Law of total probability and Bayes’ theorem in Riesz spaces

Liang Hong

Abstract. This note generalizes the notion of conditional probability to Riesz spaces using the order-theoretic approach. With the aid of this concept, we establish the law of total probability and Bayes’ theorem in Riesz spaces; we also prove an inclusion-exclusion formula in Riesz spaces. Several examples are provided to show that the law of total probability, Bayes’ theorem and inclusion-exclusion formula in probability theory are special cases of our results.

1. Introduction

The study of stochastic processes in an abstract space can date back to as early as [2]. Works along this line include [4], [5], [15], [26] and [27]. About a decade ago, [17] initiated the study of stochastic processes in a measure-free setting. They carefully investigated the fundamental properties of conditional expectations on a probability space and abstract them to Riesz spaces; this allows one to generalize many concepts from probability theory to Riesz spaces. As a result, an order-theoretic approach to stochastic process can be formed to generalize the classical theory of stochastic processes. Since then this order-theoretic approach to stochastic processes has been developing fast. For order-theoretic investigation of conditional expectations and related concepts, we refer to [14], [15], [20], [31]; for discrete-time processes in Riesz spaces, we refer to [16], [17], [18], [19], [21], [22], [23], [28]; for continuous-time processes in Riesz spaces, we refer to [6], [7], [8], [9], [10], [13], [29], [30]; for stochastic integrals in Riesz spaces, we refer to [11], [12], [24].

[22] extended the notation of independence to Riesz spaces; [9] gave a slightly more general definition of independence in Riesz spaces. In this note, we follow [22] to extend the concept of conditional probability to Riesz spaces and then establish the law of total probability and Bayes’ theorem in Riesz spaces; we also prove an inclusion-exclusion formula in Riesz spaces. Several examples are given to show that the concept of conditional probability, law of total probability, Bayes’ theorem and inclusion-exclusion formula in probability theory are special cases of our results.

The remainder of this note is organized as follows. Section 2 review some basic concepts and results of Riesz spaces and conditional expectations in Riesz spaces; for further details, we refer readers to [1], [3], [20], [25], [31] and [32]. Section 3 generalizes the concept of conditional probability to Riesz spaces; we show that

2010 Mathematics Subject Classification. Primary 60A99, 62C10, 46N30; Secondary 60B99, 06F20.

Key words and phrases. Conditional probability; law of total probability; Bayes’ theorem; inclusion-exclusion formula; Riesz spaces.
this notion leads naturally to the notion of independence in Riesz spaces. Section 4 establishes the law of total probability, Bayes’ theorem and the inclusion-exclusion formula in Riesz spaces.

2. Preliminaries of Riesz space theory

A partially ordered set \((X, \leq)\) is called a **lattice** if the infimum and supremum of any pair of elements in \(X\) exist. A real vector space \(X\) equipped with a partial order \(\leq\) is called an **ordered vector space** if its vector space structure is compatible with the order structure in a manner such that

(a) if \(x \leq y\), then \(x + z \leq y + z\) for any \(z \in X\);
(b) if \(x \leq y\), then \(\alpha x \leq \alpha y\) for all \(\alpha \geq 0\).

An ordered vector space is called a **Riesz space** (or a **vector lattice**) if it is also a lattice at the same time. For any pair \(x, y\) in a Riesz space, \(x \vee y\) denotes and supremum of \(\{x, y\}\), \(x \wedge y\) denotes the infimum of \(\{x, y\}\), and \(|x|\) denotes \(x \vee (-x)\). A Riesz space \(X\) is said to be **Dedekind complete** if every nonempty subset of \(X\) that is bounded from above has a supremum. A vector subspace of a Riesz space is said to be a **Riesz subspace** if it is closed under the lattice operation \(\vee\). A subset \(Y\) of a Riesz space \(X\) is said to be **solid** if \(|x| \leq |y|\) and \(y \in Y\) imply that \(x \in Y\).

A solid vector subspace of a Riesz space is called an **ideal**. A net \((x_\alpha)_{\alpha \in A}\) in a Riesz space \(X\) is said to be **decreasing** if \(\alpha \geq \beta\) implies \(x_\alpha \leq x_\beta\). The notation \(x_\alpha \downarrow x\) means \((x_\alpha)_{\alpha \in A}\) is a decreasing net and the infimum of the set \(\{x_\alpha \mid \alpha \in A\}\) is \(x\). A net \((x_\alpha)_{\alpha \in A}\) in a Riesz space \(X\) is said to be **order-convergent** to an element \(x \in X\), often written as \(x_\alpha \rightharpoonup x\), if there exists another net \((y_\alpha)_{\alpha \in A}\) in \(X\) such that \(|x_\alpha - x| \leq y_\alpha \downarrow 0\).

