Available online at www.sciencedirect.com
ScienceDirect
ELSEVIER
Discrete Mathematics 309 (2009) 123–127
DISCRETE MATHEMATICS
www.elsevier.com/locate/disc

The effect of redundancy on probability bounds☆

Fred M. Hoppe

Department of Mathematics and Statistics, McMaster University, 1280 Main St. W., Hamilton, Ontario, L8S 4K1, Canada

Received 23 October 2006; accepted 14 December 2007
Available online 22 January 2008

Abstract

Lower bounds on the probability of a union obtained by applying optimal bounds to subsets of events can provide excellent bounds. Comparisons are made with bounds obtained by linear programming and in the cases considered, the best bound is obtained with a subset that contains no redundant events contributing to the union. It is shown that redundant events may increase or decrease the value of a lower bound but surprisingly even removal of a non-redundant event can increase the bound.

Keywords: Linear programming; Probability bound; Redundant event

1. Introduction

In an earlier paper [4] we showed that lower bounds on the probability of a union of n events could be substantially improved by reducing the number of events $\{A_i\}$ taken in the union and then maximizing. This was demonstrated by comparing the bound in [5] that uses individual and pairwise joint event probabilities with the bound from [2] based only on binomial moments and in every example of [5] not only was there improvement but the exact probability was achieved. An examination showed that the subset of events used to obtain the equality contained no redundant events (that is no events contained in the union of the other events). The present work is a continuation and examines the effect of redundancy on the quality of a lower bound, specifically a degree 2 lower bound (by degree of a bound we mean the largest number of events appearing in any intersection whose probability is used in the bound). We apply this idea of maximizing over subsets to some numerical examples in the literature where lower bounds are obtained by linear programming methods.

It is known that the problem of determining optimal bounds can be set up as a linear program where the constraints are the given probability information [6]. In the simplest case, the constraints involve only the *binomial moments* $S_{j,n} \equiv E\left[\binom{\nu}{j}\right] = \sum_{1\le i_1<\cdots<i_j\le n} P(A_{i_1}\cap A_{i_2}\cap\cdots\cap A_{i_j})$, $1 \le j \le n$, where $j \le j^\star$ so the bound is of degree $j^\star$ (ν counts the number of events occurring). An example of a bound that can be derived in this way is the Fréchet optimal [3,8] degree 2 ($j^\star = 2$) bound of Dawson and Sankoff (2). At the other extreme, the constraints are the $\{P(A_i), 1 \le i \le n, P(A_i \cap A_j), 1 \le i < j \le n\}$ and other individual joint event probabilities. Analytic expressions

☆ Supported by NSERC Discovery Grant.
E-mail address: hoppe@mcmaster.ca.

0012-365X/$ - see front matter © 2007 Elsevier B.V. All rights reserved.
doi:10.1016/j.disc.2007.12.052

are generally not available for these bounds because of the difficulty of identifying the extreme points of the feasible region.

Prekopa and Gao [7] consider a linear program that is computationally less intensive. Their constraints involve

$$S'_{i,k} = \frac{1}{k} \sum_{1 \le i_1 < \cdots < i_k \le n, i \in \{i_1, \ldots, i_k\}} P(A_{i_1} \cap \cdots \cap A_{i_k}), \quad 1 \le i, k \le n \tag{1}$$

for example for $k = 1$, $S'_{i,1} = P(A_i)$ and for $k = 2$, $S'_{i,2} = \sum_{j:j \ne i} P(A_i \cap A_j)$. The idea is to trade optimality for simplicity when n is large. We will refer to bounds that use individual probabilities but grouped into sums according to some specified symmetry as *hybrid bounds*.

The lower bound given in [4] is a hybrid bound because the individual probabilities enter only as binomial moments over subsets of events. In this paper we apply this bound to the examples in [7] and obtain improvement in all the examples. Moreover, the bound always achieves the exact probability of the union. It is also observed that the maximum is achieved over a subset of the events which contains no redundant information that is removal of redundant events increases the magnitude of the lower bounds. However, the opposite can happen and redundant events may sometimes improve the bound. Surprisingly, we show that it is possible even if $\bigcup_{i=1}^m A_i \subset \bigcup_{i=1}^n A_i$ for $m < n$ with strict inclusion for the known optimal bound based on m events to exceed the corresponding bound based on n events. We therefore examine some analytic conditions to determine how redundancy affects bounds.

