. Thus, for almost all points $\omega, P(a(\omega) \mid \mathcal{A})(\omega) = 0$. which establishes that this rcd is maximally improper.□

Here are two additional illustrations of Theorem 3, counting the rcd $P_2(\cdot \mid \mathcal{A})(\omega)$ of Example 1 as the first example. By contrast, we use a $\mathcal{B}$ that is countably generated in each of the next two examples.

Example 3. (see Blackwell and Dubins, 1975, p. 742). Let $\Omega = \{0,1\}^{\aleph_0}$, i.e., the sample space of infinite binary sequences; let $\mathcal{B}$ be the product σ-field; and let P be the product measure corresponding to independent flips of a "fair" coin, i.e. $P(0 \times \{0,1\} \times \ldots) = P(1 \times \{0,1\} \times \ldots) = 1/2$, etc. Let $\mathcal{A}$ be the tail σ-field for this process. Then, by the Kolmogorov 0-1 law, for each $A \in \mathcal{A}, P(A) = 0$ or $P(A) = 1$. The $\mathcal{A}$-atoms, a, are countable sets of points, where $\omega', \omega \in a$ if and only if they differ in at most finitely many places. These A-atoms belong to the tail field, $a \in \mathcal{A}$. Since each $\mathcal{A}$-atom is a countable set, $P(a) = 0$; hence, P is not supported by any of its A-atoms. With $P(\cdot \mid \mathcal{A}) = P(\cdot)$ the rcd on $\mathcal{B}$, given $\mathcal{A}$, we have that for each $\mathcal{A}$-atom, a, $P\{\omega : P(a \mid \mathcal{A})(\omega) = 0\} = 1$. In particular, $P\{\omega : P(a(\omega) \mid \mathcal{A})(\omega) = 0\} = 1$, and this rcd is maximally improper. The example has a natural generalization to i.i.d. binomial "weighted" coin flipping. $P_\theta(1 \times \{0,1\} \times \ldots) = \theta$, for $0 < \theta < 1$, which we pursue in Corollary 2 for symmetric measures.

Example 4. (see Billingsley (1995), Example 33.11). Let $\Omega = [0,1]$, let $\mathcal{B} =$ the Borel subsets of Ω, and let P be Lebesgue measure. Let $\mathcal{A}$ be the sub-σ-field of all countable and co-countable sets in $[0,1]$. Clearly, $P(A) = 0$ or $P(A) = 1$, for each $A \in \mathcal{A}$. Equally obviously, $P(A) = 0$ for each countable set A. Note also that the $\mathcal{A}$-atoms, which in fact belong to $\mathcal{A}$, are just the singleton sets consisting of the points of Ω, $\{\{x\} : 0 \leq x \leq 1\}$. Hence, according to Theorem 3, the rcd on $\mathcal{B}$ given $\mathcal{A}$, $P(\cdot \mid \mathcal{A})$, satisfies

$$P\{x : P(x' \mid \mathcal{A})(\omega) = 0, \text{ for } 0 \leq x, x' \leq 1\} = 1.$$

Thus, $P(\{x : P(\{x\} \mid \mathcal{A})(\omega) = 0\}) = 1$.

Next, we discuss the σ-field of symmetric events, as covered by the 0-1 law of Hewitt and Savage (1955). We use the space of sequences of Cartesian products of binary events, as in Example 3; however, Theorem 3 generalizes directly to products of an arbitrary finite set. Thus, 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, 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. Thus, $T : \Omega \to \Omega$, is 1-1, onto, and leaves all but finitely many coordinates of a point ω unchanged. Given T, define the set $T^{-1}B$ as $\{\omega : T(\omega) \in B\}$. P is called a *symmetric probability* if $P(T^{-1}B) = P(B)$, for each $B \in \mathcal{B}$ and each T. If $B = T^{-1}B$ for all (finite) permutations T, B is called a *symmetric event*. Hewitt and Savage (1955, Theorem 6.3) shows (duplicating deFinetti's representation theorem) that each symmetric probability P is an average (integral) of "extreme" symmetric probabilities of the form

