Available online at www.sciencedirect.com
ScienceDirect
Statistics & Probability Letters 78 (2008) 311–319
STATISTICS & PROBABILITY LETTERS
www.elsevier.com/locate/stapro

Fréchet optimal bounds on the probability of a union with supplementary information

Fred M. Hoppe[a,*,1], Mikhail Nediak[b,2]

[a]Department of Mathematics and Statistics, McMaster University, 1280 Main St. W., Hamilton, Ont., Canada L8S 4K1
[b]Queen's School of Business, Goodes Hall, Queen's University, 143 Union St., Kingston, Ont., Canada K7L 3N6

Received 14 June 2006; accepted 10 February 2007
Available online 19 August 2007

Abstract

We show that bounds for the probability of a union involving either lower order binomial moments or lower order probabilities of events may be considerably improved in the presence of supplementary information such as a bound on the number of events that can occur simultaneously or bounds on their probabilities. An example shows how such additional information often may be provided naturally. We also prove the Fréchet optimality of these bounds using linear programming.

MSC: primary 60C05; secondary 90C05

Keywords: Boole's inequality; Degree one; Fréchet optimal; Linear programming; Probability bound; Supplementary information

1. Introduction

Let $\{A_i,\ i=1,\ldots,n\}$ be a collection of events on some probability space and denote $S_1=\sum_{i=1}^n P(A_i)$. The degree one lower probability bound

$$P\left(\bigcup_{i=1}^n A_i\right) \geqslant \frac{S_1}{n} \tag{1}$$

is optimal in the sense of Fréchet (1935), (see also Seneta, 1992, Section 7) meaning that it can be achieved on some probability space. This paper considers bounds under restrictions on the number of events $v=\{\#i : A_i \text{ occurs}\}$. In Proposition 1, assuming that $P(v\geqslant k+1)\leqslant B$ for some $1\leqslant k\leqslant n$, we prove

$$P\left(\bigcup_{i=1}^n A_i\right) \geqslant \frac{S_1}{k} - \left(\frac{n}{k}-1\right)\min\left\{B, \frac{S_1}{n}\right\},$$

*Corresponding author. Tel.: +1 905 529 7070; fax: +1 905 522 0935.
E-mail addresses: hoppe@mcmaster.ca (F.M. Hoppe), mnediak@business.queensu.ca (M. Nediak).
[1]Supported by NSERC Discovery Grant.
[2]Supported by SHARCnet Postdoctoral Fellowship at McMaster University.

0167-7152/$ - see front matter
doi:10.1016/j.spl.2007.02.019

which, in the special case $B=0$, corresponding to $P(v \leqslant k)=1$ as in Example 1 below, reduces to the pleasant form

$$P\left(\bigcup_{i=1}^{n} A_i\right) \geqslant \frac{S_1}{k} \tag{2}$$

paralleling (1). Then in Proposition 2, assuming $P(0<v\leqslant k-1)\leqslant C$, we show that Boole's degree one upper bound

$$P\left(\bigcup_{i=1}^{n} A_i\right) \leqslant \min\{1, S_1\} \tag{3}$$

can be improved to

$$P\left(\bigcup_{i=1}^{n} A_i\right) \leqslant \min\left\{1,\ \frac{S_1}{k}+\left(1-\frac{1}{k}\right)\min\{C, S_1\}\right\}.$$

Finally, Proposition 3 considers improvements on bounds that involve the individual $\{P(A_i), 1\leqslant i\leqslant n\}$. Situations where $v\leqslant k$ do occur naturally.

Example 1. Consider the rectangular array $\{A_{ij},\ i=1,\dots,m,\ j=1,\dots,k\}$ of disjoint families of events $\{A_{i1}, A_{i2},\dots,A_{ik}\}$, $i=1,\dots,m : (\bigcup_{j=1}^{k} A_{ij})\cap(\bigcup_{j=1}^{k} A_{i'j})=\emptyset$, whenever $i\neq i'$. Then, clearly, $v\leqslant k$ and, moreover,

$$\begin{aligned} P\left(\bigcup_{i,j} A_{ij}\right) &= \sum_{i=1}^{m} P\left(\bigcup_{j=1}^{k} A_{ij}\right) \geqslant \sum_{i=1}^{m} \frac{1}{k}\sum_{j=1}^{k} P(A_{ij}) \quad \text{from (1)} \\ &= \frac{1}{k}\sum_{i=1}^{m}\sum_{j=1}^{k} P(A_{ij}) \equiv \frac{S_1}{k}, \end{aligned}$$

an improvement over (1). Of course, this example is trivial since it involves the disjoint families of events. However, it does give insight into how the denominator can be improved in (1) from n to k.