2. Maximizing bounds over subsets and examples

The bound that we propose to maximize over subsets is

$$P\left(\bigcup_{i=1}^n A_i\right) \ge \frac{\theta S_{1,n}^2}{2S_{2,n} + (2-\theta)S_{1,n}} + \frac{(1-\theta)S_{1,n}^2}{2S_{2,n} + (1-\theta)S_{1,n}} \tag{2}$$

where $\theta = \frac{2S_{2,n}}{S_{1,n}} - \lfloor \frac{2S_{2,n}}{S_{1,n}} \rfloor$ and $\lfloor x \rfloor$ is the largest integer in x. This bound is due to Dawson and Sankoff [2], is easy to compute, and is the sharpest bound that involves only the binomial moments. It is therefore an appropriate choice.

Let I be a subset of $\{1, 2, \ldots, n\}$ and define $S_1(I) = \sum_{i \in I} P(A_i)$, $S_2(I) = \sum\sum_{j<i \in I} P(A_i \cap A_j)$, and $\theta(I) = \frac{2S_2(I)}{S_1(I)} - \lfloor \frac{2S_2(I)}{S_1(I)} \rfloor$.

Theorem (*[4]*).

$$P\left(\bigcup_{i=1}^n A_i\right) \ge \max_I \left(\frac{\theta(I) S_1^2(I)}{2S_2(I) + (2-\theta(I))S_1(I)} + \frac{(1-\theta(I)) S_1^2(I)}{2S_2(I) + (1-\theta(I))S_1(I)} \right). \tag{3}$$

In order to examine how well (3) performs in a setting with large n we apply it to the examples used by Prekopa and Gao [7]. They considered one of the systems of events in [5] when $n = 6$ and also randomly generated three sets of events for $n = 20$ to which they applied their methodology to show improvement with existing bounds. In each case, they tabulated six bounds: the degree $d = 2$ bound (2), their $d = 2$ and $d = 3$ bounds using constraints involving (1), and three additional $d = 3$ bounds referred to as mixture (passive, order, greedy) bounds that arise from computational methods in which the events are split into two groups, to one of which they apply their $d = 2$ bound and to the other their $d = 3$ bound (Tables 1 and 2). We show improvement in all cases using (3).

Example 1. The bounds in our first example are shown in Table 1, which is reproduced from [7] (the data come from system III of [5]), but include an additional last column giving the value of (3). Although n is only 6, it is still worthwhile to make the comparison. It is seen that the bound 0.7890 given by (3) improves on all the rest, and is in fact the exact probability of the union. It arises by applying (2) to the subset of three events $\{A_1, A_2, A_4\}$. As discussed in [4] the events $\{A_3, A_5, A_6\}$ are redundant.

Example 2. Our second example comprises the three larger systems in [7] where $n = 20$ and the corresponding lower bounds are presented in Table 2. Again, (3) improves the other bounds and, as well, since the sample space is provided in [7], it can be checked that (3) achieves the exact probability of the union. (We note in passing that even the weak second-degree lower Bonferroni bound maximized over subsets also results in the exact probabilities.)

Table 1
Lower bound comparison for Example 1

Lower bound	(2) $d=2$	$S'_{i,1}, S'_{i,2}$ $d=2$	Passive $d=3$	Order $d=3$	Greedy $d=3$	$S'_{i,1}, S'_{i,2}, S'_{i,3}$ $d=3$	(3) $d=2$
	0.6933333	0.7221667	0.7221667	0.73145	0.73145	0.73145	0.7890

Table 2
Lower bound comparisons for Example 2

Lower bound	(2) $d=2$	$S'_{i,1}, S'_{i,2}$ $d=2$	Passive $d=3$	Order $d=3$	Greedy $d=3$	$S'_{i,1}, S'_{i,2}, S'_{i,3}$ $d=3$	(3) $d=2$
1	0.8275266	0.8580833	0.86123	0.8698107	0.8832994	0.886446	0.99999995
2	0.8658182	0.9100646	0.9111695	0.9264307	0.9343052	0.93541	0.999989954
3	0.8985498	0.9435812	0.9446198	0.9537189	0.9577441	0.9587778	0.99999189

Furthermore, in all systems, the union $\bigcup_{i=1}^{20} A_i$ can be written as a union of fewer events and the maximum obtains when (2) is applied directly to the union of these fewer events. For instance, in System 1 the union of all 20 events is the same as the union of the four events $\{A_2, A_3, A_4, A_{12}\}$, for System 2 use the four events $\{A_6, A_7, A_{15}, A_{19}\}$, and for System 3 use $\{A_3, A_6, A_9, A_{10}, A_{11}, A_{13}, A_{14}, A_{16}, A_{19}, A_{20}\}$ to obtain the union of all 20 events. In each case, the lower bound (3) applied to the reduced systems of events results in a lower bound which is exact. Thus we have not only improved upon [7] but have obtained the exact probability of the union in their examples by a simple method.