$$P(\cdot) = \int_{\Theta} P_\theta(\cdot) d\mu(\theta)$$

where $0 \leq \theta \leq 1$, where $P_\theta(\cdot)$ is the i.i.d. (binomial) product probability on $\mathcal{B}$, with $P_\theta\{1 \times \{0,1\} \times \ldots\} = \theta$, and where $\mu(\cdot)$ is a "prior" probability on Borel subsets of the unit interval. The representation is unique in μ. Let $\mathcal{A}$ be the sub-σ-field of $\mathcal{B}$ generated by the class T of all (finite) permutations of the coordinates of Ω, i.e., $\mathcal{A}$ is the σ-field of the symmetric events. Denote by a the $\mathcal{A}$-atoms. These are denumerable sets of points, which are elements of $\mathcal{A}$. That is, all but two $\mathcal{A}$-atoms are countably infinite sets of points related by the equivalence relation that elements differ by a finite permutation of their sequences. The two distinguished A-atoms are the two constant sequences $< 0, 0, \ldots >$ and $< 1, 1, \ldots >$.

We establish our result for the class of symmetric probabilities as a corollary to the following theorem, which itself generalizes Theorem 3.

THEOREM 4. *Let $(\Theta, \mathcal{D})$ be a Borel space. For each $\theta \in \Theta$, let P_θ be a probability on $\mathcal{B}$. Let $P(\cdot)$ be defined on $\mathcal{B}$, by the integral, $P(\cdot) = \int_\Theta P_\theta(\cdot)d\mu(\theta)$. Let $\mathcal{A}$ be a sub-σ-field of $\mathcal{B}$ for which there exists a marginal rcd on $\mathcal{B}$ given $\mathcal{A}$, denoted by $P(\cdot|\mathcal{A})$ and assume that $P_\theta(\cdot|\mathcal{A})$ is maximally improper for P-almost all θ. Then $P(\cdot|\mathcal{A})$ is maximally improper as well.*

The proof of Theorem 4 is straightforward from the following lemma.

LEMMA 3. *Let $(\Theta, \mathcal{D})$ be a Borel space, with a probability measure μ. For each $\theta \in \Theta$, let P_θ be a probability on $\mathcal{B}$ such that for every $B \in \mathcal{B}$, $P_\theta(B)$*

is a measurable function of θ. Define the probability P on $\mathcal{B}$ by $P(B) = \int_\Theta P_\theta(B)d\mu(\theta)$. Let $P(\cdot \mid \mathcal{A})$ be an rcd given a sub-σ-field $\mathcal{A}$ of $\mathcal{B}$. Also, let $P_\theta(\cdot \mid \mathcal{A})$ denote an rcd for each P_θ. Then, for each ω there exists a probability ν_ω on $\mathcal{D}$ such that for all $D \in \mathcal{D}$

$$P(D \mid \mathcal{A})(\omega) = \int_\Theta P_\theta(D \mid \mathcal{A})(\omega)d\nu_\omega(\theta), \tag{1}$$

almost surely with respect to P.

PROOF. Let $\mathcal{E}$ be the product σ-field $\mathcal{B} \otimes \mathcal{D}$. For each $E \in \mathcal{E}$, define

$$E_\theta = \{\omega : (\omega, \theta) \in E\},$$

the θ-section of E. It is easy to see that, if E is a product set, i.e., $E = B \times D$ for $B \in \mathcal{B}$ and $D \in \mathcal{D}$, then $E_\theta \in \mathcal{B}$ for all θ, and $P_\theta(E_\theta)$ is a measurable function of θ. The $\pi - \lambda$ theorem of Dynkin (see Billingsley, 1995, Theorem 3.2) implies that for all $E \in \mathcal{E}$, $E_\theta \in \mathcal{B}$ for all θ and $P_\theta(E_\theta)$ is a measurable function of θ. Define