Example 2. Consider a finite sample space on N outcomes each having equal probability. Let $\{A_i,\ i=1,\dots n\}$ be a collection of distinct two-point events, call them edges. Events intersect at a point and there are at most $N-1$ events (edges) that can intersect at any point (vertex). Therefore, $v\leqslant N-1$, while n can be as large as $N(N-1)/2$.

2. Improvements to bounds of degree one

The notion of optimality adopted is from Fréchet (1935) (see Chen, 1998; Seneta and Chen, 1996, 2002):

Definition 1. Let $\mathscr{L}(S_1, n)$ be the class of functions $g(s)$ such that for any probability space and any collection of events $\{A_1,\dots,A_n\}$

$$P\left(\bigcup_{i=1}^{n} A_i\right) \geqslant g(S_1).$$

Suppose that $\bar{g}\in\mathscr{L}(S_1,n)$ is such that for any collection $\{A_1,\dots,A_n\}$, there exists a collection $\{A_1^*,\dots,A_n^*\}$, possibly on another probability space, for which the value $S_1^*\equiv\sum_{i=1}^{n} P(A_i^*)$ is the same as $S_1\equiv\sum_{i=1}^{n} P(A_i)$, and such that the following equality holds:

$$P\left(\bigcup_{i=1}^{n} A_i^*\right) = \bar{g}(S_1^*), \tag{4}$$

then $\bar{g}$ is said to be a Fréchet optimal lower bound of degree one.

We will also use the term Fréchet optimal to refer to any bound that can be achieved subject to additional restrictions on the probabilities.

Remark 1. The reason for the terminology "optimal" in Definition 1 is derived from the fact that $\bar{g}$ satisfies $\bar{g} \geqslant g$ for all $g \in \mathscr{L}(S_1, n)$, so that $\bar{g}$ is the best of all such lower bounds. This follows from

$$\bar{g}(S_1) = \bar{g}(S_1^*) = P\left(\bigcup_{i=1}^{n} A_i^*\right) \geqslant g(S_1^*) = g(S_1)$$

for any $g \in \mathscr{L}(S_1, n)$. One way to show Fréchet optimality of a bound is to describe the n events $\{A_1^*, \ldots, A_n^*\}$ in (4). When $B = 0$, if the $\{A_i^*\}$ are such that the only non-zero probabilities are given by

$$P(A_1^* A_2^* \ldots A_k^* \bar{A}_{k+1}^* \ldots \bar{A}_n^*) = \frac{S_1}{k},$$

$$P(\bar{A}_1^* \bar{A}_2^* \ldots \bar{A}_n^*) = 1 - \frac{S_1}{k},$$

then $P(A_i^*) = S_1/k,\ 1 \leqslant i \leqslant k,\ P(A_i^*) = 0,\ k+1 \leqslant i \leqslant n$, so that $\sum_{i=1}^{n} P(A_i^*) = kS_1/k = S_1$ and $P(\bigcup_{i=1}^{n} A_i^*) = P(A_1^* A_2^* \ldots A_k^* \bar{A}_{k+1}^* \ldots \bar{A}_n^*) = S_1/k$, satisfying (4). In fact the original derivation was a completely probabilistic proof of (2) followed by this argument. However, $B \neq 0$ does not seem amenable to this technique and instead we will approach the general problem through the theory of *linear programming* (see, e.g., Schrijver, 1986), which is a very well-studied class of optimization problems. Linear programming has been used to obtain optimal bounds on various probabilities since the pioneering paper of Hailperin (1965). Its successful applications to probability bounding problems include, among others, work by Dawson and Sankoff (1967), Kounias and Marin (1976), Kwerel (1975), and Prékopa (1988).