3. Effect of redundancy

A referee of [4] has suggested that a bound based on m events may be better than the one based on $n > m$ events because the additional terms in $S_{1,n}, S_{2,n}$ may be superfluous as they involve events already included in the union of the m events. We therefore examine the extent to which this explanation may be valid. To develop the analysis begin with an alternative form for (2):

$$P\left(\bigcup_{i=1}^{n} A_i\right) \geq \max_{1\leq k\leq n-1}\left(\frac{2S_{1,n}}{k+1} - \frac{2S_{2,n}}{k(k+1)}\right) \tag{4}$$

where the maximum is achieved for the integer $k \equiv \kappa_n = 1 + \lfloor 2S_{2,n}/S_{1,n}\rfloor$ (note that $1 \leq k \leq n-1$). Seneta and Chen [8] have shown that k can range over $k \geq 1$ in (4) with the maximum still achieved in the range $1 \leq k \leq n-1$, an extension that we will require, in particular for $k = n$.

Next, consider a general collection of $n-1$ events $\{A_i\}$ augmented by an included (redundant) $A_n \subseteq \bigcup_{i=1}^{n-1} A_i$. Clearly $S_{1,n} = S_{1,n-1} + P(A_n)$, $S_{2,n} = S_{2,n-1} + \sum_{j=1}^{n-1} P(A_n \cap A_j)$, and

$$\begin{aligned}\frac{2S_{1,n}}{k+1} - \frac{2S_{2,n}}{k(k+1)} &= \frac{2S_{1,n-1} + 2P(A_n)}{k+1} - \frac{2S_{2,n-1} + 2\sum_{j=1}^{n-1} P(A_n \cap A_j)}{k(k+1)} \\ &= \frac{2S_{1,n-1}}{k+1} - \frac{2S_{2,n-1}}{k(k+1)} + \frac{2}{k+1}\left(P(A_n) - \frac{1}{k}\sum_{j=1}^{n-1} P(A_n \cap A_j)\right).\end{aligned} \tag{5}$$

Let B_n and B_{n-1} denote the bounds on the right side of (4) based on $\{n, S_{1,n}, S_{2,n}\}$ and $\{n-1, S_{1,n-1}, S_{2,n-1}\}$ respectively, with κ_n and κ_{n-1} the corresponding optimal values of k. Hence, setting $k = \kappa_n$ in (5),

$$B_n = \frac{2S_{1,n-1}}{\kappa_n+1} - \frac{2S_{2,n-1}}{\kappa_n(\kappa_n+1)} + \frac{2}{\kappa_n+1}\left(P(A_n) - \frac{1}{\kappa_n}\sum_{j=1}^{n-1} P(A_n \cap A_j)\right). \tag{6}$$

Table 3
Elementary conjunctions and their probabilities for Examples 3 and 4

x	$p(x)$	A_1	A_2	A_3	A_4	A_5	A_6	A_7
x_0	0.012	×		×		×		×
x_1	0.022		×		×		×	×
x_2	0.023	×		×		×		×
x_3	0.033		×					×
x_4	0.034	×				×	×	×
x_5	0.044		×	×		×		×
x_6	0.045		×			×	×	×
x_7	0.055		×	×	×		×	×
x_8	0.056	×		×				×
x_9	0.066				×	×		×
x_{10}	0.067		×		×	×		×
x_{11}	0.077		×		×			×
x_{12}	0.078	×			×		×	×
x_{13}	0.088		×					×
x_{14}	0.089	×		×		×	×	

Since $\kappa_n \geq 1$ we can assert (recall the earlier remark concerning the extension by Seneta and Chen [8] of the range of k in (4)) that $\frac{2S_{1,n-1}}{\kappa_n+1} - \frac{2S_{2,n-1}}{\kappa_n(\kappa_n+1)} \leq B_{n-1}$. Thus always

$$B_n \leq B_{n-1} + \frac{2}{\kappa_n + 1}\left(P(A_n) - \frac{1}{\kappa_n}\sum_{j=1}^{n-1} P(A_n \cap A_j)\right) \tag{7}$$

and if $P(A_n) - \frac{1}{\kappa_n}\sum_{j=1}^{n-1} P(A_n \cap A_j) < 0$ then $B_n < B_{n-1}$ in which case removal of the redundant A_n improves the bound.

