A Correlation Inequality and Its Application to a Word Problem

Dimitris Achlioptas[*§] (optas@cs.toronto.edu)
Lefteris M. Kirousis[†§] (kirousis@fryni.ceid.upatras.gr)
Evangelos Kranakis[‡§] (kranakis@scs.carleton.ca)
Danny Krizanc[‡§] (krizanc@scs.carleton.ca)
Michael S.O. Molloy[*§] (molloy@cs.toronto.edu)

Abstract

We give upper bounds for the probability that a random word of a given length contains at least one letter from each member of a given collection of sets of letters. We first show a correlation inequality that bounds the probability of the conjunction of a number of pairwise dependent events. The bound takes into account only pairs of events that are positively correlated, yielding significantly tighter bounds in some interesting cases.

Key Words and Phrases: Correlation inequality, Finite alphabet, Random word.

[*]University of Toronto, Department of Computer Science, Toronto, ON M5S 3G4, Canada.
[†]University of Patras, Department of Computer Engineering and Informatics, Rio, 265 00 Patras, Greece.
[‡]Carleton University, School of Computer Science, Ottawa, ON K1S 5B6, Canada.
[§]The research of the first author was supported in part by a fellowship from NSERC (Natural Sciences and Engineering Research Council). The research of the other authors was supported in part by NSERC grants. The research of the second author was performed while he was visiting Carleton University and was supported in part by a grant for sabbatical leaves from the University of Patras.

1 Introduction

Let Σ be a finite non-empty alphabet with n letters, and let $\mathcal{A} = (A_i)_{i=1,\ldots,r}$ be a collection of non-empty subsets of Σ. An m-letter random word w on the alphabet Σ is a sequence of m letters from Σ selected at random independently, uniformly and with replacement. We give upper bounds on the probability that a random word contains at least one letter from each of the sets $A_i, i = 1, \ldots, r$. More specifically, if p_i is the probability that w contains at least one letter from A_i and if q_{ij} is the probability that w contains no letter from $A_i \cup A_j$, then our upper bounds are expressed as a product of $p_1 \cdot p_2 \cdots p_r$ with a correlation factor that is a function only of the p_is and the sum $\sum_{i \sim j} q_{ij}$, where $i \sim j$ denotes that the *intersection* $A_i \bigcap A_j \neq \emptyset$ and that $i \neq j$. Notice that $p_i = 1 - (1 - (|A_i|/|\Sigma|))^m$ and $q_{ij} = (1 - (|A_i \cup A_j|/|\Sigma|))^m$.

Our motivation comes from the work in [4] concerning the satisfiability problem of random Boolean formulas, where the question of bounding the probability that a random formula is not satisfied by a given collection of truth assignments was encountered.

For each $i = 1, \ldots, r$, let E_i be the event of at least one letter from A_i occuring in w. Let also $1_{\neg E_i}$ be the indicator variable for the complement of E_i, i.e. $1_{\neg E_i}(w) = 0$, if w contains a letter from A_i, and $1_{\neg E_i}(w) = 1$, otherwise. Then, obviously, w contains a letter from each of the A_is iff $\sum_i 1_{\neg E_i}(w) = 0$. Observe now that *all* pairs of events E_i and E_j are dependent. This is so even in the extreme case where all A_i are pairwise disjoint. Therefore, it is unlikely that we can apply Chernoff bounds to bound the probability of $\sum_i 1_{\neg E_i}(w) = 0$, as Chernoff bounds assume independence among the events E_i. Similarly with the Schmidt, Siegel, and Srinivasan method [5], where it is assumed that the events are k-wise independent, for $k \ll r$. Neither Janson's inequality [3], even in its general form presented in Spencer's book [6], can be applied directly, as the conditions that must be assumed are not true in our case. More crucial though than the fact that the necessary assumptions for Janson's inequality do not hold is that this inequality, when applicable to a collection of events $J_i, i = 1, \ldots, r$, gives an upper bound for $\mathbf{Pr}[\wedge_i J_i]$ which is a product of $\prod_i \mathbf{Pr}[J_i]$ together with a correlation factor that is a function of the sum $\sum_{i,j} \mathbf{Pr}[\neg J_i \wedge \neg J_j]$, where this sum is taken over all possible pairs J_i and J_j $(i \neq j)$ of *dependent* events. As in our case *all* pairs of events are dependent, Janson's inequality, even if it were directly applicable, would yield an upper bound with a large correlation factor. In this paper, we prove a variant of Janson's result which is applicable to the word problem we consider. This variant also has the nice property that when applied to the word problem gives a bound where the correlation factor involves only the sum $\sum_{i \sim j} q_{ij}$. In other words, we reduce the range of the sum to pairs of events for which $A_i \bigcap A_j \neq \emptyset$ and thus we get a smaller correlation factor.