An outline of the LP technique in probability bounding is as follows:

1. All given information about the probability measure is described by linear constraints. A bound on the probability of an event is represented as a linear (objective) function.
2. The *linear programming dual* of the resulting optimization problem is obtained, for which any feasible solution, expressed as a function of the given information, is a probability bound (not necessarily optimal).
3. The linear programming *optimality conditions* are used to arrive at the optimal solution. The primal and dual solutions will have equal value due to the *strong duality theorem* of linear programming.
4. Finally, the primal optimal solution represents a measure whose existence is required in the definition of Fréchet optimality.

While it is sometimes possible to skip step 3 in the presentation, it is often essential in arriving at the proof of optimality.

The linear programming formulation begins with the occurrence or non-occurrence of events $\{A_i\}$, the elementary conjunctions. These may be represented by n-dimensional $\{0,1\}$-vectors (denoted by $\mathbf{y} \in \{0,1\}^n$); for instance, for $n = 4$, the event $A_1 \cap \bar{A}_2 \cap A_3 \cap A_4$ is given by the point $\mathbf{y} = (1,0,1,1)$. Let $p(\mathbf{y})$ be the corresponding probability induced on $\{0,1\}^n$ by a probability P. This measure $p(\cdot)$ will determine the probability space needed to establish Fréchet optimality of a bound. In accordance with our definition of random v, $v(\mathbf{y})$ denotes $\{\#i : y_i = 1\}$. Then

$$S_1 = \sum_{\mathbf{y} \in \{0,1\}^n} v(\mathbf{y}) p(\mathbf{y}), \tag{5}$$

since

$$P(A_i) = \sum_{\mathbf{y}: y_i = 1} p(\mathbf{y}), \tag{6}$$

and each $p(\mathbf{y})$ will be counted in S_1 exactly $v(\mathbf{y})$ times (by the number of 1's in $\mathbf{y}$). Any probability measure has to satisfy a normalization condition

$$\sum_{\mathbf{y}\in\{0,1\}^n} p(\mathbf{y}) = 1. \tag{7}$$

The probability of more than k events occurring simultaneously not exceeding B can be written as

$$P(v \geqslant k+1) = \sum_{\mathbf{y}:v(\mathbf{y})>k} p(\mathbf{y}) \leqslant B, \tag{8}$$

and the probability of the union of all n events as

$$P\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{\mathbf{y}:v(\mathbf{y})>0} p(\mathbf{y}). \tag{9}$$

The problem of finding the optimal lower bound then reduces to minimizing (9) over all possible non-negative measures $p(\cdot)$ on $\{0,1\}^n$ subject to constraints (5), (7), and (8). This is a linear programming problem. Constraint (7) is redundant, since it simply forces objective (9) not to exceed 1. Because we are minimizing in Proposition 1, the optimal value of (9) must not exceed 1 as long as the problem data are consistent. We can simplify our problem further if we make the following observation:

Observation 1. *Linear expressions in the constraints and the objective can be rewritten in terms of*

$$\pi_i = \sum_{\mathbf{y}:v(\mathbf{y})=i} p(\mathbf{y})$$

(*the probability that exactly i events occur*), $i = 0, \ldots, n$. *Any assignment of values to* $p(\mathbf{y})$, $\mathbf{y} \in \{0,1\}^n$ *corresponding to the same feasible assignment of* π_i, $i = 0, \ldots, n$ *will result in the same value of the objective.*

The formulation becomes

$$\min \quad \sum_{i=1}^{n} \pi_i \tag{10}$$

$$\text{s.t.} \quad \sum_{i=1}^{n} i\pi_i = S_1, \tag{11}$$

$$\sum_{i=k+1}^{n} \pi_i \leqslant B, \tag{12}$$

$$\pi_i \geqslant 0, \quad i = 0, \ldots, n. \tag{13}$$

Remark 2. A probability measure $p(\mathbf{y})$, $\mathbf{y} \in \{0,1\}^n$, satisfying constraints (5) and (8) exists if and only if there exists an assignment of π_i, $i = 0, \ldots, n$, such that (11)–(13) and $\sum_{i=0}^{n} \pi_i = 1$ hold. Such a measure can be obtained by arbitrarily splitting a value of π_i among $\mathbf{y}$'s such that $v(\mathbf{y}) = i$, $i = 0, \ldots, n$. Therefore, an optimal solution to (10)–(13) immediately provides a Fréchet optimal bound. We now obtain the following.