The next two examples show that redundancy can work in both directions. They are based on System I of [5] followed by a slightly modified version. The original structure is given by the 15 points $\{x_0, \ldots, x_{14}\}$ that determine the six events $\{A_1, \ldots, A_6\}$ as shown in Table 3. Symbol $\times$ in the table indicates that event A_j contains point x_i whose probability is $p(x_i)$. We have included a seventh event A_7 for Example 4.

Example 3. From Table 3, the Dawson–Sankoff lower bound (2) for the original system of 15 points and six events, based on $\{n = 6, S_{1,n}, S_{2,n}\}$ for $P(\bigcup_{i=1}^{6} A_i)$, is 0.7007 and the maximizing value of k in (4) is achieved for $\kappa_6 = 2$. Next observe that $A_3 \subset \bigcup_{i\neq 3} A_i$. Compute (2) based on the five events $\{A_1, A_2, A_4, A_5, A_6\}$ obtained by removing A_3 from the original six events. The result is 0.7300. Thus the optimal bound (2) based on all six events has been improved by removing an event that is redundant. By noting that $P(A_n) - \frac{1}{\kappa_n}\sum_{j=1}^{n-1} P(A_n \cap A_j) = -0.0440 < 0$ this example can be placed in the context of (7). $P(A_3) = 0.2790$, $B_6 = 0.7007$ and $\sum_{i\neq 3} P(A_3 \cap A_i) = 0.6460$, and therefore, if B_5 is the bound (2) with A_3 removed (this requires re-ordering the events for consistent notation only so that A_3 comes last), then from (6), $B_6 \leq B_5 + \frac{2}{3}(0.2790 - \frac{0.6460}{2}) = 0.7300 - 0.0293 = 0.7007$ giving $B_6 < B_5$ and showing how inclusion of A_3 decreases (2). In fact, in this example (7) becomes an equality.

Example 4. On the other hand, if $P(A_n) - \frac{1}{\kappa_n}\sum_{j=1}^{n-1} P(A_n \cap A_j) > 0$ then use of redundant information may possibly improve the bound. To construct such an example append an additional redundant set A_7 to Example 3 comprised of all the points in $\bigcup_{i=1}^{6} A_i$ with the exception of x_{14}. Now (2) based on $\{n = 7, S_{1,n}, S_{2,n}\}$ for $P(\bigcup_{i=1}^{7} A_i)$ is 0.7597, which is achieved for the maximizing value $\kappa_7 = 3$, and this exceeds 0.7007 obtained using $\{A_1, \ldots, A_6\}$. Observe that $P(A_7) - \frac{1}{\kappa_7}\sum_{j=1}^{6} P(A_7 \cap A_j) = 0.7000 - \frac{0.6170}{3} = 0.4943 > 0$ allowing the possibility for the redundant A_7 to increase (2). In Examples 1–3, the maximum bound occurred when a redundant event was removed while in Example 4, inclusion of a redundant event increases the bound (equivalently, removal of a redundant event decreases the bound). It would therefore be interesting to find an example where removal of a non-redundant event improves the bound (equivalently, inclusion of a non-redundant event decreases the bound). Here is such a case.

Example 5. Consider a sample space with $n = 4$ events where the only non-empty elementary conjunctions have the following probabilities: $P(A_1 \cap \bar{A}_2 \cap \bar{A}_3 \cap \bar{A}_4) = 0.02$, $P(A_1 \cap A_2 \cap A_3 \cap A_4) = 0.37$, $P(\bar{A}_1 \cap A_2 \cap A_3 \cap A_4) =$