A subset \(A\) of a Riesz space \(X\) is said to be **order-closed** if a net \((x_\alpha)\) in \(A\) order-converges to \(x_0 \in X\) implies that \(x_0 \in A\). An order-closed ideal is called a **band**. Two elements \(x\) and \(y\) in a Riesz space \(X\) is said to be **disjoint** if \(|x| \wedge |y| = 0\) and is denoted by \(x \perp y\). The **disjoint complement** of a subset \(A\) of a Riesz space \(X\) is defined as \(A^d = \{x \in X \mid x \wedge y \forall y \in A\}\). A band \(B\) in a Riesz space is said to be a **projection band** if \(X\) equals the direct sum of \(B\) and \(B^d\), that is, \(X = B \oplus B^d\).

Every band in a Dedekind complete Riesz space is a projection band. If \(B\) is a projection band of a Riesz space \(X\) and \(x \in X\), then \(x = x_1 + x_2\), where \(x_1 \in B\) and \(x_2 \in B^d\). In this case, the map \(P_B : X \to X\) defined by \(P_B(x) = x_1\) is called the **band projection** associated with \(B\).

An additive group \(G\), equipped with a partial order \(\leq\) is called a **partially ordered group** if its group operation is compatible with the order structure in a manner such that

(a) if \(x \leq y\), then \(x + z \leq y + z\) for any \(z \in G\);
(b) if \(x \leq y\), then \(z + x \leq z + y\) for any \(z \in G\).

A partially ordered group is said to be a **lattice-ordered group** if it is also a lattice at the same time. Lattice-ordered rings and lattice-ordered fields are defined in a similar manner. Specifically, a ring \(R\) equipped with a partial order \(\leq\) is called a **partially ordered ring** if its ring operations are compatible with the order structure in a manner such that
(a) if $x \leq y$, then $x + z \leq y + z$ for any $z \in R$;
(b) if $x \leq y$ and $z > 0$, then $xz \leq yz$ and $zx \leq yz$.

A partially ordered ring is said to be a lattice-ordered ring if its additive group is a lattice-ordered group. Likewise, a field $F$ equipped with a partial order $\leq$ is called a partially ordered field if its field operations are compatible with the order structure in a manner such that

(a) if $x \leq y$, then $x + z \leq y + z$ for any $z \in F$;
(b) if $x > y$ and $z > 0$, then $xz > yz$;
(c) $1 > 0$, where $1$ is the identity element of $F$.

A partially ordered field is said to be a lattice-ordered field if its additive group is a lattice-ordered group.

A linear operator $T$ on a vector space $X$ is said to be a projection if $T^2 = T$. A linear operator $T$ between two Riesz spaces $X$ and $Y$ is said to be positive if $x \in X$ and $x \geq 0$ implies $T(x) \geq 0$; $T$ is said to be strictly positive if $x \in X$ and $x > 0$ implies $T(x) > 0$; $T$ is said to be order-continuous if $x \xrightarrow{o} 0$ in $X$ implies $T(x) \xrightarrow{o} 0$ in $Y$. A linear operator $T$ on a vector space $V$ is called an averaging operator if $T(yT(x)) = T(y)T(x)$ for any pair $x, y \in V$ such that $yT(x) \in V$.

Following [17] and [20], we say a positive order-continuous projection on a Riesz space $X$ with a weak order unit $e$ is a conditional expectation if the range $R(T)$ of $T$ is a Dedekind complete Riesz subspace of $X$ and $T(e)$ is a weak order unit of $X$ for every weak order unit $e$ of $X$. If $T$ is a conditional expectation on a Riesz space $X$, then there is a weak order unit $e$ such that $T(e) = e$.

3. Conditional probability in Riesz spaces

To extend the notion of conditional probability in probability theory to Riesz spaces, we first recall the following definition.

Definition 3.1. Let $(\Omega, \mathcal{F}, P)$ be a probability space. Suppose $A$ and $B$ be two events such that $P(B) > 0$. Then the conditional probability $P(A \mid B)$ of $A$ given $B$ is defined as

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}.$$

Intuitively, $P(A \mid B)$ measures the relative size of the projection of $A$ onto $B$. This motivates the following definition.