Proposition 1. *For any probability measure on the space of possible outcomes satisfying* (5) *and* (8) *for some* $1 \leqslant k \leqslant n$

$$P\left(\bigcup_{i=1}^{n} A_i\right) \geqslant \frac{S_1}{k} - \left(\frac{n}{k} - 1\right)\min\left\{B, \frac{S_1}{n}\right\}. \tag{14}$$

The bound can be achieved.

Proof. When $k = n$, condition (8) does not provide additional information since it is automatically satisfied for any $B \geqslant 0$ and (14) reduces to (1), its Fréchet optimality as a special case.

When $1 \leqslant k < n$, we employ the duality theory of linear programming. Introduce dual variables λ and μ for constraints (11) and (12), respectively. The problem dual to (10)–(13) is

$$\max \quad S_1\lambda + B\mu \tag{15}$$
$$\text{s.t.} \quad i\lambda \leqslant 1, \quad i = 1, \ldots, k, \tag{16}$$
$$i\lambda + \mu \leqslant 1, \quad i = k+1, \ldots, n, \tag{17}$$
$$\mu \leqslant 0. \tag{18}$$

Any dual-feasible solution (to the maximization problem (15)–(18)) provides a lower bound on the probability of the union. Consider, for example, a solution such that $\lambda^* = 1/n$, $\mu^* = 0$ when $B \geqslant S_1/n$, and $\lambda^* = 1/k$, $\mu^* = 1 - n/k$ when $B < S_1/n$. The solution is feasible. Indeed, in the first case, $i\lambda^* = i/n \leqslant 1$, $i = 1, \ldots, n$. In the second case, the solution is also feasible since $i\lambda^* = i/k \leqslant 1$, $i = 1, \ldots, k$, and $i\lambda^* + \mu^* = 1 - (n-i)/k \leqslant 1$, $i = k+1, \ldots, n$. Immediately, $S_1\lambda^* + B\mu^* = S_1/k - (n/k - 1)\min\{B, S_1/n\}$.

We now verify optimality using linear programming. This involves the construction of an optimal probability distribution $\boldsymbol{\pi}^*$ (whose existence, by Remark 2, immediately implies the existence of an optimal probability measure $p^*(\cdot)$ required in the definition of Fréchet optimality). The optimality of $\boldsymbol{\pi}^*$ will follow from the strong duality theorem of linear programming as long as $\boldsymbol{\pi}^*$ is (primal) feasible and has the same value as a dual feasible (λ^*, μ^*).

Consider first the case $B \geqslant S_1/n$. Let $\pi_i^* = 0$, $i = 1, \ldots, n-1$, $\pi_n^* = S_1/n$. Primal feasibility holds since $n\pi_n^* = S_1$ and $\pi_n^* \leqslant B$. The values of $\boldsymbol{\pi}^*$ and (λ^*, μ^*) are both equal to S_1/n. Next, suppose $B < S_1/n$ and let $\pi_i^* = 0$, $i \neq k, n$, $\pi_n^* = B$, $\pi_k^* = (1/k)(S_1 - n\pi_n^*) = (1/k)(S_1 - nB)$. Primal feasibility clearly holds. Also, the values of $\boldsymbol{\pi}^*$ and (λ^*, μ^*) are both equal to $S_1/k - (n/k - 1)B$. □

