Submodularity and pairwise independence

Arjun Ramachandra ARJUN_RAMACHANDRA@SUTD.EDU.SG
Singapore University of Technology and Design, Singapore

Karthik Natarajan KARTHIK_NATARAJAN@SUTD.EDU.SG
Singapore University of Technology and Design, Singapore

Abstract

In this paper, we provide a characterization of the expected value of submodular set functions with pairwise independent random input. The set of pairwise independent (uncorrelated) probability distributions contains the mutually independent distribution and is contained within the set of arbitrarily dependent (correlated) distributions. We study the ratio of the maximum expected value of a function with arbitrary dependence among the random input with given marginal probabilities to the maximum expected value of the function with pairwise independent random input with the same marginal probabilities. The definition of the ratio is inspired from the "correlation gap" ratio of Agrawal et al. (2012) and Calinescu et al. (2007). Our results show that for any monotone submodular set function defined on n variables, the ratio is bounded from above by $4/3$ in the following cases: (a) for small n (specifically $n = 2, 3$) with general marginal probabilities, and (b) for general n with small marginal probabilities. The bound is tight in cases (a) and (b). This contrasts with the $e/(e-1)$ bound on the correlation gap ratio for monotone submodular set functions with mutually independent random input which is known to be tight in case (b). Our results illustrate a fundamental difference in the behavior of submodular functions with weaker notions of independence. We discuss an application in distributionally robust optimization and end the paper with a conjecture.

Keywords: submodularity, pairwise independence, correlation gap, distributionally robust optimization

1. Introduction

Submodular set functions play an important role in machine learning problems (Bach, 2013; Krause and Jegelka, 2013). An important notion which describes the behavior of these functions under random input is the "correlation gap", which was introduced in Agrawal et al. (2012), building on the work of Calinescu et al. (2007). The correlation gap is defined as the ratio of the maximum expected value of a set function with arbitrary dependence among the random input with fixed marginal probabilities to the expected value of the function with mutually independent random input with the same marginal probabilities. The key result in this area is that for any monotone submodular set function, the correlation gap is always bounded from above by $e/(e-1) \approx 1.582$ and this bound is tight (Calinescu et al., 2007; Agrawal et al., 2012). This result has been applied in many settings including content resolution schemes (Chekuri et al., 2014; Feldman et al., 2021), mechanism design (Chawla et al., 2010; Yan, 2011), combinatorial prophet inequalities (Rubinstein and Singla, 2017; Chekuri and Livanos, 2021) and distributionally robust optimization (Agrawal et al., 2012; Staib et al., 2019). In this paper, we introduce a modification of the notion of correlation gap to set functions with pairwise independent random input and derive a new bound of $4/3$ for monotone submodular set functions in several cases. This adds to a recent result in Ramachandra and Natarajan (2021) where the bound was identified for a specific set function.

1.1. Preliminaries

Let $[n] = \{1, 2, \ldots, n\}$ be the ground set and $f : 2^{[n]} \to \mathbb{R}_+$ be a nonnegative set function. The function f is submodular if $f(S) + f(T) \geq f(S \cap T) + f(S \cup T)$ for all $S, T \subseteq [n]$ or equivalently $f(S \cup \{i\}) - f(S) \geq f(T \cup \{i\}) - f(T)$ for all $S \subseteq T \subseteq [n]$, $i \notin T$. We let $f(i|S) = f(S \cup \{i\}) - f(S)$ denote the marginal contribution of adding i to S. The function f is monotone if $f(S) \leq f(T)$ for all $S \subseteq T \subseteq [n]$. The multilinear extension of f is defined for any $\boldsymbol{x} \in [0,1]^n$ as:

$$F(\boldsymbol{x}) = \sum_{S \subseteq [n]} f(S) \prod_{i \in S} x_i \prod_{i \notin S} (1 - x_i), \tag{1}$$

and is simply the expected value of the set function where each input $i \in [n]$ is selected with probability x_i independently. The concave closure of f is defined for any $\boldsymbol{x} \in [0,1]^n$ as the maximum expected value of the set function over all joint distributions of the random input where each $i \in [n]$ is selected with probability x_i. It is computed as the optimal value of the following primal-dual pair of linear programs:

$$\begin{aligned} f^+(\boldsymbol{x}) = \max \quad & \sum_S \theta(S) f(S) & = \quad \min \quad & \sum_{i \in [n]} \lambda_i x_i + \lambda_0 \\ \text{s.t} \quad & \sum_S \theta(S) = 1, & \text{s.t.} \quad & \sum_{i \in S} \lambda_i + \lambda_0 \geq f(S), \quad \forall S \subseteq [n]. \\ & \sum_{S: S \ni i} \theta(S) = x_i, \quad \forall i \in [n], & & \\ & \theta(S) \geq 0, \qquad \forall S \subseteq [n], & & \end{aligned} \tag{2}$$

Both $F : [0,1]^n \to \mathbb{R}_+$ and $f^+ : [0,1]^n \to \mathbb{R}_+$ define continuous extensions of the set function $f : 2^{[n]} \to \mathbb{R}_+$. The functions satisfy $f(S) = F(1_S) = f^+(1_S)$ for all $S \subseteq [n]$ where 1_S denotes the indicator vector of the set S. The correlation gap is defined as the ratio $f^+(\boldsymbol{x})/F(\boldsymbol{x})$ where we define $0/0 = 1$.

Theorem 1 *[Calinescu et al. (2011); Agrawal et al. (2012)] For any monotone submodular function $f : 2^{[n]} \to \mathbb{R}_+$ and any $\boldsymbol{x} \in [0,1]^n$, $f^+(\boldsymbol{x})/F(\boldsymbol{x}) \leq e/(e-1)$. The upper bound is tight for $f(S) = \min(|S|, 1)$ and $\boldsymbol{x} = (1/n, \ldots, 1/n)$ as $n \uparrow \infty$.*

Even for $n = 2$, violating the assumption of either monotonicity or submodularity might make the correlation gap unbounded (see Appendix A.1 for examples). Another important extension of submodular functions is the convex closure (denoted by $f^-(\boldsymbol{x})$) where we minimize over the $\theta(S)$ variables in (2), rather than maximize. We focus only on $F(\boldsymbol{x})$ and $f^+(\boldsymbol{x})$ in this paper.

2. Set functions with pairwise independent random inputs

2.1. Pairwise independence

Given a Bernoulli random vector $\tilde{\boldsymbol{c}} = (\tilde{c}_1, \ldots, \tilde{c}_n)$, mutual independence is defined by the condition $\mathbb{P}(\tilde{c}_i = c_i, \forall i \in [n]) = \prod_{i=1}^n \mathbb{P}(\tilde{c}_i = c_i)$ for all $\boldsymbol{c} \in \{0,1\}^n$. Pairwise independence is a weaker notion of independence where only pairs of random variables are independent. It is defined by the condition $\mathbb{P}(\tilde{c}_i = c_i, \tilde{c}_j = c_j) = \mathbb{P}(\tilde{c}_i = c_i)\mathbb{P}(\tilde{c}_j = c_j)$ for all $(c_i, c_j) \in \{0,1\}^2$, $i \neq j$. While mutual independence implies pairwise independence, the reverse is not true (Bernstein, 1946). For example, consider a distribution on three random variables that assigns a probability of $1/4$ to each of the scenarios $(0,0,1)$, $(0,1,0)$, $(1,0,0)$ and $(1,1,1)$. The three random variables are pairwise independent but not mutually independent. Another pairwise independent distribution that has the same marginal probabilities of $(1/2,1/2,1/2)$ is given by the scenarios $(0,0,0)$, $(0,1,1)$, $(1,0,1)$ and $(1,1,0)$, each occurring with a probability of $1/4$. For general n, constructions of pairwise independent random variables can be found in the works of Geisser and Mantel (1962), Karloff and Mansour (1994) and Koller and Meggido (1994).