The intuition behind our improvement is the following: a pair of events E_i and E_j for which $A_i \bigcap A_j = \emptyset$ is nonpositively correlated, i.e., $\mathbf{Pr}[E_i \wedge E_j] \leq \mathbf{Pr}[E_i]\mathbf{Pr}[E_j]$. Therefore it is plausible that we can avoid having such a pair contribute to the correlation factor of Janson's inequality. On the other hand, when $A_i \bigcap A_j \neq \emptyset$, then there is a "strong positive component" in the correlation of E_i and E_j, so such pairs must contribute to the correlation factor.

In the next section we will formally describe and prove a correlation inequality, which, as we subsequently prove in Section 3, is applicable in the case of the word problem. In the latter section, we will also give our upper bounds.

2 A Correlation Inequality

We start with a definition:

DEFINITION 1 *Let $\mathcal{J} = \{J_i : i = 1, \ldots, r\}$ be a finite collection of events in an arbitrary probability space. We say that J_j is nonpositively correlated to J_i under any conjunction of conditions from $\mathcal{J}$ iff for any conjunction S of events from $\mathcal{J}$,*

$$\mathbf{Pr}[(J_i \mid J_j) \mid S] \leq \mathbf{Pr}[J_i \mid S].$$

Notice that it is not in general true that if J_j is independent from J_i, then J_j is nonpositively correlated to J_i under any conditions, as the independence may be destroyed under certain conditions. It is also easy to see that if J_j is nonpositively correlated to J_i under any conjunction of conditions from $\mathcal{J}$, then so is J_i to J_j.

Now for each i, let P_i be any subset of $\{1, \ldots, i-1\}$ such that for any $j \in \{1, \ldots, i-1\} \setminus P_i$, J_j and J_i are nonpositively correlated under any conjunction of conditions from $\mathcal{J}$. Let

$$\Delta = \sum_i \sum_{j \in P_i} \mathbf{Pr}[\neg J_i \wedge \neg J_j],$$

and let ϵ be such that $1 - \epsilon \leq \mathbf{Pr}[J_i]$, for all i. Also let $\mu = \sum_i \mathbf{Pr}[\neg J_i]$. Then the following holds.

THEOREM 1

$$\mathbf{Pr}[\wedge_i J_i] \leq \left(\prod_i \mathbf{Pr}[J_i] \right) \cdot e^{\Delta/(1-\epsilon)}. \tag{1}$$

Moreover, if $\Delta \geq \mu(1-\epsilon)$, then

$$\mathbf{Pr}[\wedge_i J_i] \leq e^{-\frac{\mu^2(1-\epsilon)}{4\Delta}}. \tag{2}$$

PROOF Notice that, unlike the case of Janson's inequalities as presented in [6], we do not make any correlation assumptions about the events J_i. For the proof, we start with the first inequality. Since,

$$\mathbf{Pr}[\wedge_i J_i] = \prod_i \mathbf{Pr}[J_i \mid \wedge_{j=1,\ldots,i-1} J_j],$$

we will try to find an upper bound for $\mathbf{Pr}[J_i \mid \wedge_{j=1,\ldots,i-1} J_j]$. We first notice that

$$\mathbf{Pr}[J_i \mid \wedge_{j=1,\ldots,i-1} J_j] \leq \mathbf{Pr}[J_i \mid \wedge_{j \in P_i} J_j]. \tag{3}$$

To prove the last inequality, say, without loss of generality, that J_{i-1} is nonpositively correlated to J_i under any conjunction of conditions from $\mathcal{J}$, and notice that by definition it follows that $\mathbf{Pr}[J_i \mid J_1 \cdots J_{i-1}] \leq \mathbf{Pr}[J_i \mid J_1 \cdots J_{i-2}]$; repeat this as necessary to get inequality (3).