Example 3. Bound (2) can be used in a very natural way. Let $\Omega = \{1, \ldots, N\}$ be a finite sample space on N outcomes and let $\Pi = \{\{i,j\} : 1 \leqslant i < j \leqslant N\}$ be a set of all undirected pairs from Ω. Let the event collection be a subset of n distinct members of $\Pi : E = \{A_1, \ldots, A_n\} \subseteq \Pi$. Observe that the pair (Ω, E) forms an undirected graph with N vertices and n edges. Suppose that each outcome from Ω occurs with the same probability $1/N$. Then $P(A_i) = 2/N$, $i = 1, \ldots, n$ and $S_1 = 2n/N$. The event $A = \bigcup_{i=1}^n A_i$ can be interpreted as "a randomly selected vertex not being isolated". Using (1), its probability can be bounded from below by

$$P(A) \geqslant \frac{1}{n}\frac{2n}{N} = \frac{2}{N}.$$

We can improve on this bound if we observe that the elements of E intersect if and only if they have a vertex in common, and there are at most $k = N - 1$ distinct edges that can be incident to a given vertex. Thus, if $n \geqslant k$, bound (2) provides an immediate improvement by the factor n/k:

$$P(A) \geqslant \frac{1}{k}\frac{2n}{N} = \frac{2}{N}\frac{n}{k}.$$

If, for instance, $n \geqslant (N-1)(\log_2 N + 1)$, then $n/k \geqslant \log_2 N + 1$.

Similarly, Boole's inequality (3) can be improved in the presence of

$$P(0 < v \leqslant k-1) = \sum_{\mathbf{y}:0 < v(\mathbf{y}) < k} p(\mathbf{y}) \leqslant C. \tag{19}$$

Due to Observation 1, the corresponding linear program has the form

$$\max \quad \sum_{i=1}^{n} \pi_i \tag{20}$$
$$\text{s.t.} \quad \sum_{i=1}^{n} i\pi_i = S_1, \tag{21}$$

$$\sum_{i=1}^{k-1} \pi_i \leqslant C, \tag{22}$$

$$\pi_i \geqslant 0, \quad i = 1 \ldots n. \tag{23}$$

To be absolutely precise, the formulation needs to include a normalization condition (7), which, when expressed in terms of the π_i's, assumes the form $\sum_{i=0}^{n} \pi_i = 1$. Because of the non-negativity of π_0, which does not participate in other constraints or the objective, we can also write this as

$$\sum_{i=1}^{n} \pi_i \leqslant 1. \tag{24}$$

Constraint (24) has a simple effect of forcing (20) not to exceed 1. Thus, we should analyze (20)–(23) with understanding that the true Fréchet optimal bound is a minimum of its optimal value and 1.

Proposition 2. *For any probability measure on the space of possible outcomes satisfying* (5) *and* (19) *for some* $2 \leqslant k \leqslant n$

$$P\left(\bigcup_{i=1}^{n} A_i\right) \leqslant \min\left\{1,\ \frac{S_1}{k} + \left(1 - \frac{1}{k}\right)\min\{C, S_1\}\right\}. \tag{25}$$

The bound is Fréchet optimal.

Proof. It is straightforward to verify by using the dual of (20)–(23) that the optimal measure is given by

$$\pi_1^* = \begin{cases} S_1 & \text{if } C \geqslant S_1, \\ C & \text{if } C < S_1, \end{cases}$$

$\pi_k^* = (S_1 - \pi_1^*)/k$, and $\pi_i^* = 0$, $i \neq 1, k$. The optimal value $\pi_1^* + \pi_k^*$ is then equal to $S_1/k + (1 - 1/k)\min\{C, S_1\}$. The statement of the proposition follows. □

Observation 2. *Substitution of* $k = 1$ *in the right-hand side of* (25) *gives Boole's inequality* (3). *Likewise,* $C = 1$ *always gives a valid bound in* (19) *and also leads to Boole's inequality because if* $S_1 > 1$ *then* (25) *becomes*

$$P\left(\bigcup_{i=1}^{n} A_i\right) \leqslant \min\left\{1, \frac{S_1}{k} + 1 - \frac{1}{k}\right\} = 1,$$

while if $S_1 < 1$ *then* (14) *becomes*

$$P\left(\bigcup_{i=1}^{n} A_i\right) \leqslant \min\left\{1,\ \frac{S_1}{k} + S_1 - \frac{S_1}{k}\right\} = S_1$$

again resulting in (3).