2.2. Applications of pairwise independence

One of the motivations for studying constructions of pairwise independent random variables is that there is always a joint distribution that has a low cardinality support (polynomial in n), in contrast to mutual independence (exponential in n). The low cardinality of the distribution has important ramifications in efficiently derandomizing algorithms for combinatorial optimization problems (Wigderson, 1994; Luby and Wigderson, 2005). Pairwise independence is useful in modeling situations where the underlying randomness has zero or close to zero correlations but complex higher-order dependencies. Such behavior has been experimentally observed in cortical neurons in Schneidman et al. (2006), where weak correlations between pairs of neurons coexist with strong collective behaviour of the joint response of multiple neurons and approximations based on the assumption of mutual independence are weak.

2.3. A new continuous extension

Inspired by the definition of the concave closure and to characterize the behavior of set functions with pairwise independent random inputs, we consider the following continuous extension of a set function.

Definition 2 *The maximum expected value of a set function f over all pairwise independent distributions of the random input where each $i \in [n]$ is selected with probability x_i is computed as the optimal value of the following primal-dual pair of linear programs:*

$$
\begin{aligned}
f^{++}(\boldsymbol{x}) = \max \quad & \sum_{S} \theta(S) f(S) & = \min \quad & \sum_{i<j\in[n]} \lambda_{ij} x_i x_j + \sum_{i\in[n]} \lambda_i x_i + \lambda_0 \\
\text{s.t} \quad & \sum_{S} \theta(S) = 1, & \text{s.t.} \quad & \sum_{i<j\in S} \lambda_{ij} + \sum_{i\in S} \lambda_i + \lambda_0 \geq f(S), \\
& \sum_{S:S\ni i} \theta(S) = x_i, \quad \forall i \in [n], & & \forall S \subseteq [n]. \\
& \sum_{S:S\ni i,j} \theta(S) = x_i x_j, \quad \forall i<j \in [n], & & \\
& \theta(S) \geq 0, \quad \forall S \subseteq [n], & &
\end{aligned}
\tag{3}
$$

where $f^{++}(\boldsymbol{x})$ is referred to as the upper pairwise independent extension. We define the pairwise independent correlation gap as the ratio $f^{+}(\boldsymbol{x})/f^{++}(\boldsymbol{x})$.

Clearly $f(S) = f^{++}(1_S)$ for all $S \subseteq [n]$. Instead of maximizing, if we minimize over the $\theta(S)$ variables in (3), we obtain a lower pairwise independent extension (denoted by $f^{--}(\boldsymbol{x})$). In this paper, we only focus on analyzing $f^{++}(\boldsymbol{x})$ for monotone submodular functions and its relation to $f^{+}(\boldsymbol{x})$ and $F(\boldsymbol{x})$. Clearly, $F(\boldsymbol{x}) \leq f^{++}(\boldsymbol{x}) \leq f^{+}(\boldsymbol{x})$ and hence $f^{+}(\boldsymbol{x})/f^{++}(\boldsymbol{x}) \leq f^{+}(\boldsymbol{x})/F(\boldsymbol{x})$.

Our results show that for monotone submodular set functions in several cases, the maximum possible value which the ratio $f^+(\boldsymbol{x})/f^{++}(\boldsymbol{x})$ takes is provably smaller than the maximum possible value of the ratio $f^+(\boldsymbol{x})/F(\boldsymbol{x})$ and we conjecture this to hold more generally. We next discuss applications and some recent related work.

2.4. Applications and related work

1. *Distributionally robust optimization*: Using Theorem 1, one can show that the loss in performance by using this approach is bounded by a constant factor for specific optimization problems even if the true joint distribution is not independent (Agrawal et al., 2012). In some instances, the distributionally robust optimization problem given only the marginal distributions of the random variables is itself solvable in polynomial time (Meilijson and Nádas, 1979; Bertsimas et al., 2004; Mak et al., 2015; Chen et al., 2022). For example, given nonnegative weights $c_1, \ldots, c_n$ and a set $\mathcal{Y} \subseteq \{0,1\}^n$, the deterministic k-sum combinatorial optimization problem for a fixed $k \in [n]$ is formulated as:

$$\min_{\boldsymbol{y} \in \mathcal{Y} \subseteq \{0,1\}^n} \max_{S \subseteq [n]:|S| \leq k} \sum_{i \in S} c_i y_i.$$

For $k = 1$, this reduces to a bottleneck combinatorial optimization problem. Suppose the weights $\tilde{c}_i$ are random where $\tilde{c}_i = c_i$ with probability x_i and $\tilde{c}_i = 0$ with probability $1 - x_i$. Define the set function:

$$f_{\boldsymbol{y}}(S) = \max_{T \subseteq S, |T| \leq k} \sum_{i \in T} c_i y_i, \ \forall S \subseteq [n].$$

For a fixed $\boldsymbol{y} \in \mathcal{Y} \subseteq \{0,1\}^n$, the function $f_{\boldsymbol{y}} : 2^{[n]} \to \mathbb{R}_+$ is a monotone submodular function and the corresponding value of the concave closure $f^+_{\boldsymbol{y}}(\boldsymbol{x})$ (worst-case expected cost with given marginals) is computable in polynomial time (Calinescu et al., 2007; Natarajan, 2021). The distributionally robust k-sum combinatorial optimization problem is then formulated as:

$$\min_{\boldsymbol{y} \in \mathcal{Y} \subseteq \{0,1\}^n} f^+_{\boldsymbol{y}}(\boldsymbol{x}).$$

The optimal solution for this problem (denoted by $\boldsymbol{y}^+$) can be found in polynomial time under the assumption that optimizing a linear function over $\mathcal{Y}$ is possible in polynomial time. If the random weights are mutually independent, the stochastic optimization problem is formulated as:

$$\min_{\boldsymbol{y} \in \mathcal{Y} \subseteq \{0,1\}^n} F_{\boldsymbol{y}}(\boldsymbol{x}).$$

Using Theorem 1, we obtain a $e/(e-1)$ approximation algorithm for the stochastic optimization problem since:

$$F_{\boldsymbol{y}^+}(\boldsymbol{x}) \leq f^+_{\boldsymbol{y}^+}(\boldsymbol{x}) = \min_{\boldsymbol{y} \in \mathcal{Y} \subseteq \{0,1\}^n} f^+_{\boldsymbol{y}}(\boldsymbol{x}) \leq (e/(e-1)) \min_{\boldsymbol{y} \in \mathcal{Y} \subseteq \{0,1\}^n} F_{\boldsymbol{y}}(\boldsymbol{x}).$$

Now suppose the random weights are pairwise independent. The distributionally robust k-sum combinatorial optimization problem given a fixed marginal probability vector $\boldsymbol{x}$ and assuming pairwise independence is formulated as:

$$\min_{\boldsymbol{y} \in \mathcal{Y} \subseteq \{0,1\}^n} f^{++}_{\boldsymbol{y}}(\boldsymbol{x}).$$

We then obtain a $4/3$ approximation algorithm for this problem in the following two cases: (a) for $k = 1$ and general values of $\boldsymbol{x}$, and (b) for general k and small values of $\boldsymbol{x}$. Specifically:

$$\min_{\boldsymbol{y} \in \mathcal{Y} \subseteq \{0,1\}^n} f_{\boldsymbol{y}}^{++}(\boldsymbol{x}) \leq f_{\boldsymbol{y}^+}^{++}(\boldsymbol{x}) \leq 4/3 \min_{\boldsymbol{y} \in \mathcal{Y} \subseteq \{0,1\}^n} f_{\boldsymbol{y}}^{++}(\boldsymbol{x}).$$

The precise conditions on the marginal probabilities under which the approximation guarantee in case (b) holds is provided in Theorems 9 and 10. Thus we obtain improved approximation guarantees for a class of distributionally robust optimization problems where only weaker notions of independence are assumed to hold.