$$Q(E) = \int_\Theta P_\theta(E_\theta)d\mu(\theta),$$

which is easily seen to be a probability on $\mathcal{E}$. Let $\mathcal{A}' = \{B \times \Theta : B \in \mathcal{A}\}$, which is a sub-$\sigma$-field of $\mathcal{E}$. Let $Q(\cdot \mid \mathcal{A}')$ be an rcd. Clearly, $Q(E \mid \mathcal{A}')(\omega, \theta)$ is a function of ω only since it is $\mathcal{A}'$-measurable. It is easy to see that for all D, $P(D \mid \mathcal{A})$ is a version of $Q(\Omega \times D \mid \mathcal{A}')$. Next, let $\mathcal{A}'' = \mathcal{A} \otimes \mathcal{D}$ so that $\mathcal{A}'$ is a sub-σ-field of $\mathcal{A}''$. It is easy to see that for all D $P_\theta(D \mid \mathcal{A})$ is a version of $Q(\Omega \times D \mid \mathcal{A}'')$. For each $D \in \mathcal{D}$ and $\omega \in \Omega$, define $\nu_\omega(D) = Q(\Omega \times D \mid \mathcal{A}')(\omega)$. The law of total probability (see Schervish, 1995, Theorem B.70, p. 632) now says that (1) holds. □

COROLLARY 2. *Each rcd $P(\cdot|\mathcal{A})$ on $\mathcal{B}$ given $\mathcal{A}$, for a symmetric probability P, is maximally improper provided that the two distinguished $\mathcal{A}$-atoms are P-null events, $P\{< 0, 0, 0, \ldots >\} = P\{< 1, 1, 1 \ldots >\} = 0$.*

PROOF. We apply Theorem 4 to the Hewitt-Savage representation for a symmetric probability P, using the sub-σ-field of symmetric events as $\mathcal{A}$. By the Hewitt-Savage 0-1 law, $P_\theta(A) = 0$ or $P_\theta(A) = 1$ for each $A \in \mathcal{A}$ and each extreme measure $P_\theta(\cdot)$. Evidently, for each $0 < \theta < 1$, and for each $\mathcal{A}$-atom a, $P_\theta(a) = 0$, so that P_θ is not supported by any of the $\mathcal{A}$-atoms. Then, by Theorem 3, P_θ-almost surely, $P_\theta(a \mid \mathcal{A})(\omega) = 0$ for each $\mathcal{A}$-atom a and so $P_\theta(\cdot|A)(\omega)$ is maximally improper. In fact, for this case θ is an $\mathcal{A}$-measurable function. (Note that for a symmetric probability P, almost surely the infinite sequence of binary events has a limiting frequency for 1's, say, which is an $\mathcal{A}$-event of P-measure 1. Almost surely, θ of the Hewitt-Savage representation equals this limiting frequency; hence, θ is $\mathcal{A}$-measurable.) Thus, in the conclusion of Lemma 3 as applied to our situation, almost surely $\nu_\omega(\cdot)$ is a point-distribution concentrated at the value of θ consistent with ω. $\square$

4. Impropriety of rcd's and some non-measurable sets. Dubins (1971) identifies a different argument from the one of Theorem 2, establishing that there cannot be everywhere proper rcd's for $\mathcal{B}$ given $\mathcal{A}$ in Example 3. His insight is this. In Example 3, were the rcd $P(\cdot \mid \mathcal{A})(\omega)$ for $\mathcal{B}$ given $\mathcal{A}$ everywhere proper, then there would then be an $\mathcal{A}$-measurable selection function on the atoms of $\mathcal{A}$ whose range is an analytic (hence Lebesgue measurable) set. As the $\mathcal{A}$-atoms are denumerable sets, a proper rcd $P(\cdot \mid \mathcal{A})(\omega)$ is a discrete distribution that lives on the atom $a(\omega)$ that contains ω. For instance, the mode of each distribution, $P(\cdot \mid \mathcal{A})(\omega)$, could serve to define a selection function – a function that picks out exactly one element from each $\mathcal{A}$-atom. However, the range of such a selection function is a Vitali-style non-measurable set, which is a contradiction. That is, the "fair coin" product measure is invariant to changes in a finite number of the coordinates in each binary sequence of a measurable set – corresponding to the fact that Lebesgue measure is (translation) invariant under the addition/subtraction of a fixed (binary rational) number to each real number in a measurable set. However, in Example 3, Ω is covered by countably many such changes to the range of any selection function on the $\mathcal{A}$-atoms. But as P cannot be uniform over a countably infinite set, this contradicts the fact that the range of the selection function is analytic.