3. Lower bound in the case of known $P(A_i)$

Let us suppose now that we know the individual probabilities $\{P(A_i), i = 1, \ldots, n\}$, ordered, without loss of generality, so that $P(A_i) \geqslant P(A_j),\ i < j$. Knowledge of these values generally leads to better bounds. Actually, as our proof below shows, the individual probabilities need not be known, only their maximum. A well-known (optimal) lower bound is (Boole, 1854, pp. 297–300)

$$P\left(\bigcup_{i=1}^{n} A_i\right) \geqslant \max_{i=1,\ldots,n} P(A_i) = P(A_1). \tag{26}$$

In Section 2, we obtained the lower bound (2) when $B = 0$. Combining (2) with (26), we immediately obtain

$$P\left(\bigcup_{i=1}^{n} A_i\right) \geqslant \max\left\{P(A_1),\ \frac{S_1}{k}\right\} \tag{27}$$

which reduces to (26) with $k = n$. It is a surprising fact that, given $\max_i P(A_i)$, a number $k \leqslant n$, and the information that $B = 0$, this bound is Fréchet optimal.

Note that $B = 0$ implies $p(\mathbf{y}) = 0$ for all $\mathbf{y} : v(\mathbf{y}) > k$. Using relations (6) and (9), we can formulate the problem as

$$\min \quad \sum_{\mathbf{y}:0<v(\mathbf{y})\leqslant k} p(\mathbf{y}) \tag{28}$$

$$\text{s.t.} \quad \sum_{\mathbf{y}:y_i=1,\ v(\mathbf{y})\leqslant k} p(\mathbf{y}) = P(A_i), \tag{29}$$

$$p(\mathbf{y}) \geqslant 0, \quad \mathbf{y} \in \{0,1\}^n, \ \ v(\mathbf{y}) \leqslant k. \tag{30}$$

Here, the $\{P(A_i)\}$ are any collection of probabilities whose maximum is the given $P(A_1)$. Their actual individual values will not be relevant. Let λ_i, $i = 1, \ldots, n$, be the dual variables. A linear program, dual to (28)–(30), has the form

$$\max \quad \sum_{i=1}^{n} \lambda_i P(A_i) \tag{31}$$

$$\text{s.t.} \quad \sum_{i:y_i=1} \lambda_i \leqslant 1, \quad \mathbf{y} : v(\mathbf{y}) \leqslant k. \tag{32}$$

When $k = n$, a primal–dual pair of optimal solutions is given by

$$\begin{aligned} p^*(1,0,0,\ldots,0) &= P(A_1) - P(A_2), \\ p^*(1,1,0,\ldots,0) &= P(A_2) - P(A_3), \\ &\vdots \\ p^*(1,1,1,\ldots,1) &= P(A_n), \end{aligned}$$

with $p^*(\mathbf{y}) = 0$ for all other $\mathbf{y}$, and

$$\lambda_1^* = 1, \quad \lambda_i^* = 0, \quad i = 2, \ldots, n.$$

According to the theory of linear programming, a primal–dual pair of feasible solutions is optimal if and only if it satisfies *a complementary slackness condition* (which is equivalent to the condition that both the solutions have equal value). Note that $p^*(\cdot)$ is primal feasible since the only non-zero terms in expression (29) for each of $P(A_i)$'s are

$$p^*(\underbrace{1,\ldots,1,}_{i\text{ elements}}0,\ldots,0) = P(A_i) - P(A_{i+1}), \ldots, p^*(1,\ldots,1) = P(A_n).$$

Note also that $\boldsymbol{\lambda}^*$ is trivially dual feasible. The complementary slackness conditions

$$p^*(\mathbf{y})\left(\sum_{i:y_i=1} \lambda_i^* - 1\right) = 0, \quad \mathbf{y} \in \{0,1\}^n, \ \ v(\mathbf{y}) \leqslant k \tag{33}$$

hold, since $\boldsymbol{\lambda}^*$ satisfies as equality those and only those constraints in (32), corresponding to $\mathbf{y} : y_1 = 1$, and $p^*(\mathbf{y})$ may be non-zero only for $\mathbf{y} : y_1 = 1$. Thus, the pair is optimal with corresponding value of $P(A_1)$.