2. *Properties of the continuous relaxation of the weighted coverage function*: Let $w_1, \ldots, w_m$ denote nonnegative weights and $T_1, \ldots, T_n$ be subsets of $[m]$. The weighted coverage function $f : 2^{[n]} \to \mathbb{R}_+$ is defined by $f(S) = \sum_{j \in \cup_{i \in S} T_i} w_j$. For each $j \in [m]$, let $U_j = \{i \in [n] \mid j \in T_i\}$ represent the subsets that cover j. The multilinear extension for any $\boldsymbol{x} \in [0,1]^n$ is given by:

$$F(\boldsymbol{x}) = \sum_{j=1}^{m} w_j \mathbb{P}(j \text{ is covered by some } T_i) = \sum_{j=1}^{m} w_j (1 - \prod_{i \in U_j} (1 - x_i)).$$

An upper bound on $F(\boldsymbol{x})$ is obtained by using the concave closure for each individual term:

$$F(\boldsymbol{x}) \leq \sum_{j=1}^{m} w_j f_j^+(\boldsymbol{x}) \text{ where } f_j^+(\boldsymbol{x}) = \min(1, \sum_{i \in U_j} x_i).$$

The upper bound can maximized efficiently over a polytope $\mathcal{P} \subseteq [0,1]^n$ in polynomial time using linear optimization or subgradient methods (Calinescu et al., 2011; Karimi et al., 2017) to obtain an optimal solution $\boldsymbol{x}^+$. Theorem 1 ensures $F(\boldsymbol{x}^+) \geq (1 - 1/e) \max_{\boldsymbol{x} \in \mathcal{P} \subseteq [0,1]^n} F(\boldsymbol{x})$. In prior work, pipage rounding has been applied to $\boldsymbol{x}^+$ to find good solutions to the deterministic combinatorial optimization problem while preserving the quality of the approximation in expectation (Karimi et al., 2017). In fact, the solution $\boldsymbol{x}^+$ possesses an additional performance guarantee, namely if we use the upper pairwise independent extension for each individual term then $\sum_{j=1}^{m} w_j f_j^{++}(\boldsymbol{x}^+) \geq 3/4 \sum_{j=1}^{m} w_j \max_{\boldsymbol{x} \in \mathcal{P} \subseteq [0,1]^n} f_j^{++}(\boldsymbol{x})$.

Next, we discuss a recent result where a tight closed form upper bound was developed for pairwise independent random variables for a specific monotone submodular function.

Theorem 3 *[(Ramachandra and Natarajan, 2021)] For $f(S) = \min(|S|, 1)$ and any $\boldsymbol{x} \in [0,1]^n$, $f^{++}(\boldsymbol{x}) = \min(1, \sum_{i=1}^{n} x_i (1 - \max_{i \in [n]} x_i) + \max_{i \in [n]} x_i^2)$. In this case, $f^+(\boldsymbol{x})/f^{++}(\boldsymbol{x}) \leq 4/3$ and the upper bound is tight for $\sum_{i=1}^{n} x_i = 1$ and $\max_{i \in [n]} x_i = 1/2$.*

For the function $f(S) = \min(|S|, 1)$, the ratio $f^+(\boldsymbol{x})/F(\boldsymbol{x})$ attains the $e/(e-1)$ upper bound for specific values of $\boldsymbol{x}$ and n (see Theorem 1). Theorem 3 shows that in contrast, the ratio $f^+(\boldsymbol{x})/f^{++}(\boldsymbol{x})$ is provably smaller in the worst-case for this specific function. In Appendix A.2, we show that the related ratio of $f^{++}(\boldsymbol{x})/F(\boldsymbol{x})$ in the worst-case can be as large as the original $e/(e-1)$ bound. It is hence a natural question to ask if the bound of $4/3$ on the ratio $f^+(\boldsymbol{x})/f^{++}(\boldsymbol{x})$ generalizes to other monotone submodular functions. In the next two sections, we show that this is true for all monotone submodular functions in two cases: (a) small n (specifically $n = 2$ and 3) with general marginal probabilities, and (b) general n with small marginal probabilities. We discuss some of the technical challenges that arise in the analysis of the pairwise independent correlation gap next.

2.5. Technical challenges

Without loss of generality, assume $f(\emptyset) = 0$ and $f([n]) = 1$ since translating and scaling the function by defining $g(S) = (f(S) - f(\emptyset))/(f([n]) - f(\emptyset))$ preserves monotonicity and submodularity. Define the set of functions of interest as follows:

$$\mathcal{F}_n = \{f : 2^{[n]} \to \mathbb{R}_+ \mid f \text{ is monotone submodular}, f(\emptyset) = 0, f([n]) = 1\}. \tag{4}$$

Computing $f^+(\boldsymbol{x})$ and $f^{++}(\boldsymbol{x})$ for a given $\boldsymbol{x} \in [0,1]^n$ and an arbitrary function $f \in \mathcal{F}_n$ are NP-hard (see Appendix A.3). In particular, the analysis of the ratio is challenging since the optimal probability distributions which attain the values $f^+(\boldsymbol{x})$ and $f^{++}(\boldsymbol{x})$ are not oblivious to the choice of the function $f \in \mathcal{F}_n$. This is in contrast to the multilinear extension $F(\boldsymbol{x})$ and the convex closure $f^-(\boldsymbol{x})$ (Dughmi, 2009). Table 1 provides an example of two such functions $f_1, f_2 \in \mathcal{F}_3$ where the optimal distributions $\theta^+_*(S)$ and $\theta^{++}_*(S)$ that attain the values $f^+(\boldsymbol{x})$ and $f^{++}(\boldsymbol{x})$ change from f_1 to f_2. This makes the analysis challenging since the optimal distributions in the numerator and the denominator of the ratio $f^+(\boldsymbol{x})/f^{++}(\boldsymbol{x})$ are sensitive to the choice of $f \in \mathcal{F}_n$.

Table 1: Optimal distributions change with the function for $(x_1, x_2, x_3) = (1/2, 1/2, 1/2)$.

S	$f_1(S)$	$\theta^+_*(S)$	$\theta^{++}_*(S)$	$f_2(S)$	$\theta^+_*(S)$	$\theta^{++}_*(S)$
$\emptyset$	0	0	0	0	0	1/4
$\{1\}$	1/3	0	1/4	1/3	0	0
$\{2\}$	1/2	0	1/4	1/2	1/2	0
$\{3\}$	3/5	1/2	1/4	1/2	0	0
$\{1,2\}$	3/4	1/2	0	3/4	0	1/4
$\{1,3\}$	4/5	0	0	4/5	1/2	1/4
$\{2,3\}$	5/6	0	0	5/6	0	1/4
$\{1,2,3\}$	1	0	1/4	1	0	0
$f^+(\boldsymbol{x})$	-	27/40	-	-	13/20	-
$f^{++}(\boldsymbol{x})$	-	-	73/120	-	-	143/240

