Operations Research Letters 45 (2017) 242–247

Contents lists available at ScienceDirect

Operations Research Letters

journal homepage: www.elsevier.com/locate/orl

Computing the probability of union in the n-dimensional Euclidean space for application of the multivariate quantile: p-level efficient points

Jinwook Lee *

Decision Sciences and MIS, LeBow College of Business, Drexel University, Philadelphia, PA 19104, United States

ARTICLE INFO

Article history:
Received 15 February 2017
Accepted 11 March 2017
Available online 22 March 2017

Keywords:
Probability of union in R^n
Binomial moments
Multivariate quantiles
p-efficient points (Multivariate Value-at-Risk)

ABSTRACT

As recently studied in Lee and Prékopa (2017), in the general case of $R^n, n \geq 3$, if certain conditions among the projected sets are met, the probability of union of such sets can be efficiently calculated by the modified inclusion–exclusion formula in a recursive manner. In this paper, we study with the fullest generality how to construct such suitable conditions on the partial order relation. Upon this basis we can use the modified inclusion–exclusion formulas in any case. Numerical examples are presented.

1. Introduction

Probabilities of the type $P(A_{i_1} \dots A_{i_r})$ are called r-order probabilities and if only i-order probabilities are used in the calculation or bounding where $1 \leq i \leq r$, then we call it of order r. Computing and bounding of probabilities of Boolean functions of events, repeated by sets in R^n, is typically done in such a way that we compute low order probabilities and infer to higher order ones. This is because the calculation of the binomial moments gets very complicated as the number of events increases (except for a couple of lowest and highest moments).

Recently, [7] has discovered an efficient way for the calculation of the probability of union in the n-dimensional Euclidean space. In the bivariate case, the new formula, called "the modified inclusion–exclusion formula", makes it possible to efficiently find the probability of union as that only requires the first binomial moment and the incomparable pairwise intersections. For $R^n, n \geq 3$, however, we need a special condition on the subset relationship among the projections onto lower dimensional spaces in order to use the method employed in the bivariate case.

The above mentioned condition on the partial order relation is crucial for the new formulas of [7] in $R^n, n \geq 3$. The main purpose of this paper is to show that there exists a way for such a partial order relation construction, based on which we present the new formulas in the fullest generality. We also study duality theorems in connection with system of distinct representative (SDR) of our setting. It is strongly recommended for the readers to see [7] (and references therein, e.g., [1,4,5,8], etc.) for more detailed background of this paper.

Our events in this paper are orthants in the n-space, designated by

$$A(z^{(1)}), \dots, A(z^{(N)}), \tag{1}$$

where $z^{(1)}, \dots, z^{(N)}$ are the vertices of the orthants ($z^{(i)} \in R^n, i = 1, \dots, N$). We assume that $z^{(1)}, \dots, z^{(N)}$ is an antichain in the partially ordered set R^n, i.e., for no i, j ($i \neq j$) do we have $z^{(i)} \leq z^{(j)}$. An important example for such sets in connection with stochastic programming is the multivariate quantiles, known as the p-efficient points of a discrete distribution (or a discretized continuous distribution), introduced by [10] as the following:

Definition 1. The point $z \in Z$ is a p-efficient point of $\xi \in R^n$, or its distribution, if $F(z) \geq p$ and there is no $y \leq z, y \neq z$ such that $F(y) \geq p$, where F is the joint c.d.f. of $\xi \in R^n$.

If $\{z^{(1)}, \dots, z^{(N)}\}$, where $z^{(i)} \in R^n, i = 1, \dots, N$ is the set of p-efficient points of ξ, then it is obviously an antichain in R^n and typically $N \gg n$.

It is well known that the set of p-efficient points is a multivariate counterpart of Value-at-Risk (VaR) which is in fact, a univariate quantile. For this reason, the set of p-efficient points is called Multivariate Value-at-Risk (MVaR) and we refer the reader to [13,6] for more details.

* Fax: +1 215 895 2907.
E-mail address: jinwook.lee@drexel.edu.

http://dx.doi.org/10.1016/j.orl.2017.03.007
0167-6377/

2. Functional representation of the set $\{z^{(1)}, \ldots, z^{(N)}\}$

Suppose there are N sets, $A_i = A(z^{(i)}), i = 1, \ldots, N$. Let $A(z) = \{v \mid v \leq z\} \subset R^n$. Then $A(z) \supset \Pi_{i=1}^{n} A(z^{(i)})$ if and only if $z \geq (\min_i(z_1^{(i)}), \ldots, \min_i(z_n^{(i)})), i = 1, \ldots, N$.