The probability measure which has just been described corresponds to a situation when $A_1 \supseteq A_2 \supseteq \cdots \supseteq A_n$. Clearly, when $k < n$, this is impossible. It is not obvious how to construct a probability measure for which (27) holds as an equality (which would imply optimality). Before proving the optimality result, we need to establish the following lemma:

Lemma 1. *A polyhedral cone* $Q = \{\mathbf{q} : q_i \leqslant (1/k)\sum_{j=1}^n q_j,\ \mathbf{q} \geqslant 0\}$ *is the conic hull of the vectors in the set* $Y_k = \{\mathbf{y} \in \{0,1\}^n : v(\mathbf{y}) = k\}$.

Proof. Observe that Q is the set of all vectors in $\mathbb{R}^n$ satisfying the system of linear inequalities

$$A\mathbf{q} = \begin{bmatrix} I - \frac{1}{k}E \\ -I \end{bmatrix}\mathbf{q} \leqslant 0,$$

where E is an $n \times n$ matrix of 1's. The extremal rays of Q are contained in the (one-dimensional) linear subspaces of the form $A'\mathbf{q} = 0$, where A' is a rank $n-1$ submatrix of A. Thus, up to a permutation of indices, any $\mathbf{q}$ on the extremal ray must satisfy a system of equations

$$q_i = \frac{1}{k}\sum_{j=1}^n q_j, \quad i = 1, \ldots, l-1,$$

$$q_i = 0, \quad i = l+1, \ldots, n,$$

for some $l = 1, \ldots, n-1$. Denote a common value of $q_1, \ldots, q_{l-1}$, and $(1/k)\sum_{j=1}^n q_j$ as ρ. Then, it follows that, $\rho = (1/k)[(l-1)\rho + q_l]$ and, expressing q_l via ρ, we get $q_l = (k-l+1)\rho$. However, for $\mathbf{q}$ to belong to Q, we must have $0 \leqslant q_l \leqslant \rho = (1/k)\sum_{j=1}^n q_j$. It follows, that $0 \leqslant k-l+1 \leqslant 1$ or, equivalently, $k \leqslant l \leqslant k+1$. If $l = k$, then $q_l = \rho$, and, if $l = k+1$, then $q_l = 0$, thus, both values of l correspond to the same ray of the form

$$(\underbrace{\rho, \ldots, \rho}_{k}, \underbrace{0, \ldots, 0}_{n-k}).$$

Thus, the directions of all extremal rays are given by the vectors in Y_k. Since a pointed polyhedral cone is the conic hull of its extremal rays, the statement of the lemma follows. For the relevant results on polyhedra, see Chapter 8 of Schrijver (1986). Probabilistically, this lemma means that when the vector of probabilities $P(A_i)$, $i = 1, \ldots, n$ belongs to Q, there exists a probability measure such that exactly k events always occur simultaneously.

Proposition 3. *Given* $P(A_1) \equiv \max_i P(A_i)$, *a number* $k \leqslant n$ *with* $B = 0$, *then* (27) *holds and is Fréchet optimal.*

Proof. We start by considering the case when $P(A_1) \leqslant S_1/k$. This condition implies that the value of the bound is S_1/k and that

$$P(A_i) \leqslant \frac{1}{k}\sum_{j=1}^n P(A_j), \quad i = 1, \ldots, n. \tag{34}$$

A dual-feasible solution with the value of S_1/k has the form

$$\lambda_i^* = \frac{1}{k}, \quad i = 1, \ldots, n.$$

To prove optimality, we need to show that there exists a primal-feasible solution that satisfies the complementary slackness conditions (33). Observe that all constraints in (32) corresponding to $\mathbf{y}$ with $0 < v(\mathbf{y}) < k$ are satisfied by $\boldsymbol{\lambda}^*$ as strict inequalities. Thus, the corresponding primal solution must have $p^*(\mathbf{y}) = 0$ for all $\mathbf{y} : 0 < v(\mathbf{y}) < k$. We must therefore show the existence of non-negative $p^*(\cdot)$ such that, from (29),

$$\sum_{\mathbf{y}:y_i=1,\ v(\mathbf{y})=k} p^*(\mathbf{y}) = \sum_{\mathbf{y}:v(\mathbf{y})=k} y_i p^*(\mathbf{y}) = P(A_i), \quad i = 1, \ldots, n.$$

In the above sum, $p^*(\mathbf{y})$'s are just the coefficients in the conic combination representation of the vector of all $P(A_i)$'s via vectors of Y_k set of Lemma 1. However, from (34), it follows that the vector of $P(A_i)$'s belongs to the cone Q defined in the lemma and, consequently, such coefficients exist.