3. Upper bound for small n, general $\boldsymbol{x}$

Our technique of analysis for small values of n is based on explicitly partitioning the probability space into v regions given by $[0,1]^n = R_1 \cup R_2 \cup \cdots \cup R_v$, such that for each region, we are able to compute $f^+(\boldsymbol{x})$ and $f^{++}(\boldsymbol{x})$ or else approximate it well enough to prove the bound. For $n = 2$, it is straightforward to find $v = 2$ regions such that the probability distributions which attain the bound $f^+(\boldsymbol{x})$ in each of these regions are oblivious to the particular $f \in \mathcal{F}_2$. This is because the worst-case distribution is given by two perfectly negatively dependent random variables arising from the intrinsic connection between submodular functions and substitutability (Topkis, 1998). Furthermore for $n = 2$, $f^{++}(\boldsymbol{x})$ is simply $F(\boldsymbol{x})$ and hence $f^+(\boldsymbol{x})/f^{++}(\boldsymbol{x})$ is exactly the original correlation gap. For $n = 3$, the problem is significantly more challenging since pairwise independence is no longer equivalent to mutual independence. Perfect negative dependence among three or more random variables is also not well defined (Joe, 1997) and developing a theory of negative dependence has been a topic of intense research over the past two decades (Pemantle, 2000; Borcea et al., 2009). As illustrated in Table 1, computing $f^{++}(\boldsymbol{x})$ and $f^{+}(\boldsymbol{x})$ is not oblivious to the choice of $f \in \mathcal{F}_3$. Furthermore, for the function f_1, inputs 1 and 2 are perfectly negatively dependent to input 3, making 1 and 2 perfectly positively dependent. To overcome the challenge in characterizing the optimal distributions, we partition the probability space $[0,1]^3$ into $v = 5$ regions. In each region, we compute valid lower bounds on $f^{++}(\boldsymbol{x})$ by generating feasible pairwise independent distributions for the primal linear program in (3) and valid upper bounds on $f^{+}(\boldsymbol{x})$ by generating feasible solutions for the dual linear program in (2). While these solutions are not always optimal, they turn out to be sufficient to prove the $4/3$ bound. The $n = 3$ case is important since it illustrates a fundamental difference in the behavior of the ratios $f^{+}(\boldsymbol{x})/F(\boldsymbol{x})$ and $f^{+}(\boldsymbol{x})/f^{++}(\boldsymbol{x})$ for monotone, submodular functions that is explained by the difference in pairwise and mutual independence (Bernstein, 1946). We next provide a characterization of the set of extremal monotone submodular functions for $n = 2$ and 3 and two key inequalities which are used in the proofs.

3.1. Extremal monotone submodular functions and two key ineuqalities

The set $\mathcal{F}_n$ in (4) is polytope contained in $[0,1]^{2^n}$. The extreme points of this polytope define extremal monotone submodular functions and are characterized in the lemma below for $n = 2$ and 3 (Shapley, 1971; Rosenmuller and Weidner, 1974; Kashiwabara, 2000).

Lemma 4 *The extremal set functions in $\mathcal{F}_2$ are given by $\mathcal{E}(\mathcal{F}_2) = \{e_1, e_2, e_3\}$ where:*

$$e_1 = (0,1,0,1), \quad e_2 = (0,0,1,1), \quad e_3 = (0,1,1,1),$$

with the entries of the vector denoting the values $(f(\emptyset), f(1), f(2), f(1,2))$. The extremal set functions in $\mathcal{F}_3$ are given by $\mathcal{E}(\mathcal{F}_3) = \{E_1, E_2, E_3, E_4, E_5, E_6, E_7, E_8\}$ where:

$$\begin{array}{llll} E_1 = (0,1,0,0,1,1,0,1), & E_2 = (0,0,1,0,1,0,1,1), & E_3 = (0,0,0,1,0,1,1,1), & E_4 = (0,1,1,0,1,1,1,1), \\ E_5 = (0,1,0,1,1,1,1,1), & E_6 = (0,0,1,1,1,1,1,1), & E_7 = (0,1,1,1,1,1,1,1), & E_8 = (0,\frac{1}{2},\frac{1}{2},\frac{1}{2},1,1,1,1), \end{array}$$

with the entries of the vector denoting the values $(f(\emptyset), f(1), f(2), f(3), f(1,2), f(1,3), f(2,3), f(1,2,3))$. In addition, the extreme points of the polytopes $\mathcal{F}_3^1 = \mathcal{F}_3 \cap \{f | f(1) + f(2,3) \geq f(2) + f(1,3), f(1) + f(2,3) \geq f(3) + f(1,2)\}$ are given by $\mathcal{E}(\mathcal{F}_3^1) = \mathcal{E}(\mathcal{F}_3) \backslash \{E_6\}$, the extreme points of $\mathcal{F}_3^2 = \mathcal{F}_3 \cap \{f | f(2) + f(1,3) \geq f(1) + f(2,3), f(2) + f(1,3) \geq f(3) + f(1,2)\}$ are given by $\mathcal{E}(\mathcal{F}_3^2) = \mathcal{E}(\mathcal{F}_3) \backslash \{E_5\}$ and the extreme points of $\mathcal{F}_3^3 = \mathcal{F}_3 \cap \{f | f(3) + f(1,2) \geq f(1) + f(2,3), f(3) + f(1,2) \geq f(2) + f(1,3)\}$ are given by $\mathcal{E}(\mathcal{F}_3^3) = \mathcal{E}(\mathcal{F}_3) \backslash \{E_4\}$.

Proof See Appendix A.4. ■

The following two inequalities (see Appendix A.5) are used in the proofs:

$$\begin{array}{lll} (I_1) & \alpha + \beta - 4\alpha\beta \geq 0, & \forall \alpha, \beta \in [0,1], \alpha + \beta \leq 1 \\ (I_2) & 4\alpha + 4\beta - 4\alpha\beta - 3 \geq 0, & \forall \alpha, \beta \in [0,1], \alpha + \beta \geq 1. \end{array} \tag{5}$$

3.2. Main results for $n = 2$ and $n = 3$

Theorem 5 *For any monotone submodular function $f : 2^{[2]} \to \mathbb{R}_+$ and any $\boldsymbol{x} \in [0,1]^2$:*

$$f^{+}(\boldsymbol{x})/f^{++}(\boldsymbol{x}) \leq 4/3.$$

Proof For $n = 2$, $f^{++}(\boldsymbol{x}) = F(\boldsymbol{x})$ and the proof simply reduces to showing the original correlation gap is itself upper bounded by $4/3$. Here $F(\boldsymbol{x}) = x_1(1-x_2)f(1) + x_2(1-x_1)f(2) + x_1x_2$. It is straightforward to find $f^+(\boldsymbol{x})$ in closed form based on a partition of the probabilities into $v = 2$ regions (see Figure 1). Table 2 provides the distributions $\theta^+_{1,*}(S)$ and $\theta^+_{2,*}(S)$ for the primal linear program (2) that are optimal in regions R_1 and R_2 respectively. Here the optimal distributions are oblivious to the specific choice of $f \in \mathcal{F}_2$. The optimality can be verified by observing that $(\lambda_0, \lambda_1, \lambda_2) = (0, f(1), f(2))$ is dual feasible in region R_1, $(\lambda_0, \lambda_1, \lambda_2) = (f(1) + f(2) - 1, 1 - f(2), 1 - f(1))$ is dual feasible in region R_2 and the dual objective values match the primal objective values for $\theta^+_{1,*}(S)$ and $\theta^+_{2,*}(S)$ in these regions respectively. Let $\delta = 4F(\boldsymbol{x}) - 3f^+(\boldsymbol{x})$. For $x_1 + x_2 \leq 1$ (region R_1), $f^+(\boldsymbol{x}) = x_1 f(1) + x_2 f(2)$. In this case $\delta = x_1 f(1) + x_2 f(2) + 4x_1x_2(1 - f(1) - f(2))$. The minimum value of this expression over all functions in $\mathcal{F}_2$ is attained at one of the extremal set functions e_1, e_2 or e_3. This gives the expression $\delta = \min(x_1, x_2, x_1 + x_2 - 4x_1x_2)$. Using the inequality I_1, we get the desired result in R_1. For $x_1 + x_2 > 1$ (region R_2), $f^+(\boldsymbol{x}) = (1-x_2)f(1) + (1-x_1)f(2) + x_1 + x_2 - 1$. In this case, $\delta = (4x_1 + 3x_2 - 4x_1x_2 - 3)f(1) + (3x_1 + 4x_2 - 4x_1x_2 - 3)f(2) + 4x_1x_2 - 3x_1 - 3x_2 + 3$. The minimum value of this expression over all functions in $\mathcal{F}_2$ is attained at one of the extremal set functions e_1, e_2 or e_3 which gives the expression $\delta = \min(x_1, x_2, 4x_1 + 4x_2 - 4x_1x_2 - 3)$. Using the inequality I_2, we get the desired result in R_2.