Definition 3.2. Let $E$ be a Dedekind complete lattice-ordered field with a conditional expectation $T$ and a weak unit $e$ such that $T(e) = e$. Suppose $B_1$ and $B_2$ are two projection bands, $P_{B_1}$ and $P_{B_2}$ are the corresponding band projections of $B_1$ and $B_2$, respectively, and $B_1 \cap B_2 \neq \phi$. Then the conditional probability $P_{B_1 \mid B_2}$ of $B_1$ given $B_2$ with respect to $T$ is defined as

$$P(B_1 \mid B_2)(f) = [TP_{B_1}P_{B_1}(f)][TP_{B_2}(f)]^{-1}, \quad \forall f \in E.$$  

provided $TP_{B_1}(f) \neq 0$.  

Remark. We require $E$ to be a lattice-ordered field because we need the multiplication operation on $E$ to be invertible. It is evident that if $P_{B_1} = P_{B_2}$ then $P_{B_1|B_2} = I$, where $I$ is the identity operator on $E$. Since two band projections on a Riesz space are commutative, Equation (3.1) may be replaced by

$$P(B_1 | B_2)(f) = [TP_{B_1}P_{B_2}(f)] [TP_{B_2}(f)]^{-1}, \quad \forall f \in E,$$

provided $TP_{B_2}(f) \neq 0$.

The following example shows that the classical definition of conditional probability is a special case of Definition 2.1.

Example 2.1. Let $E$ be the family of all random variables with finite mean on a probability space $(\Omega, \mathcal{F}, P)$, that is, $E = L^1(\Omega, \mathcal{F}, P)$. Then $e = 1_\Omega$, where $1_A$ denotes the indicator function of an event $A$. Let $T$ be the expectation operator. For two events $A, B \in \mathcal{F}$ with $B \neq \emptyset$, the bands generated by $1_A$ and $1_B$ are $B_1 = L^1(A, \mathcal{F} \cap A, P)$ and $B_2 = L^1(B, \mathcal{F} \cap B, P)$, respectively. Therefore, the corresponding band projections $P_{B_1}$ and $P_{B_2}$ are defined as $P_{B_1}(f) = f1_A$ and $P_{B_2}(f) = f1_B$. We have

$$TP_{B_2}(e) = E[1_\Omega 1_B] = P(B),$$

and

$$TP_{B_2}P_{B_1}(e) = E[1_\Omega 1_B 1_A] = P(A \cap B).$$

It follows that $P(B_1 | B_2) = P(A | B)$ which is consistent with the classical definition of conditional probability of $A$ given $B$.

Recall that in probability theory two events $A$ and $B$ are defined to be independent if $P(A | B)$ does not depend on $B$ and $P(B | A)$ does not depend on $A$, that is $P(A | B) = P(A)$ and $P(B | A) = P(B)$. To accommodate the case $P(B) = 0$, we usually define two events $A$ and $B$ to be independent if $P(A \cap B) = P(A)P(B)$. Thus, Definition 3.2 leads naturally to the following concept of independence in Riesz spaces.

Definition 3.3. Let $E$ be a Dedekind complete lattice-ordered ring with a conditional expectation $T$ and a weak unit $e$ such that $T(e) = e$ and the multiplication operation coincide with the lattice operation $\land$. Suppose $B_1$ and $B_2$ are two projection bands of $E$ and $P_{B_1}$ and $P_{B_2}$ are the associated band projections of $B_1$ and $B_2$, respectively. We say $B_1$ and $B_2$ are independent with respect to $T$ if

$$TP_{B_1}P_{B_2}(e) = TP_{B_2}(e)TP_{B_1}(e).$$

Remark. In Definition 3.3, we assume that $E$ is a Dedekind complete lattice-ordered ring and its multiplication is the lattice operation $\land$. Definition 4.1 in [22] defines two projection bands $P_{B_1}$ and $P_{B_2}$ to be independent with respect to a conditional expectation $T$ in a Dedekind complete Riesz space if

$$TP_{B_1}TP_{B_2}(e) = TP_{B_1}P_{B_2}(e) = TP_{B_2}TP_{B_1}(e).$$
As [23] pointed out, the space of $e$ bounded elements in $E$ will admit a natural $f$-algebra structure if we set $P(e)Q(e) = PQ(e)$ for two band projections $P$ and $Q$. In view of this, our definition and Definition 4.1 in [22] are consistent.