Therefore it is enough to find an upper bound for $\mathbf{Pr}[J_i \mid \wedge_{j \in P_i} J_j]$, or alternatively a lower bound for $\mathbf{Pr}[\neg J_i \mid \wedge_{j \in P_i} J_j]$. But $\mathbf{Pr}[\neg J_i \mid \wedge_{j \in P_i} J_j] \geq \mathbf{Pr}[\neg J_i \wedge \wedge_{j \in P_i} J_j]$.

The rest of our proof follows the steps of the corresponding proof in [6] (page 82). By inclusion-exclusion

$$\mathbf{Pr}[\neg J_i \wedge \wedge_{j \in P_i} J_j] \geq \mathbf{Pr}[\neg J_i] - \sum_{j \in P_i} \mathbf{Pr}[\neg J_i \wedge \neg J_j].$$

Taking complements, we conclude that

$$\mathbf{Pr}[J_i \mid \wedge_{j=1,\ldots,i-1} J_j] \leq \mathbf{Pr}[J_i] + \sum_{j \in P_i} \mathbf{Pr}[\neg J_i \wedge \neg J_j].$$

Therefore, by the choice of ϵ, we conclude that

$$\mathbf{Pr}[J_i \mid \wedge_{j=1,\ldots,i-1} J_j] \leq \mathbf{Pr}[J_i] \left(1 + \frac{1}{1-\epsilon} \sum_{j \in P_i} \mathbf{Pr}[\neg J_i \wedge \neg J_j]\right).$$

Multiplying out the last inequalities and using the fact that $1 + x \leq e^x$, we obtain the first inequality of the theorem. The second one may be proved by repeating verbatim the corresponding proof in [6] (page 83). Only a word of caution for the factor 4 that appears in the denominator of the exponent of e in inequality (2): this factor is there because of the non-symmetric form in which we wrote the range of the sum in the definition of Δ. For the same reason, inequality (1), contrary to the corresponding Janson's inequality in [6], does not have the factor 2 in the denominator of the exponent of e. □

3 The Bounds

We first prove the following theorem about the family of events $E_i, i = 1, \ldots, r$ defined in the Introduction.

THEOREM **2** *If $A_{i_1} \bigcap A_{i_2} = \emptyset$ then the events E_{i_1} and E_{i_2} are nonpositively correlated for any conjunction of conditions from $\{E_i : i = 1, \ldots, r\}$.*

PROOF Let S be an arbitrary conjunction of events in the family of the E_is. To make the notation simpler, the conditioned on S probability of an event X will be denoted by $\mathbf{Pr}_S[X]$. We also denote by E_i^l the event that there is a letter from the set A_i at the lth position of the word. Clearly, $\neg E_i = \wedge_{l=1}^m \neg E_i^l$. We have to prove that $\mathbf{Pr}_S[E_{i_1}|E_{i_2}] \leq \mathbf{Pr}_S[E_{i_1}]$, assuming the corresponding sets A_{i_1} and A_{i_2} are disjoint. The inequality to be proved is equivalent to:

$$\mathbf{Pr}_S[E_{i_2}|\neg E_{i_1}] \geq \mathbf{Pr}_S[E_{i_2}] \tag{4}$$

Assume first that that S, which is a set of m-tuples from the alphabet Σ, is a Cartesian product $S_1 \times \cdots \times S_m$, where the S_ls ($l = 1, \ldots, m$) are subsets of the alphabet Σ. For an arbitrary position l in the word, let $x_l = \mathbf{Pr}_{S_l}[E_{i_2}^l]$ and let $y_l = \mathbf{Pr}_{S_l}[E_{i_2}^l|\neg E_{i_1}^l]$. Since the sets A_{i_1} and A_{i_2} are disjoint, $y_l = \mathbf{Pr}_{S_l}[E_{i_2}^l|\neg E_{i_1}^l] = \mathbf{Pr}_{S_l}[E_{i_2}^l \wedge \neg E_{i_1}^l]/\mathbf{Pr}_{S_l}[\neg E_{i_1}^l] = x_l/\mathbf{Pr}_{S_l}[\neg E_{i_1}^l]$, and therefore $y_l \geq x_l$. Now observe that:

$$\begin{aligned}
& \mathbf{Pr}_S[\neg E_{i_2}|\neg E_{i_1}] = \frac{\mathbf{Pr}_S[\neg E_{i_2} \wedge \neg E_{i_1}]}{\mathbf{Pr}_S[\neg E_{i_1}]} \\
= \quad & \frac{\mathbf{Pr}_S[\wedge_{l=1}^m \neg E_{i_2}^l \wedge \wedge_{l=1}^m \neg E_{i_1}^l]}{\mathbf{Pr}_S[\wedge_{l=1}^m \neg E_{i_1}^l]} = \frac{\mathbf{Pr}_S[A_{i_2}^c \cap A_{i_1}^c \times \cdots \times A_{i_2}^c \cap A_{i_1}^c]}{\mathbf{Pr}_S[A_{i_1}^c \times \cdots \times A_{i_1}^c]} \\
= \quad & \frac{\Pi_{l=1}^m \mathbf{Pr}_{S_l}[A_{i_2}^c \cap A_{i_1}^c]}{\Pi_{l=1}^m \mathbf{Pr}_{S_l}[A_{i_1}^c]} = \Pi_{l=1}^m \mathbf{Pr}_{S_l}[\neg E_{i_2}^l|\neg E_{i_1}^l] = \Pi_{l=1}^m (1 - y_l).
\end{aligned}$$

Above we made use of the identity

$$\mathbf{Pr}_{S_1 \times \cdots \times S_m}[X_1 \times \cdots \times X_m] = \Pi_{l=1}^{m} \mathbf{Pr}_{S_l}[X_l],$$

which holds for arbitrary subsets S_l and X_l of Σ and follows by trivial set-theoretic manipulations.

From the above series of equalities it follows that $\mathbf{Pr}_S[E_{i_2} | \neg E_{i_1}] = 1 - \Pi_{l=1}^{m}(1 - y_l)$. Also, $\mathbf{Pr}_S[E_{i_2}] = 1 - \Pi_{l=1}^{m}(1 - x_l)$. Since we have proved that $y_l \geq x_l$, $\forall l = 1, \ldots, m$, inequality (4) follows.

Now assume that S is not necessarily a Cartesian product. Any arbitrary S however is the pairwise disjoint union of Cartesian products (e.g., it is the pairwise disjoint union of singletons, and a set having as its only element an m-tuple is the Cartesian product of singletons with elements in Σ). The theorem now immediately follows from the following claim that holds in any probability space:

Claim: For any events X, Y, and Z such that Z is the pairwise disjoint union of of a family of events $Z_1, \ldots, Z_r$, if $\forall j = 1, \ldots, r, \mathbf{Pr}_{Z_j}[X] \leq \mathbf{Pr}_{Z_j}[Y]$, then $\mathbf{Pr}_Z[A] \leq \mathbf{Pr}_Z[B]$.

The proof of the above claim follows by trivial set-theoretic manipulations. □

In the framework of our problem, let $\Delta = \sum_{i \sim j} \mathbf{Pr}[\neg E_i \wedge \neg E_j]$. Recall that $i \sim j$ means that $A_i \bigcap A_j \neq \emptyset$. It does not mean that E_i and E_j are dependent (after all, in our case, all pairs of events are dependent). Also let ϵ be such that $1 - \epsilon \leq \mathbf{Pr}[E_i]$, for all i. Moreover, let $\mu = \sum_i \mathbf{Pr}[\neg E_i]$. It is easy to see by standard indicator variable arguments that μ is the expected number of sets A_i that the random word avoids. Finally, recall that p_i denotes the probability that the random word contains at least one letter from A_i. Then