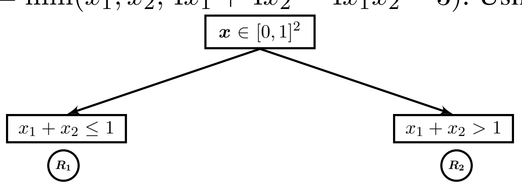

Figure 1: Partition of $[0,1]^2$ into $v = 2$ regions

Table 2: Optimal distributions that attain $f^+(\boldsymbol{x})$

S	$\theta^+_{1,*}(S)$	$\theta^+_{2,*}(S)$
$\emptyset$	$1 - x_1 - x_2$	0
$\{1\}$	x_1	$1 - x_2$
$\{2\}$	x_2	$1 - x_1$
$\{1,2\}$	0	$x_1 + x_2 - 1$
Region	R_1	R_2

■

Theorem 5 shows that for all monotone submodular functions defined on two variables:

$$f^+(\boldsymbol{x})/f^{++}(\boldsymbol{x}) = f^+(\boldsymbol{x})/F(\boldsymbol{x}) \leq 4/3.$$

The bound is tight (simply choose $f(S) = \min(|S|, 1)$ and $(x_1, x_2) = (1/2, 1/2)$). Our next theorem shows that this bound holds even for $n = 3$ where pairwise independence and mutual independence are different.

Theorem 6 *For any monotone submodular function $f : 2^{[3]} \to \mathbb{R}_+$ and any $\boldsymbol{x} \in [0,1]^3$:*

$$f^+(\boldsymbol{x})/f^{++}(\boldsymbol{x}) \leq 4/3.$$

Proof The proof is based on finding lower bounds on $f^{++}(\boldsymbol{x})$ (denoted by $\underline{f}^{++}(\boldsymbol{x})$) and upper bounds on $f^+(\boldsymbol{x})$ (denoted by $\overline{f}^+(\boldsymbol{x})$). Our choice of bounds verify the inequality $4\underline{f}^{++}(\boldsymbol{x}) - 3\overline{f}^+(\boldsymbol{x}) \geq 0$ for all $f \in \mathcal{F}_3$ which in turn guarantee $4f^{++}(\boldsymbol{x}) - 3f^+(\boldsymbol{x}) \geq 0$ for all $f \in \mathcal{F}_3$. To find the bounds, we partition the space $[0,1]^3$ of all possible marginal probabilities into regions where $\underline{f}^{++}(\boldsymbol{x})$ and $\overline{f}^+(\boldsymbol{x})$ are linear in the function values f. This is useful since we can then verify the inequality $4\underline{f}^{++}(\boldsymbol{x}) - 3\overline{f}^+(\boldsymbol{x}) \geq 0$ for all functions in $\mathcal{F}_3$ by simply verifying it for the extremal set functions in $\mathcal{F}_3$. We outline the construction of the bounds and the main steps of the proof next (the omitted parts are provided in the appendix).

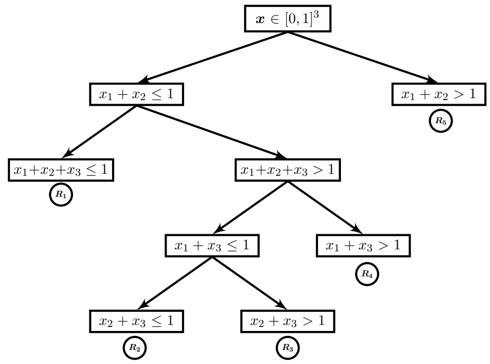

Figure 2: Partition of $[0, 1]^3$ into $v = 5$ regions

Table 3: Pairwise independent distributions.

S	$\theta_1^{++}(S)$	$\theta_2^{++}(S)$
$\emptyset$	$(1-x_3)(1-x_1-x_2)$	$(1-x_2)(1-x_3)$
$\{1\}$	$x_1(1-x_3)$	0
$\{2\}$	$x_2(1-x_3)$	$(1-x_3)(x_2-x_1)$
$\{3\}$	$x_3(1-x_1-x_2)+x_1x_2$	$(1-x_2)(x_3-x_1)$
$\{1,2\}$	0	$x_1(1-x_3)$
$\{1,3\}$	$x_1(x_3-x_2)$	$x_1(1-x_2)$
$\{2,3\}$	$x_2(x_3-x_1)$	$x_2x_3-x_1(x_2+x_3-1)$
$\{1,2,3\}$	x_1x_2	$x_1(x_2+x_3-1)$
REGION	R_1 TO R_4	R_5

1. We partition $[0, 1]^3$ into $v = 5$ regions as follows: Sort the marginal probabilities (x_1, x_2, x_3) in increasing value such that $0 \leq x_1 \leq x_2 \leq x_3 \leq 1$. Then $x_1 + x_2 \leq x_1 + x_3 \leq x_2 + x_3 \leq x_1 + x_2 + x_3$. Divide $[0, 1]^3$ based on the relative position of the number 1 among these sums (see Figure 2). For example region R_2 is defined by $x_2 + x_3 \leq 1 < x_1 + x_2 + x_3$. In each of these regions, we generate primal feasible distributions to find lower bounds for $f^{++}(\boldsymbol{x})$ and dual feasible solutions to find upper bounds for $f^+(\boldsymbol{x})$.

2. We evaluate $\underline{f}^{++}(\boldsymbol{x})$ by considering two pairwise independent feasible distributions for the primal linear program in (3) (see Table 3). It is easy to verify the probability distribution θ_1^{++} is feasible if $x_1 + x_2 \leq 1$ (in regions R_1 to R_4). Similarly it is easy to verify the probability distribution θ_2^{++} is feasible if $x_1 + x_2 > 1$. The set of all pairwise independent distributions for three Bernoulli random variables is completely characterized in Derriennic and Kłopotowski (2000). For our purposes, these two distributions suffice.

3. We evaluate $\overline{f}^+(\boldsymbol{x})$ by creating feasible solutions for the dual linear program in (2). For region R_1 (small probabilities), we use the dual feasible solution $(\lambda_0, \lambda_1, \lambda_2, \lambda_3) = (0, f(1), f(2), f(3))$ which gives the upper bound $\overline{f}^+(\boldsymbol{x}) = x_1 f(1) + x_2 f(2) + x_3 f(3)$. For region R_5 (large probabilities), we use the dual feasible solution $(\lambda_0, \lambda_1, \lambda_2, \lambda_3) = (f(1, 2) + f(1, 3) + f(2, 3) - 2, 1 - f(2, 3), 1 - f(1, 3), 1 - f(1, 2))$ which gives the upper bound $\overline{f}^+(\boldsymbol{x}) = x_1 + x_2 + x_3 - 2 + (1 - x_3)f(1, 2) + (1 - x_2)f(1, 3) + (1 - x_1)f(2, 3)$. It is easy to check these solutions are dual feasible for all $f \in \mathcal{F}_3$. For regions R_2 to R_4 (moderate probabilities), the details of the dual feasible solutions are provided in Appendix A.6.