0.51, $P(\bar{A}_1 \cap \bar{A}_2 \cap A_3 \cap A_4) = 0.06$, $P(A_1 \cap A_2 \cap A_3 \cap \bar{A}_4) = 0.01$, $P(\bar{A}_1 \cap A_2 \cap \bar{A}_3 \cap \bar{A}_4) = 0.01$, $P(\bar{A}_1 \cap \bar{A}_2 \cap \bar{A}_3 \cap \bar{A}_4) = 0.01$. In this space, the Dawson–Sankoff bound applied to all four events yields 0.96. It can be seen that event A_3 is redundant since $A_3 \subset A_1 \cup A_4$. Removal of this event and application of the Dawson–Sankoff bound to $A_1 \cup A_2 \cup A_4$ decreases the bound to 0.9567. Event A_2 is not redundant, and yet removal of A_2 and application of the Dawson–Sankoff bound to $A_1 \cup A_3 \cup A_4$ increases the bound to 0.97. Finally, removal of both A_2, A_3 gives the lower bound 0.98. Note that the actual probability of the union of all four events is 0.99. This simple example simultaneously demonstrates the non-intuitive behaviour that removal of a redundant event may decrease a bound, while removal of a non-redundant event may increase a bound (it is trivial to show, using a disjoint event, that removal of a non-redundant event may decrease a bound — this is the expected behaviour).

Here $P(A_1) = 0.4$, $P(A_2) = 0.9$, $P(A_3) = 0.95$, $P(A_4) = 0.95$, $P(A_1 \cap A_2) = 0.38$, $P(A_1 \cap A_3) = 0.38$, $P(A_2 \cap A_3) = 0.89$, $P(A_1 \cap A_4) = 0.37$, $P(A_2 \cap A_4) = 0.88$, $P(A_3 \cap A_4) = 0.94$, which are the same as the corresponding values in Example 2 of [6] for which the probabilities of all elementary conjunctions were not provided. In that paper, two different degree 2 lower bounds, 0.96 and 0.97, were obtained corresponding to general numerical algorithms for solving the linear programming problem. However, we have just seen that the lower bound (3), maximized at $I = \{1, 4\}$, is 0.98. This provides another manifestation of the simplicity in the use of optimization over subsets.

Example 6. This last example is from Chen [1] who uses not linear programming but Hamiltonian-type circuits to derive hybrid degree 2 lower bounds based upon $\{S_{1,n}, P(A_i \cap A_j), 1 \le i < j \le n\}$. His example for $n = 6$ yields the lower bound 0.925 while (3) is 0.930 which is based on the events A_2, A_5 which are disjoint with $P(A_2) = 0.72$, $P(A_5) = 0.21$ and $A_2 \cup A_5 = \bigcup_{i=1}^{6} A_i$ rendering the other four events redundant.

4. Final remarks

Maximization bounds involving only binomial moments over subsets appears to provide a simple yet effective approach to finding good lower bounds on the probability of a union $P(\cup_{i=1}^{n} A_i)$. In the examples considered, taken from the literature, the resulting bounds improved those obtained by linear programming methods designed to cope with an exponentially large number of constraints as n gets large. For moderate n, the maximum can quickly be found by searching all cases. For large n it would be necessary to find an algorithm that finds the maximum. For instance (6) directly compares B_n with B_{n-1} and depending on the sign of the term in parentheses, inclusion of A_n may increase or decrease the bound. We may thus sequentially check which events should be included as a means of approaching the maximizing bound.

Acknowledgments

Thanks to the referees for very useful suggestions.

References

[1] J.T. Chen, A lower bound using Hamilton-type circuits and its applications, J. Appl. Probab. 40 (2003) 1121–1132.
[2] D.A. Dawson, D. Sankoff, An inequality for probabilities, Proc. Amer. Math. Soc. 18 (1967) 504–507.
[3] M. Fréchet, Généralisations du théorème des probabilités totales, Fund. Math. 25 (1935) 379–387.
[4] F.M. Hoppe, Improving probability bounds by optimization over subsets, Discrete Math. 306 (2006) 526–530.
[5] H. Kuai, F. Alajaji, G. Takahara, A lower bound on the probability of a finite union of events, Discrete Math. 215 (2000) 147–158.
[6] A. Prékopa, Boole–Bonferroni inequalities and linear programming, Oper. Res. 36 (1) (1988) 145–162.
[7] A. Prékopa, L. Gao, Bounding the probability of the union of events by aggregation and disaggregation in linear programs, Discrete Appl. Math. 145 (2005) 444–454.
[8] E. Seneta, T. Chen, On explicit and Fréchet-optimal lower bounds, J. Appl. Probab. 39 (2002) 81–90.