4. Law of total probability and Bayes’ theorem in Riesz spaces

In probability theory, the law of total probability and Bayes’ theorem are two fundamental theorems involving conditional probability. In this section, we will extend these two theorems to Riesz spaces. We will also give examples to show that the classical law of total probability and Bayes’ theorem are special cases of law of total probability and Bayes’ theorem in Riesz spaces. To motivate our study, we first recall the law of total probability in probability theory:

**Theorem 4.1.** Suppose $B_1, ..., B_n$ are mutually disjoint events on a probability space $(\Omega, \mathcal{F}, P)$ such that $\Omega = \cup_{i=1}^{n} B_i$ and $P(B_i) > 0$ for each $1 \leq i \leq n$. Then for any event $A$

$$P(A) = \sum_{i=1}^{n} P(A \mid B_i)P(B_i).$$

Since $P(A \mid B_i)P(B_i) = P(A \cap B_i)$, the intuition of law of total probability can be described as follows: If an event $A$ is the union of $n$ mutually disjoint sub-events $A \cap B_i$, then the projection onto $A$ is equivalent to the sum of all projections onto $A \cap B_i$’s, that is,

$$P(A) = \sum_{i=1}^{n} P(A \cap B_i).$$

This intuition extends naturally to the abstract case of Riesz spaces as the next theorem shows. To this end, we first state and prove an inclusion-exclusion formula for projection bands; this result extends the classical inclusion-exclusion theorem for events to Riesz spaces.

**Lemma 4.1 (Inclusion-exclusion formula for projection bands).** Let $B_1, ..., B_n$ be $n$ projection bands in a Riesz space. Then $B_1 + ... + B_n$ is a projection band and

$$P_{B_1 + ... + B_n} = \sum_{i=1}^{n} P_{B_i} - \sum_{1 \leq i_1 < i_2 \leq n} P_{B_{i_1} \cap B_{i_2}} + ... + (-1)^{k-1} \sum_{1 \leq i_1 < ... < i_k \leq n} P_{B_{i_1} \cap ... \cap B_{i_k}} + ... + (-1)^{n-1} P_{B_1 \cap ... \cap B_n}.$$  \(4.1\)

**Proof.** The fact that $B_1 + ... + B_n$ is a projection band is trivial. We show Equation (4.1) by induction on $n$. The case $n = 2$ follows from Theorem 1.45 of
[1]. Suppose Equation (4.1) holds for \( n - 1 \). Then the induction hypothesis implies

\[
P_{B_1 + ... + B_n} = P_{B_1 + ... + B_{n-1}} + P_{B_n} - P_{B_1 + ... + B_{n-1}} P_{B_n}
\]

\[
= \sum_{i=1}^{n-1} P_{B_i} - \sum_{1 \leq i_1 < i_2 \leq n} P_{B_{i_1} \cap B_{i_2}} + ... \\
+(-1)^{k-1} \sum_{1 \leq i_1 < ... < i_k \leq n} P_{B_{i_1} \cap ... \cap B_{i_k}} + ... + (-1)^{n-2} P_{B_1 \cap ... \cap B_{n-1}} \\
+ P_{B_n} - P_{B_1 + ... + B_{n-1}} P_{B_n}
\]

\[
= \sum_{i=1}^{n-1} P_{B_i} - \sum_{1 \leq i_1 < i_2 \leq n} P_{B_{i_1} \cap B_{i_2}} + ... \\
+(-1)^{k-1} \sum_{1 \leq i_1 < ... < i_k \leq n} P_{B_{i_1} \cap ... \cap B_{i_k}} + ... + (-1)^{n-2} P_{B_1 \cap ... \cap B_{n-1}} \\
+ (-1)^{k-2} \sum_{1 \leq i_1 < ... < i_{k-1} \leq n} P_{B_{i_1} \cap ... \cap B_{i_{k-1}}} + ... + (-1)^{n-2} P_{B_1 \cap ... \cap B_{n-1}} \left[ P_{B_n} \right]
\]

\[
= \sum_{i=1}^{n} P_{B_i} + P_{B_n} \\
- \left( \sum_{i=1}^{n} P_{B_i} \right) P_{B_n} - \sum_{1 \leq i_1 < i_2 \leq n} P_{B_{i_1} \cap B_{i_2}} \\
+ ... + \\
(-1)^{k-1} \sum_{1 \leq i_1 < ... < i_k \leq n} P_{B_{i_1} \cap ... \cap B_{i_k}} + (-1)^{k-1} \sum_{1 \leq i_1 < ... < i_{k-1} \leq n} P_{B_{i_1} \cap ... \cap B_{i_{k-1}}}
\]

\[
+ ... + \\
+ (-1)^{n-1} P_{B_1 \cap ... \cap B_{n-1}} P_{B_n}
\]

\[
= \sum_{i=1}^{n} P_{B_i} - \sum_{1 \leq i_1 < i_2 \leq n} P_{B_{i_1} \cap B_{i_2}} \\
+ ... + (-1)^{k-1} \sum_{1 \leq i_1 < ... < i_k \leq n} P_{B_{i_1} \cap ... \cap B_{i_k}} \\
+ ... + (-1)^{n-1} P_{B_1 \cap ... \cap B_n}
\]