We adapt this line of reasoning involving non-measurable sets to establish the following:

THEOREM 5. *Let $\mathcal{B}$ be the Borel subsets of Ω, let $\mathcal{A}$ be the sub-σ-field of symmetric events, and let P be a symmetric probability that assigns 0 to the two distinguished atoms. Then, with respect to elements of $\mathcal{A}$, the P-lower bound is 0 on the set of points where $P(\cdot \mid \mathcal{A})$ can be even modestly proper.*

The proof of Theorem 5 uses the following result of Blackwell and Ryll-Nardzewski (1963):

THEOREM 6. *(Theorem 2 of Blackwell and Ryll-Nardzewski, 1963) Let* X, Y *be Borel subsets of complete separable metric spaces, let* $\mathcal{C}$ *be a countably generated sub-*σ*-field of the* σ*-field of Borel subsets of* X *and let* $\mathcal{B}$ *be the class of Borel subsets of* Y*. For any function* μ *on* $\mathcal{B} \times X$ *such that (a)* $\mu(\cdot, x)$ *is for each* x *a probability measure on* $\mathcal{B}$ *and (b) for each* $B \in \mathcal{B}$*,* $\mu(B, \cdot)$ *is an* $\mathcal{C}$*-measurable function on* X*, and any set* $S \in \mathcal{B} \times X$ *such that* $\mu(S_x, x) > 0$ *for all* $x \in X$*, where* S_x *denotes the* x*-section of* S*, i.e.,* $S_x = \{y : (y, x) \in S\}$*, then there is a* $\mathcal{C}$*-measurable function* g *from* X *into* Y *whose graph is a subset of* S*, i.e.,* $(g(x), x) \in S$ *for all* $x \in X$*.*

PROOF OF THEOREM 5. Let F be the set of points ω where the rcd $P(\cdot \mid \mathcal{A})(\omega)$ for $\mathcal{B}$ given $\mathcal{A}$ is modestly proper. Assume for an indirect proof that, with respect to sets in $\mathcal{A}$, $\underline{P}(F) > 0$. Then let $A \in \mathcal{A}$, $F \supseteq A$ denote a set of positive measure. We use Theorem 6 iteratively to find a countable sequence of selection functions whose ranges, though measurable sets, each behaves as a Vitali-styled non-measurable set. These sets lead to a countable partition of A into sets of measure 0 events, which contradicts the fact that $P(A) > 0$.

Reason as follows. Let $\mathcal{A}^*$ be the smallest sub-σ-field with respect to which $P(\cdot \mid \mathcal{A})$ over $\mathcal{B}$ is measurable. Trivially, $\mathcal{A}^* \subseteq \mathcal{A}$. In the case considered, $\mathcal{A}^*$ is countably generated (hence atomic), because $\mathcal{B}$ is. Recall that each $\mathcal{A}$-atom is a countable set and that each $\mathcal{A}^*$-atom, a^*, consists of that union of $\mathcal{A}$-atoms $a^* = \cup a_\alpha$ such that each point $\omega \in a^*$ yields the same distribution $P(\cdot \mid \mathcal{A})(\omega)$ over $\mathcal{B}$ as do the other points in a^*. As $P(\cdot \mid \mathcal{A})(\omega)$ is modestly proper over A, each atom a^* contains a finite or at most denumerable union of $\mathcal{A}$-atoms from A. However, a^* may contain uncountably many $\mathcal{A}$-atoms from A^c.