Theorem 1. *If we have $A_1, \ldots, A_N \subset R^n$ with $N > n$, then at least one of them contains the intersection of the others, i.e., for at least one i,*

$$A_i \supset \Pi_{j \in \{1,\ldots,N\} \setminus \{i\}} A_j.$$

Proof. The smallest one out of N orthants in R^n can be represented by at most n orthants. But we have $N > n$. □

Theorem 2. *If we have $A_1, \ldots, A_N \subset R^n$ with $N > n$, then the inclusion–exclusion formula for the union of them will have the following terms only: singles, pairs, triples, ..., n-tuples, but not the entire members of their family.*

Proof. Any intersection of $A_1, \ldots, A_N \subset R^n$ with $N > n$ can be expressed by n-tuples at most, i.e., there exists j such that $\Pi_{j \in \{1,\ldots,n\}} A_j = \Pi_{i \in \{1,\ldots,N\}} A_i$. □

Definition 2 (*System of Distinct Representatives (SDR)*)**.** Suppose that $A_1, A_2, \ldots, A_N$ are sets. The family of sets $A_1, A_2, \ldots, A_N$ has a system of distinct representatives (SDR) if and only if there exist distinct elements $z^{(1)}, z^{(2)}, \ldots, z^{(N)}$ such that $z^{(i)} \in A_i$ for each $i = 1, \ldots, N$.

Theorem 3 (*Duality*)**.** *The minimum number of non-redundant events in $A_{i_1}, \ldots, A_{i_r}$ is equal to the maximum number of distinct representatives in $z^{(i_1)}, \ldots, z^{(i_r)}$.*

$\{A_1, A_2, \ldots, A_N\}$ is called a discrete convex set if none of the points $z^{(1)}, \ldots, z^{(N)}$ is in the relative interior of the convex hull of $\{z^{(1)}, \ldots, z^{(N)}\}$. Note that we can write, for the ith component of a vector $z \in R^n, z_i = z_i(z_1, \ldots, z_{i-1}, z_{i+1}, \ldots, z_n), i = 1, \ldots, n$. Also note that the sets $(z_1, \ldots, z_n)$ and $(z_1, \ldots, z_{i-1}, z_i(z_1, \ldots, z_{i-1}, z_{i+1}, \ldots, z_n), z_{i+1}, \ldots, z_n)$ are the same.

Theorem 4 (*Second Duality Theorem for The Case of a Convex Set $\{z^{(1)}, \ldots, z^{(N)}\}$*)**.** *Consider $z_n = z_n(z_1, \ldots, z_{n-1})$, or any other subscript instead of n. Consider the projection of the set $\{z^{(1)}, \ldots, z^{(N)}\}$ onto the space of $z_1, \ldots, z_{n-1}$. Then in $A_{i_1}, \ldots, A_{i_r}$ the event i_j is redundant if and only if $z^{(i_j)} \leq \sum_{k=1}^{r} \lambda_k z^{(i_k)}$, without the component $z_n^{(i_k)} = z_n^{(i_k)}(z_1^{(i_k)}, \ldots, z_{n-1}^{(i_k)})$, where $\lambda_k \geq 0$, $k = 1, \ldots, r$ and $\sum_{k=1}^{r} \lambda_k = 1$.*

We can fully eliminate the redundant events by the use of the first duality theorem. The second theorem helps because we can look at sets in the space of $z_1, \ldots, z_{n-1}$ and if a z is inside the convex hull of such a set in the $n-1$-space, then we can eliminate it.

Example 1. We consider a SDR (System of Distinct Representative) of the set $\{z^{(1)}, \ldots, z^{(5)}\}$, where $z^{(i)} \in R^3, i = 1, \ldots, 5$ and their related sets (orthants) in R^3: $A_1 = A(z^{(1)}), A_2 = A(z^{(2)}), A_3 = A(z^{(3)}), A_4 = A(z^{(4)}), A_5 = A(z^{(5)})$. We use the following notations: $p_i = P(A_i),\ i = 1, \ldots, 5$. $p_{ij} = P(A_i A_j)$ for $1 \leq i < j \leq 5$, $p_{ijk} = P(A_i A_j A_k)$ for $1 \leq i < j < k \leq 5$, $p_{ijkl} = P(A_i A_j A_k A_l)$ for $1 \leq i < j < k < l \leq 5$ and $p_{12345} = A_1 A_2 A_3 A_4 A_5$.