By mathematical induction, (4.1) holds for all positive integer \( n \).

\[ \square \]

**Theorem 4.2** (Law of total probability in Riesz spaces). Let \( E \) be a Dedekind complete lattice-ordered field with a conditional expectation \( T \) and a weak unit \( e \) such that \( T(e) = e \). Suppose \( B_1, ..., B_n \) are \( n \) mutually disjoint nonempty projection bands such that \( E = B_1 \oplus B_2 \oplus ... \oplus B_n \) and \( P_{B_i} \) is the band projection onto \( B_i \).
Then for any projection band \( D \) and its associated band projection \( P_D \)

\[
TP_D(f) = \sum_{i=1}^{n} (P_{D|B_i}(f))(TP_{B_i}(f)), \quad \forall f \in E.
\]

provided \( TP_{B_i}(f) \neq 0 \) for all \( 1 \leq i \leq n \).

**Proof.** Definition 3.2 shows that the right-hand side of Equation (4.2) equals

\[
\sum_{i=1}^{n} T(P_{D|B_i}) = T \left[ \sum_{i=1}^{n} (P_{D\cap B_i}) \right]
\]

Since \( B_i \)'s are mutually disjoint, \( P_{D\cap B_i} \cap P_{D\cap B_j} = 0 \) for \( i \neq j \). It follows from Lemma 4.1 that

\[
T \left[ \sum_{i=1}^{n} (P_{D\cap B_i}) \right] = T \left[ (P_{D\cap (\bigcup_{i=1}^{n} B_i)}) \right]
= TP_D.
\]

Therefore, the theorem is established. \( \square \)

Theorem 4.1 and Definition 3.2 immediately imply the following Bayes’ Theorem in Riesz spaces.

**Corollary 4.1 (Bayes’ Theorem in Riesz spaces).** Let \( E \) be a Dedekind complete lattice-ordered field with a conditional expectation \( T \) and a weak unit \( e \) such that \( Te = e \). Suppose \( B_1, \ldots, B_n \) are mutually disjoint nonempty projection bands such that \( E = B_1 \oplus B_2 \oplus \ldots \oplus B_n \) and \( P_{B_i} \) is the band projection onto \( B_i (1 \leq i \leq n) \). Then for any projection band \( D \) and any \( 1 \leq j \leq n \)

\[
P_{B_j|D}(f) = \left[ (P_{D|B_j}(f))(TP_{B_j}(f)) \right] \left[ \sum_{i=1}^{n} (P_{D|B_i}(f))(TP_{B_i}(f)) \right]^{-1}, \quad \forall f \in E,
\]

provided \( TP_{B_i}(f) \neq 0 \) for all \( 1 \leq i \leq n \).

**Example 3.1.** Take \( E = L^1(\Omega, F, \mathbb{P}) \). Then \( e = 1_\Omega \). Suppose \( T \) is the expectation operator and \( P_{B_i} \)'s are defined as \( P_{B_i}(f) = 1_{B_i}f \) for \( f \in E \), where \( B_1, B_2, \ldots, B_n \) are \( n \) disjoint nonempty events such that \( \Omega = \bigcup_{i=1}^{n} B_i \). Also, for any event \( D \) define \( P_D(f) = 1_Df \) for \( f \in E \). Then \( TP_D(e) = P(D), TP_{B_i}(e) = P(B_i) \) and \( P_{D|B_i}(e) = P(D \mid B_i) \); thus, Theorem 4.2 specializes to Theorem 4.1. Moreover, we have \( P_{B_1 + \ldots + B_n}(e) = P(\bigcup_{i=1}^{n} B_i), P_{B_i}(e) = P(B_i), P_{B_i \cap B_j}(e) = P(B_i \cap B_j), \ldots \); hence Lemma 4.1 specializes to the classical inclusion-exclusion formula. Finally, in this case Corollary 4.1 yields

\[
P(B_j \mid D) = \frac{P(D \mid B_j)P(B_j)}{\sum_{i=1}^{n} P(D \mid B_i)P(B_i)},
\]

therefore, Corollary 4.1 specializes to the classical Bayes’ theorem.
Acknowledgements

I thank Coenraad Labuschagne and George Stoica for providing me with some references.

References


Department of Mathematics, Robert Morris University, Moon, PA 15108, USA
E-mail address: hong@rmu.edu