In our first application of Theorem 6, let $X_1 = Y_1 = A$. Let $\mathcal{B}_1 = \mathcal{B}/A$ and $\mathcal{C}_1 = \mathcal{A}^*/A$, the quotient σ-fields, respectively of $\mathcal{B}$ and $\mathcal{A}^*$ given A. Clearly, $\mathcal{C}_1$ is countably generated with (uncountably many) atoms c_1. Let $\mu_1(\cdot, \omega) = \frac{P(\cdot \cap A \mid \mathcal{A}^*)}{P(A \mid \mathcal{A}^*)}(\omega)$ for $\omega \in A$. Last, let

$$S_1 = \{(\omega', \omega) : \omega' \in c_1(\omega) \text{ and } \omega \in A\}.$$

Evidently, $\mu_1(\cdot, \omega)$ satisfies the requisite conditions in Theorem 6. Then apply Theorem 6 to argue that there is a $\mathcal{C}_1$-measurable selection function $g_1(c_1(\omega))$ that picks out one element from each $\mathcal{C}_1$-atom $c_1(\omega)$ for each $\omega \in A$.

Let $V_{1,1}$ be the range of this function. (We use $'V'$ to remind the reader of the Vitali-like properties of this range.) We argue that, as $V_{1,1}$ is $\mathcal{B}_1$-measurable, $P(V_{1,1}) = 0$ using P's symmetries under finite permutations of the binary sequences that are the points of Y. Consider the countable set of finite permutations of a binary sequence, which we write as $PER = \{per_j : j = 1, ...\}$. For simplicity we let per_1 be the identity function. Then, as P is is a symmetric probability, it is invariant under the application of each element of PER to a measurable set. Thus P assigns equal probability to each of the countably many disjoint sets $per_j(V_{1,1}) = V_{1,j}$. We can say more. Let $V_1 = \cup\, V_{1,j}$. Then $P(V_1) = 0$. Moreover, we see that V_1 is an uncountable union of $\mathcal{A}$-atom a_α, $V_1 = \cup a_\alpha$, where each such $\mathcal{A}$-atom a_α is a subset of a distinct $\mathcal{C}_1$-atom $c_{1,\alpha}$ and where each $\mathcal{C}_1$-atom has one such an $\mathcal{A}$-atom as its witness. Let $A_1 = A - V_1$

We iterate the application of Theorem 6 by induction through a countable set of countable ordinals as follows, until we arrive at a stage ε where $\mathrm{X}_\varepsilon = \emptyset$.

For a successor ordinals $\beta + 1$ set $X_{\beta+1} = Y_{\beta+1} = A_\beta$. Let $\mathcal{B}_{\beta+1} = \mathcal{B}/A_\beta$ and $\mathcal{C}_{\beta+1} = \mathcal{A}^*/A_\beta$. Let $\mu_{\beta+1}(\cdot, \omega) = \frac{P(\cdot \cap A_\beta | \mathcal{A}^*)}{P(A_\beta | \mathcal{A}^*)}(\omega)$ for $\omega \in A_\beta$. Last, let $S_{\beta+1} = \{(\omega', \omega) : \omega' \in c_{\beta+1}(\omega) \text{ and } \omega \in A_\beta\}$.

For γ a countable limit ordinal, define the respective sets by intersections in the usual fashion for such constructions, as follows. With $\beta < \gamma$, let $X_\gamma = Y_\gamma = A_\gamma = \cap A_\beta$. Set $\mathcal{B}_\gamma = \mathcal{B}/A_\gamma$ and $\mathcal{C}_\gamma = \mathcal{A}^*/A_\gamma$. Let $\mu_\gamma(\cdot, \omega) = \frac{P(\cdot A\gamma)|\mathcal{A}^*)}{P(A_\gamma)|\mathcal{A}^*)}(\omega)$ for $\omega \in A\gamma$. Last, let $S_\gamma = \{(\omega', \omega) : \omega' \in c_\gamma(\omega) \text{ and } \omega \in A_\gamma\}$.