4. Let $\delta = 4\underline{f}^{++}(\boldsymbol{x}) - 3\overline{f}^+(\boldsymbol{x})$. For regions R_1 and R_5, we need to verify $\delta \geq 0$ at each of the extremal set functions E_1 to E_8 provided in Lemma 4.

   (a) Region R_1: Here $\overline{f}^+(\boldsymbol{x}) = x_1 f(1) + x_2 f(2) + x_3 f(3)$. For $E_1 = (0, 1, 0, 0, 1, 1, 0, 1)$, we obtain:

$$\begin{aligned} \delta &= 4(\theta_1^{++}(1) + \theta_1^{++}(1,2) + \theta_1^{++}(1,3) + \theta_1^{++}(1,2,3)) - 3x_1 \\ &= 4(x_1(1-x_3) + x_1(x_3-x_2) + x_1x_2) - 3x_1 \\ &= x_1 \\ &\geq 0. \end{aligned}$$

Similar computations for E_2 and E_3 give $\delta = x_2 \geq 0$ and $\delta = x_3 \geq 0$. For $E_4 = (0,1,1,0,1,1,1,1)$, we obtain:

$$\begin{aligned}
\delta &= 4(\theta_1^{++}(1) + \theta_1^{++}(2) + \theta_1^{++}(1,2) + \theta_1^{++}(1,3) + \theta_1^{++}(2,3) + \theta_1^{++}(1,2,3)) \\
&\quad -3(x_1 + x_2) \\
&= 4((x_1 + x_2)(1 - x_3) + x_1(x_3 - x_2) + x_2(x_3 - x_1) + x_1 x_2) - 3(x_1 + x_2) \\
&= x_1 + x_2 - 4x_1 x_2 \\
&\geq 0 \text{ [using } I_1 \text{ since } x_1 + x_2 \leq 1 \text{ in } R_1].
\end{aligned}$$

Similar computations for E_5 and E_6 give $\delta = x_1 + x_3 - 4x_1x_3$ and $x_2 + x_3 - 4x_2x_3$ which are nonnegative since $x_1 + x_3 \leq 1$ and $x_2 + x_3 \leq 1$ in region R_1. For E_7, we get:

$$\begin{aligned}
\delta &= 4(\theta_1^{++}(1) + \theta_1^{++}(2) + \theta_1^{++}(3) + \theta_1^{++}(1,2) + \theta_1^{++}(1,3) + \theta_1^{++}(2,3) + \theta_1^{++}(1,2,3)) \\
&\quad -3(x_1 + x_2 + x_3) \\
&= x_1 + x_2 + x_3 - 4x_1(x_2 + x_3) \\
&\geq 0 \text{ [using } I_1 \text{ since } x_1 + x_2 + x_3 \leq 1 \text{ in } R_1 \text{ and by setting } \alpha = x_1 + x_2, \beta = x_3].
\end{aligned}$$

For E_8, we get:

$$\begin{aligned}
\delta &= 4(\tfrac{1}{2}\theta_1^{++}(1) + \tfrac{1}{2}\theta_1^{++}(2) + \tfrac{1}{2}\theta_1^{++}(3) + \theta_1^{++}(1,2) + \theta_1^{++}(1,3) + \theta_1^{++}(2,3) + \theta_1^{++}(1,2,3)) \\
&\quad -3(x_1 + x_2 + x_3) \\
&= \tfrac{1}{2}(x_1 + x_2 + x_3 - 4x_1x_2) \\
&\geq 0 \text{ [using } I_1 \text{ since } x_1 + x_2 \leq 1 \text{ in conjunction with } x_3 \geq 0].
\end{aligned}$$

Hence $\delta \geq 0$ for all $f \in \mathcal{F}_3$ and marginal probabilities defined by region R_1.

(b) Region R_5: Here $\overline{f}^+(\boldsymbol{x}) = x_1 + x_2 + x_3 - 2 + (1 - x_3)f(1,2) + (1 - x_2)f(1,3) + (1 - x_1)f(2,3)$. For $E_1 = (0,1,0,0,1,1,0,1)$, we obtain:

$$\begin{aligned}
\delta &= 4(\theta_2^{++}(1) + \theta_2^{++}(1,2) + \theta_2^{++}(1,3) + \theta_2^{++}(1,2,3)) - 3x_1 \\
&= 4(x_1(1 - x_3) + x_1(1 - x_2) + x_1(x_2 + x_3 - 1)) - 3x_1 \\
&= x_1 \\
&\geq 0.
\end{aligned}$$

Similarly for E_2 and E_3, we get $\delta = x_2$ and $\delta = x_3$ which is nonnegative. For $E_4 = (0,1,1,0,1,1,1,1)$, we obtain:

$$\begin{aligned}
\delta &= 4(\theta_1^{++}(1) + \theta_1^{++}(2) + \theta_1^{++}(1,2) + \theta_1^{++}(1,3) + \theta_1^{++}(2,3) + \theta_1^{++}(1,2,3)) - 3 \\
&= 4((1 - x_3)(x_2 - x_1) + x_1(2 - x_3 - x_2) + x_2x_3) - 3 \\
&= 4x_1 + 4x_2 - 4x_1x_2 - 3 \\
&\geq 0 \text{ [using } I_2 \text{ since } x_1 + x_2 \geq 1 \text{ in } R_5].
\end{aligned}$$

Similar computations for E_5 and E_6 give $\delta = 4x_1 + 4x_3 - 4x_1x_3 - 3$ and $4x_2 + 4x_3 - 4x_2x_3 - 3$ which are nonnegative since $x_1 + x_3 \geq 1$ and $x_2 + x_3 \geq 1$ in region R_5. For E_7, we get $\delta = 4x_2 + 4x_3 - 4x_2x_3 - 3$ which is nonnegative. Finally for E_8, we get $\delta = 4x_1 + 2x_2 + 2x_3 - 2x_1x_3 - 2x_1x_2 - 3$ which is nonnegative from inequality I_2 by setting $\alpha = x_1$ and $\beta = (x_2 + x_3)/2$ where $\alpha + \beta \geq 1$ for region R_5.

The proofs for the remaining three regions $R_2 - R_4$ is a bit more involved and can be found in Appendix A.6.

Table 4: Pairwise independent distribution.

S	$\theta(S)$
$\emptyset$	$\left(1-\sum_{i=1}^{n-1} x_i\right)(1-x_n)$
$\{1\}$	$x_1(1-x_n)$
$\{2\}$	$x_2(1-x_n)$
$\vdots$	$\vdots$
$\{n-1\}$	$x_{n-1}(1-x_n)$
ALL OTHER S WITH $n \notin S$	0
ALL S WITH $n \in S$	$\theta(S)$

■

For $n = 3$, consider the function $f(S) = \min(|S|, 1)$ with marginal probabilities $\boldsymbol{x} = (1/3, 1/3, 1/3)$. Here $f^+(\boldsymbol{x}) = 1$ and $F(\boldsymbol{x}) = 19/27$ which gives $f^+(\boldsymbol{x})/F(\boldsymbol{x}) = 27/19 \approx 1.421$. Theorem 6 shows in contrast, the ratio $f^+(\boldsymbol{x})/f^{++}(\boldsymbol{x})$ can never be larger than $4/3 \approx 1.333$ for all functions $f \in \mathcal{F}_3$. This illustrates that even for $f \in \mathcal{F}_3$, the upper pairwise independent bound is closer to the concave closure than the multilinear extension.