Let $z^{(1)} = (1, 1, 3), z^{(2)} = (3, 1, 2), z^{(3)} = (2, 2, 2), z^{(4)} = (1, 3, 2), z^{(5)} = (3, 3, 1)$. Note that any system of distinct representative can be used. Fig. 1 describes their projected points onto (z_1, z_2) space. Then we can write

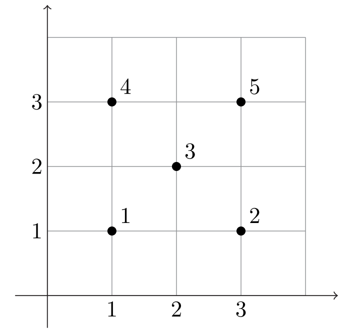

Fig. 1. Illustration of five incomparable elements $(z_1^{(i)}, z_2^{(i)}, z_3^{(i)}) \in R^3, i = 1, \ldots, 5$, projected onto (z_1, z_2)-space. Numbers next to the points mean i of A_i, $i = 1, \ldots, 5$.

$$\begin{aligned}
P\left(\bigcup_{i=1}^{5} A_i\right) &= \sum_{i=1}^{5} p_i - \sum_{i<j} p_{ij} + \sum_{i<j<k} p_{ijk} - \sum_{i<j<k<l} p_{ijkl} + p_{12345} \\
&= p_1 + p_2 + p_3 + p_4 + p_5 - p_{12} - p_{13} - p_{14} - p_{15} \\
&\quad - p_{23} - p_{24} - p_{25} - p_{34} - p_{35} - p_{45} \\
&\quad + p_{123} + p_{124} + p_{125} + p_{134} + p_{135} + p_{145} \\
&\quad + p_{234} + p_{235} + p_{245} + p_{345} \\
&\quad - p_{1234} - p_{1235} - p_{1245} - p_{1345} - p_{2345} + p_{12345} \\
&= p_1 + p_2 + p_3 + p_4 + p_5 \\
&\quad - p_{12} - p_{13} - p_{14} - p_{23} - p_{25} - p_{34} - p_{35} - p_{45} \\
&\quad + p_{123} + p_{134} + p_{235} + p_{345},
\end{aligned} \tag{2}$$

since $p_{135} = p_{15}, p_{124} = p_{1234}, p_{234} = p_{24}, p_{1235} = p_{125}, p_{1245} = p_{12345}, p_{1345} = p_{145}, p_{2345} = p_{245}$.

Note that in the last equation of (2) we have:

- all individual probabilities
- all neighboring pairs (in the (z_1, z_2) space)
- all triples without additional points in them (in the (z_1, z_2) space), i.e., no points inside a triangle.

Also note that we need, at most, triples for the calculation as Theorem 2 says.

For the general case, the method is to eliminate the redundant events from all $A_{i_1}, \ldots, A_{i_k}$ by the use of the system of distinct representatives. For the convex case we look inside the simplices, 2-dim, 3-dim, ..., n-dim and remove the subscript from all $p_{ij\ldots l}$ where some of the events form a simplex and the others are inside. This step applied for $(\{i, j, \ldots, l\}) \leq n$. See below for each case.

Algorithm (*General Case*)**.**

Step 1. Enumerate pairs, triples, ...
Step 2. In each $A_i A_j \ldots A_l$ find maximum number of distinct representatives.
Step 3. In each $A_i A_j \ldots A_l$ keep those which contain representatives; eliminate the others.
Step 4. Use the terms in the inclusion–exclusion formula to summarize the result.

Algorithm (*Convex Case*)**.**

Step 1. Enumerate pairs, triples, ...
Step 2. In each $A_i A_j \ldots A_l$ take the simplex and eliminate the insiders.
Step 3. In each $A_i A_j \ldots A_l$ find the maximum set of distinct representatives, eliminate those which do not contain distinct representatives.