In the former case we obtain a $\mathcal{C}_{\beta+1}$-measurable selection function $g_{\beta+1}(c_{\beta+1}(\omega))$ that picks out one element from each $\mathcal{C}_{\beta+1}$-atom $c_{\beta+1}(\omega)$ for each $\omega \in A_\beta$. Let $V_{\beta+1,1}$ be the range of this function. Then, $P(V_{\beta+1,1}) = 0$, and with $V_{\beta+1} = \cup_j V_{\beta+1,j}$, we have also $P(V_{\beta+1}) = 0$. In the latter case, the same argument leads to the conclusion that $P(V_\gamma) = 0$. However, as each atom a^* contains a finite or at most denumerable union of $\mathcal{A}$-atoms from A, this process exhausts A after some countable number of iterations. That is, there exists a countable ordinal ζ such that $A = \cup_{\beta<\zeta} V_\beta$. This completes the proof as $0 < P(A) = P(\cup_{\beta<\zeta} V_\beta) = \sum_{\beta<\zeta} P(V_\beta) = 0$, a contradiction. Hence, $\underline{P}(F) = 0$, and P's rcd cannot be even modestly proper over a set of positive P-measure, given the symmetric field $\mathcal{A}$. $\square$

COROLLARY 3. *Under the same conditions as Theorem 5 there is no extension of the symmetric probability P to a larger σ-field $\mathcal{B}'$ that has positive lower bound on the set of points where its rcd given the symmetric events $\mathcal{A}$ is modestly proper.*

PROOF. Apply Corollary 1 to the preceding theorem to establish that no extension of P to a σ-field $\mathcal{B}'$ that includes a P-non-measurable set admits an rcd, proper or not, given $\mathcal{B}$. $\Box$

5. Conclusions. We have examined the received theory of regular conditional distributions for certain anomalous behavior, *impropriety*, with respect to conditioning on sub-σ-fields. Impropriety was studied in papers by Blackwell (1955), Blackwell and Ryll-Nardzewski (1963), Dubins (1971) and Blackwell and Dubins (1975), where their focus was on the impossibility of everywhere proper rcds.. Here, we provide a index for the extent of impropriety in an rcd based on the measure of the set of points where the rcd is not proper, and by how much it is not proper. When rcd's exist and the sub-σ-field is countably generated, almost surely the rcd is *proper*. However, when the sub-σ-field is not countably generated, the door is opened to the possibility that almost surely $\mathrm{P}(a(\omega) \mid \mathcal{A}^*)(\omega) = 0$, a situation we call *maximally improper*. We offer sufficient conditions for an rcd to be maximally improper and show that they obtain in some familiar cases, e.g., given the tail σ-field of a mixture of i.i.d. process and given the σ-field of symmetric events for symmetric probabilites.

Insofar as the anomalous sub-σ-fields involve conditioning on (countable) Borel sets, we believe that it is unlikely that a minor variation in the theory of rcd's can avoid impropriety in its conditional distributions. A rival theory does exist, however, that assures not only that condtional probabilty is everywhere proper, but also allows conditional probability to be coherently defined *given an event* rather than given a sub-σ-field, hence solving the Borel paradox. Also, this theory does not encounter the limitation of non-measurable events; hence, coherent conditional probability always exists. We have in mind, of course, the theory of finitely additive probability, as described by Dubins (1975, Section 3).

The price for all these benefits is not insignificant. The theory of finitely additive conditional probability does not always satisfy the integral equation, condition 3, for rcd's. With finitely additive probability, it generally does not happen that, for $B \in \mathcal{B}$, $\int_{\Omega} P(B \mid \mathcal{A})(\omega) dP(\omega) = P(B)$. When this equation fails, there the finitely additve P is not *disintegrable* in the partition of the $\mathcal{A}$-atoms. In particular, Schervish, Seidenfeld and Kadane (1984) shows that each finitely but not countably additive probability P will fail to be disintegrable in some denumerable partition.