Lemma 7 *Let $f = f_1 + \ldots + f_m : 2^{[n_1+\ldots+n_m]} \to \mathbb{R}_+$, where the set functions f_i are monotone submodular functions defined on disjoint ground sets of size $n_i \leq 3$. Then $f^+(\boldsymbol{x})/f^{++}(\boldsymbol{x}) \leq 4/3$ for all $\boldsymbol{x} \in [0,1]^n$.*

Proof The proof can be found in Appendix A.7. ■

4. Upper bound for general n, small $\boldsymbol{x}$

Our technique of analysis for general n is based on a recent construction of a pairwise independent distribution provided in Ramachandra and Natarajan (2021) that was shown to be optimal for the function $f(S) = \min(|S|, 1)$. We will see in this section that this distribution suffices to prove the $4/3$ upper bound when the marginal probabilities are small for all functions in $\mathcal{F}_n$.

Theorem 8 *[Ramachandra and Natarajan (2021)] For any $\boldsymbol{x} \in [0,1]^n$, sort the values as $0 \leq x_1 \leq \ldots \leq x_n \leq 1$. Suppose $\sum_{i=1}^{n-1} x_i \leq 1$. Then there always exists a pairwise independent distribution with the marginal probability vector $\boldsymbol{x}$ of the form shown in Table 4. Specifically, the joint probabilities are given by: (a) $\theta(\emptyset) = (1 - \sum_{i=1}^{n-1} x_i)(1 - x_n)$, (b) $\theta(i) = x_i(1 - x_n)$ for all $i < n$, (c) $\theta(S) = 0$ for all other S with $n \notin S$, and (d) $\theta(S) \geq 0$ for all $S \ni n$ such that $\sum_{S:n\in S} \theta(S) = x_n$, $\sum_{S:i,n\in S} \theta(S) = x_i x_n$ for $i < n$, $\sum_{S:i,j,n\in S} \theta(S) = x_i x_j$ for all $i < j < n$.*

This brings us to the following theorem.

Theorem 9 *For any monotone submodular function $f : 2^{[n]} \to \mathbb{R}_+$ and any $\boldsymbol{x} \in [0,1]^n$ where $0 \leq x_1 \leq \ldots \leq x_n \leq 1$, under the assumption that $\sum_{i=1}^{n-1} x_i \leq 1$ and $x_n \leq 1/4$, $f^+(\boldsymbol{x})/f^{++}(\boldsymbol{x}) \leq 4/3$.*

Proof To compute the lower bound on $f^{++}(\boldsymbol{x})$, we use the pairwise independent distribution provided in Theorem 8. This distribution is feasible when $\sum_{i=1}^{n-1} x_i \leq 1$ and was shown to be optimal in Ramachandra and Natarajan (2021) for the set function $f(S) = \min(|S|, 1)$. For general functions $f \in \mathcal{F}_n$, from the non-decreasing property of f, we have $f(S) \geq f(n)$ for all $S \ni n$. This implies that:

$$f^{++}(\boldsymbol{x}) \geq \underline{f}^{++}(\boldsymbol{x}) = \sum_{i=1}^{n-1} x_i(1-x_n)f(i) + x_n f(n).$$

For the upper bound, we use the dual feasible solution $\lambda_0 = 0$ and $\lambda_i = f(i)$ for all $i \in [n]$. This is dual feasible for the linear program in (2) for all functions $f \in \mathcal{F}_n$ and gives:

$$f^{+}(\boldsymbol{x}) \leq \overline{f}^{+}(\boldsymbol{x}) = \sum_{i=1}^{n} x_i f(i).$$

Together we get

$$\begin{aligned}
\delta &= 4\underline{f}^{++}(\boldsymbol{x}) - 3\overline{f}^{+}(\boldsymbol{x}) \\
&\geq 4\sum_{i=1}^{n-1} x_i(1-x_n)f(i) + 4x_n f(n) - 3\sum_{i=1}^{n} x_i f(i) \\
&= \sum_{i=1}^{n-1} x_i(1-4x_n)f(i) + x_n f(n) \\
&\geq 0 \text{ [since } \textstyle\sum_{i=1}^{n-1} x_i \leq 1 \text{ and } x_n \leq 1/4].
\end{aligned}$$

■

For the function $f(S) = \min(|S|, 1)$ and $\boldsymbol{x} = (1/n, \ldots, 1/n)$, the correlation gap $f^{+}(\boldsymbol{x})/F(\boldsymbol{x}) \uparrow e/(e-1)$ as $n \uparrow \infty$. Theorem 9 shows that in the regime of small marginal probabilities, for any value of n, the maximum value of the ratio $f^{+}(\boldsymbol{x})/f^{++}(\boldsymbol{x})$ is provably smaller. The next result shows that with identical marginal probabilities, it is possible to replace the conditions on the marginal probability vector $\boldsymbol{x}$ in Theorem 9 with a slightly weaker condition and still obtain an upper bound of $4/3$.

Theorem 10 *Assume $x_i = x \in [0,1]$ for all $i \in [n]$. Then $f^{+}(\boldsymbol{x})/f^{++}(\boldsymbol{x}) \leq 4/3$ for any monotone submodular function $f : 2^{[n]} \to \mathbb{R}_+$ with $x \leq 1/(n-1)$.*

Proof The proof can be found in Appendix A.8. ■

5. Conclusions and a conjecture

In this paper, we provided a new definition of correlation gap for set functions with pairwise independent random input and proved a $4/3$ upper bound in several cases for monotone submodular set functions. Specifically, the results in this paper in conjunction with Ramachandra and Natarajan (2021) show that the bound of $4/3$ holds in the following cases: (a) for $f(S) = \min(|S|, 1)$ for all n and all $\boldsymbol{x} \in [0,1]^n$, (b) for $n = 2$ and $n = 3$ for all $f \in \mathcal{F}_n$ and all $\boldsymbol{x} \in [0,1]^n$, and (c) for small values of $\boldsymbol{x}$ for all n and all $f \in \mathcal{F}_n$. This brings us to the following conjecture:

Conjecture 11 *For any monotone submodular function $f : 2^{[n]} \to \mathbb{R}_+$ and any $\boldsymbol{x} \in [0,1]^n$, the pairwise independent correlation gap satisfies $f^{+}(\boldsymbol{x})/f^{++}(\boldsymbol{x}) \leq 4/3$.*

References

S. Agrawal, Y. Ding, A. Saberi, and Y. Ye. Price of correlations in stochastic optimization. *Operations Research*, 1(60):150–162, 2012.

B. Assarf, E. Gawrilow, K. Herr, M. Joswig, B. Lorenz, A. Paffenholz, and T. Rehn. Computing convex hulls and counting integer points with `polymake`. *Mathematical Programming Computation*, 9(1):1–38, 2017.

F. Bach. Learning with submodular functions: A convex optimization perspective. *Foundations and Trends in Machine Learning*, 6(2-3):145–373, 2013.

S.N. Bernstein. *Theory of Probability*. Moscow-Leningrad, 1946.

D. Bertsimas, K. Natarajan, and C-P. Teo. Probabilistic combinatorial optimization: Moments, semidefinite programming, and asymptotic bounds. *SIAM Journal on Optimization*, 15(1):185–209, 2004.

J. Borcea, P. Branden, and T. M. Liggett. Negative dependence and the geometry of polynomials. *Journal of the American Mathematical Society*, 22(2):521–567, 2009.

G. Calinescu, C. Chekuri, M. Pál, and J. Vondrák. Maximizing a monotone submodular function subject to a matroid constraint. *Proceedings of the 12th IPCO Conference*, pages 182–196, 2007.

G. Calinescu, C. Chekuri, M. Pál, and J. Vondrák. Maximizing a monotone submodular function subject to a matroid constraint. *SIAM Journal on Computing*, 40(6):1740–1766, 2011.

S. Chawla, J. D. Hartline, D. L. Malec, and B. Sivan. Multi-parameter mechanism design and sequential posted pricing. In *Proceedings of the forty-second ACM symposium on Theory of computing*, pages 311–320, 2010.