THEOREM **3** *The probability that the random word contains at least one letter from each set A_i is bounded above by*

$$p_1 \cdots p_r e^{\Delta/[2(1-\epsilon)]}.$$

Also if $2\Delta \geq \mu(1 - \epsilon)$, then this probability is bounded above by

$$e^{-\frac{\mu^2(1-\epsilon)}{2\Delta}}.$$

PROOF The theorem follows by a direct application of Theorems 1 and 2. Note that the range of the sum in Δ is now written in a symmetric way, so "the current" Δ is half of the Δ in Theorem 2. □

From the above theorem it follows immediately that:

COROLLARY **1** *If the sets A_i are pairwise disjoint then the probability that the random word contains at least one letter from each set A_i is bounded above by $p_1 \cdots p_r$.*

4 Discussion

It is not hard to bound the probability of a random word containing at least one letter from each A_i by the so called "second moment method." Indeed, we already mentioned in the Introduction that if X is the random variable $\sum_i 1_{\neg E_i}(w)$, then the probability of w containing at least one letter from each A_i is equal to $\mathbf{Pr}[X = 0]$. By Chebyshev's inequality (see, e.g., page

40 in [1]), $\mathbf{Pr}[X = 0] \leq \mathbf{var}[X]/(\mathbf{E}[X])^2$. But $\mathbf{var}[X] = \sum_i \mathbf{var}[1_{\neg E_i}] + \sum_{i \neq j} \mathbf{cov}[1_{\neg E_i}, 1_{\neg E_j}]$. Also, it can be easily seen that

$$\mathbf{var}[1_{\neg E_i}] \leq \mathbf{E}[1_{\neg E_i}], \tag{5}$$

and that

$$\mathbf{cov}[1_{\neg E_i}, 1_{\neg E_j}] = \mathbf{Pr}[\neg E_i \wedge \neg E_j] - \mathbf{Pr}[\neg E_i]\mathbf{Pr}[\neg E_j]. \tag{6}$$

Finally if for a pair E_i and E_j the corresponding sets A_i and A_j are disjoint, then it is immediate that $\mathbf{cov}[1_{\neg E_i}, 1_{\neg E_j}] \leq 0$. On the other hand, in all cases and in particular when A_i and A_j intersect we have that $\mathbf{cov}[1_{\neg E_i}, 1_{\neg E_j}] \leq \mathbf{Pr}[\neg E_i \wedge \neg E_j]$. From the last two inequalities and also by inequalities (6) and (5), we conclude that

$$\mathbf{var}[X] \leq \sum_i \mathbf{Pr}[\neg E_i] + \sum_{i,j:A_i \bigcap A_j \neq \emptyset} \mathbf{Pr}[\neg E_i \wedge \neg E_j] = \mu + \Delta.$$

Therefore by Chebyshev's inequality, the probability that w contains at least one letter from each A_i is at most $(1/\mu) + (\Delta/\mu^2)$. However, our bounds are better in many cases, as, for example, in inequality (2) of Theorem 1, where the expression Δ/μ^2 appears negated and inverted in the exponent, which is much better than having it as is.

Acknowledgments

We thank Petr Savicky for pointing out to us that the proof of Theorem 2 in a previous draft of this paper was incorrectly formulated. We also thank Yannis Stamatiou for his help in correcting that proof.

References

[1] N. Alon and J. Spencer, *The Probabilistic Method,* Wiley, New York, 1992.

[2] M.R. Garey and D.S. Johnson, *Computers and Intractability: A Guide to the Theory of NP-Completeness,* Freeman, San Francisco, 1979.

[3] S. Janson, "Poisson approximation for large deviations," *Random Structures and Algorithms* 1, pp 221–230, 1990.

[4] L.M. Kirousis, E. Kranakis, and D. Krizanc, *Approximating the Unsatisfiability Threshold of Random Formulas*, Technical Report TR-95-26, School of Computer Science, Carleton University, Ottawa, Canada, 1995.

[5] J.P. Schmidt, A. Siegel, and A. Srinivasan, "Chernoff-Hoeffding bounds for applications with limited independence," *Proceedings of the 4th Annual ACM-SIAM Symposium on Discrete Algorithms,* pp. 331–340, 1993.

[6] J. H. Spencer, *Ten Lectures on the Probabilistic Method*, 2nd edition, SIAM, Philadelphia, 1994.