For an example of what might usefully be done with finitely additive probabilities, Dubins (1977) shows there exists a coherent finitely additive probability P that extends the "fair coin" product measure of Example 3, μ, and which is disintegrable in the partition $\pi_{\mathcal{T}}$ formed by the atoms $a_{\mathcal{T}}$ of the anomalous tail-field, $\mathcal{T}$. Since P extends μ and is disintegrable in the partiton $\pi_{\mathcal{T}}$ where μ obeys the Kolmogorov 0-1 law, its conditional probability distributions behave P-almost surely like the maximally improper rcd's based on μ with respect to sets of *positive* μ-measure. Thus, one can compute values of P for sets of positive μ-measure by using the familiar, maximally improper rcd for μ given $\mathcal{T}$, which we here denote by $\mu(B \mid \mathcal{T})(\omega)$. However, since P is coherent, its conditional probability distributions are everywhere proper. That is, the conditional probability distrbution $P(\cdot|a_{\mathcal{T}})$ is supported by its conditioning event, the countable set $a_{\mathcal{T}}$. Thus, $P(\cdot|a_{\mathcal{T}})$ and $\mu(B \mid \mathcal{T})(\omega)$ are mutually singular at each element of π. To heighten the tension, $P(\cdot|a_{\mathcal{T}})$ may be a "uniform" purely finitely additive measure over its conditioning event, i.e., $P(\omega|a_{\mathcal{T}}) = 0$ for each $\omega \in a_{\mathcal{T}}$, as Dubins' result establishes.

It is an open general question, we believe, whether useful finitely additive extensions of familiar countably additive probabilties always exist for conditioning in cases where an atomic sub-σ-field $\mathcal{A}$ is not countably generated and the countably additive rcd is maximally improper, or whether, instead, failures of distintegration are unavoidable in those same partitions for all extensions of the countably additive probability of interest.

REFERENCES

Billingsley, P. (1995). *Probability and Measure* (3 ed.) Wiley, New York.

Blackwell, D. (1955). On a class of probability spaces. In *Proc. Third Berkeley Symp. Math. Statist. Prob.* (J. Neyman, ed.) 1–6. University of California Press, Berkeley.

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

Blackwell, D. and Ryll-Nardzewski, C. (1963). Non-existence of everywhere proper conditional distributions. *Ann. Math. Statist.* **34** 223–225.

Breiman, L. (1968). *Probability*, Addison-Wesley, Reading, MA.

deFinetti, B. (1974). *The Theory of Probability*, Wiley, New York.

Doob, J. L. (1953). *Stochastic Processes*, Wiley, New York.

Dubins, L. E. (1971). On conditional distributions for stochastic processes. In *Proceedings of the Symposium in Probability Theory* 72–85. University of Aarhus, Denmark.

Dubins, L. E. (1975). Finitely additive conditional probabilities, conglomerability, and disintegrations. *Ann. Probab.* **3** 89–99.

Dubins, L. E. (1977). Measurable, tail disintegrations of the haar integral are purely finitely additive. *Proc. Amer. Math. Soc.* **62** 34–36.

Halmos, P. R. (1950). *Measure Theory*, van Nostrand, New York.

Hewitt, E. and Savage, L. J. (1955). Symmetric measures on cartesian products. *Trans. Amer. Math. Soc.* **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* (P. Goel and A. Zellner, eds.) 59–76. Elsevier Science Publishers, New York.

Kolmogorov, A. (1950). *Foundations of the Theory of Probability*, Chelsea, New York.

Loeve, M. (1955). *Probability Theory*, van Nostrand, New York.

Savage, L. J. (1954). *The Foundations of Statistics*, Wiley, New York.

Schervish, M. J. (1995). *Theory of Statistics*, Springer-Verlag, New York.

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

Teddy Seidenfeld
Department of Philosophy
Carnegie Mellon University
Pittsburgh, Pennsylvania 15213
E-mail: teddy@stat.cmu.edu

Mark J. Schervish
Joseph B. Kadane
Department of Statistics
Carnegie Mellon University
Pittsburgh, Pennsylvania 15213
E-mail: mark@stat.cmu.edu,
kadane@stat.cmu.edu