Now, let us suppose that the opposite case holds, i.e., $P(A_1) > S_1/k$. The value of the bound is $P(A_1)$ and a corresponding dual-feasible solution has the form

$$\lambda_1^* = 1, \quad \lambda_i^* = 0, \quad i = 2, \ldots, n.$$

To prove optimality, we again show the existence of $p^*(\cdot)$ satisfying (29) and (33). This time, constraints in (32) satisfied as strict inequalities are exactly those corresponding to all $\mathbf{y}$ such that $y_1 = 0$, which forces $p^*(\mathbf{y}) = 0$ whenever $y_1 = 0$. In other words, we must construct a measure $p^*(\cdot)$ such that A_1 contains all other events. Our construction will use a circle in $\mathbb{R}^2$ with a circumference of $P(A_1)$. We pick an arbitrary point on the circle and start to tile it with *consecutive* arcs of lengths $P(A_i)$, $i = 1, \ldots, n$ (we will denote their set as $\mathscr{A}$). The endpoints of arcs in $\mathscr{A}$ subdivide the circle into a set $\mathscr{P}$ of arcs of non-zero lengths. Assume that each arc in $\mathscr{A}$ and $\mathscr{P}$ includes only one of its endpoints, say, in the counterclockwise direction. Then, with each arc $r \in \mathscr{P}$, we can associate a $\{0, 1\}$ vector $\mathbf{y}$ so that $y_i = 1$ if and only if r is contained in the arc $a_i \in \mathscr{A}$ corresponding to A_i. For this vector $\mathbf{y}$, we set $p^*(\mathbf{y})$ equal to the length of r. Clearly, for every $\mathbf{y}$ so obtained, we have $y_1 = 1$ since a_1 covers the entire circle. Also, we know that $v(\mathbf{y}) \leqslant k$, since $kP(A_1) > \sum_{j=1}^{n} P(A_j)$, the tiling is consecutive, and, therefore, the circle is tiled at most k times. Finally, (29) holds since the length of a_i, $i = 1, \ldots, n$, is equal to the total length of arcs of $\mathscr{P}$ contained in it. Thus, we have demonstrated the existence of $p^*(\cdot)$ forming a primal–dual optimal pair with $\boldsymbol{\lambda}^*$. $\square$

Acknowledgment

We thank the referee for suggestions that led to clarification of some details.

References

Boole, G., 1854. An Investigation of the Laws of Thought on Which are Founded the Mathematical Theories of Logic and Probabilities. Macmillan, London. Reprinted by Dover, New York, 1958.

Chen, T., 1998. Optimal lower bounds for bivariate probabilities. J. Appl. Probab. 30, 476–492.

Dawson, D., Sankoff, D., 1967. An inequality for probabilities. Proc. Amer. Math. Soc. 18, 504–507.

Fréchet, M., 1935. Généralisations du théorème des probabilités totales. Fund. Math. 25, 379–387.

Hailperin, T., 1965. Best possible inequalities for the probability of a logical function of events. Amer. Math. Monthly 72, 343–359.

Kounias, E., Marin, J., 1976. Best linear Bonferroni bounds. SIAM J. Appl. Math. 30, 307–323.

Kwerel, S., 1975. Most stringent bounds on aggregated probabilities of specified dependent probability systems. J. Amer. Statist. Assoc. 70, 472–479.

Prékopa, A., 1988. Boole–Bonferroni inequalities and linear programming. Oper. Res. 36 (1), 145–162.

Schrijver, A., 1986. Theory of Linear and Integer Programming. Wiley, Chichester, UK.

Seneta, E., 1992. On the history of the strong law of large numbers and Boole's inequality. Historia Math. 19, 24–39.

Seneta, E., Chen, T., 1996. Fréchet optimality of upper bivariate Bonferroni-type bounds. Theory Probab. Math. Statist. 52, 147–152.

Seneta, E., Chen, T., 2002. On explicit and Fréchet-optimal lower bounds. J. Appl. Probab. 39, 81–90.