C. Chekuri and V. Livanos. On submodular prophet inequalities and correlation gap. In *Proceedings of SAGT 2021*, 2021.

C. Chekuri, J. Vondrak, and R. Zenklusen. Submodular function maximization via the multilinear relaxation and contention resolution schemes. *SIAM Journal on Computing*, 43(6):1831–1879, 2014.

L. Chen, W. Ma, K. Natarajan, D. Simchi-Levi, and Z. Yan. Distributionally robust linear and discrete optimization with marginals. *Operations Research*, 70(3):1822–834, 2022.

Y. Derriennic and A. Kłopotowski. On Bernstein's example of three pairwise independent random variables. *Sankhyā: The Indian Journal of Statistics, Series A*, pages 318–330, 2000.

S. Dughmi. Submodular functions: Extensions, distributions, and algorithms. a survey, 2009.

M. Feldman, O. Svensson, and R. Zenklusen. Online contention resolution schemes with application to bayesian selection problems. *SIAM Journal on Computing*, 50(2):255–300, 2021.

S. Geisser and N. Mantel. Pairwise independence of jointly dependent variables. *The Annals of Mathematical Statistics*, 33(1):290–291, 1962.

H. Joe. *Multivariate Models and Multivariate Dependence Concepts*. CRC Press, 1997.

M. R. Karimi, M. Lucic, H. Hassani, and A. Krause. Stochastic submodular maximization: The case of coverage functions. In *31st Conference on Neural Information Processing Systems*, pages 6856–6866, 2017.

H. Karloff and Y. Mansour. On construction of k-wise independent random variables. In *Proceedings of the 26th Annual ACM Symposium on Theory of Computing*, pages 564–573, 1994.

K. Kashiwabara. Extremality of submodular functions. *Theoretical Computer Science*, 235:239–256, 2000.

D. Koller and N. Meggido. Construcing small sample spaces satisfying given constraints. *SIAM Journal on Discrete Mathematics*, 7(2):260–274, 1994.

A. Krause and S. Jegelka. Submodularity in machine learning: New directions. *ICML Tutorial*, 2013.

M. Luby and A. Wigderson. Pairwise independence and derandomization. *Foundations and Trends in Theoretical Computer Science*, 1(4):239–201, 2005.

H-Y. Mak, Y. Rong, and J. Zhang. Appointment scheduling with limited distributional information. *Management Science*, 61:316–334, 2015.

I. Meilijson and A. Nádas. Convex majorization with an application to the length of critical paths. *Journal of Applied Probability*, 16(3):671–677, 1979.

K. Natarajan. *Optimization with Marginals and Moments*. Dynamic Ideas LLC, Belmont, Massachusetts, 2021.

R. Pemantle. Towards a theory of negative dependence. *Journal of Mathematical Physics*, 41: 1371–1390, 2000.

A. Ramachandra and K. Natarajan. Tight probability bounds with pairwise independence *Preprint avaliable on arXiv (2006.00516)*. 2021.

J. Rosenmuller and H. G. Weidner. Extreme convex set functions with finite carrier: General theory. *Discrete Mathematics*, 10:384–382, 1974.

A. Rubinstein and S. Singla. Combinatorial prophet inequalities. In *Proceedings of 28th Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 1671–1687, 2017.

E. Schneidman, M. J. Berry, R., and W. Bialek. Weak pairwise correlations imply strongly correlated network states in a neural population. *Nature*, (440):1007–1012, 2006.

L. S. Shapley. Core of convex games. *International Journal of Game Theory*, 1:11–26, 1971.

M. Staib, B. Wilder, and S. Jegelka. Distributionally robust submodular maximization. In *Proceedings of the 22nd International Conference on Artificial Intelligence and Statistics*, 2019.

D. M. Topkis. *Supermodularity and Complementarity*. Princeton University Press, Princeton, New Jersey, 1998.

A. Wigderson. The amazing power of pairwise independence (abstract). In *Proceedings of the Twenty-Sixth Annual ACM Symposium on Theory of Computing, Montréal, Québec, Canada*, pages 645–647, 1994.

Q. Yan. Mechanism design via correlation gap. In *Proceedings of the 22th ACM-SIAM SODA*, pages 710–719, 2011.

Appendix A. Appendix

A.1. Examples for $n = 2$ where the correlation gap is unbounded

Let $(f(\emptyset), f(1), f(2), f(1,2)) = (0, \delta, 0, 0)$ and $(x_1, x_2) = (\epsilon, 1-\epsilon)$. For any $\delta > 0$, this function is non-monotone submodular where $f^+(\boldsymbol{x}) = \delta\epsilon$ and $F(\boldsymbol{x}) = \delta\epsilon^2$. As $\epsilon \downarrow 0$, $f^+(\boldsymbol{x})/F(\boldsymbol{x}) \uparrow \infty$ (Rubinstein and Singla, 2017). Let $(f(\emptyset), f(1), f(2), f(1,2)) = (0, 0, 0, \delta)$ and $(x_1, x_2) = (\epsilon, \epsilon)$. For any $\delta > 0$, this function is monotone but not submodular where $f^+(\boldsymbol{x}) = \delta\epsilon$ and $F(\boldsymbol{x}) = \delta\epsilon^2$. As $\epsilon \downarrow 0$, $f^+(\boldsymbol{x})/F(\boldsymbol{x}) \uparrow \infty$.

A.2. Bound on $f^{++}(\boldsymbol{x})/F(\boldsymbol{x})$

For monotone submodular functions, we have $1 \leq f^{++}(\boldsymbol{x})/F(\boldsymbol{x}) \leq f^+(\boldsymbol{x})/F(\boldsymbol{x}) \leq e/(e-1)$. We construct an example where the ratio $f^{++}(\boldsymbol{x})/F(\boldsymbol{x})$ can be as large as $f^+(\boldsymbol{x})/F(\boldsymbol{x})$ and equal $e/(e-1)$. Let $f(S) = \min(|S|, 1)$ with $\boldsymbol{x} = (1/n, \ldots, 1/n)$. The value $f^+(\boldsymbol{x}) = 1$ is attained by the distribution supported on n points where $\theta(i) = 1/n$ for all $i \in [n]$ (all other probabilities are zero). Furthermore $F(\boldsymbol{x}) = 1 - (1 - 1/n)^n$. Now consider a distribution supported on $n+2$ points where $\theta(\emptyset) = 1/n - 1/n^2$, $\theta(i) = 1/n - 1/n^2$ for all $i \in [n]$ and $\theta([n]) = 1/n^2$. It is easy to see this distribution is pairwise independent and the marginal probabilities are given by $\boldsymbol{x} = (1/n, \ldots, 1/n)$. Hence $f^{++}(\boldsymbol{x}) \geq 1 - 1/n + 1/n^2$ (Ramachandra and Natarajan, 2021). As $n \uparrow \infty$, we get $f^{++}(\boldsymbol{x})/F(\boldsymbol{x}) \geq e/(e-1)$ and hence the bound is attained.

A.3. NP-hardness of computing $f^+(\boldsymbol{x})$ and $f^{++}(\boldsymbol{x})$

The proof of the NP-hardness of computing $f^+(\boldsymbol{x})$ is provided in Agrawal et al. (2012) and Dughmi (2009). The proof in Agrawal et al. (2012) can be extended to showing the NP-hardness of computing $f^{++}(\boldsymbol{x})$. The separation problem for the dual linear program in (3) is formulated as:

$$\max_{S \subseteq [n]} f(S) - \sum_{i \in S} \lambda_i - \sum_{i<j \in S} \lambda_{ij}.$$

Given a graph $G = (V, E)$, define for any $S \subseteq V$, the value $f(S)$ as two times the number of edges that have at least one end point in S. For each $i \in V$, let λ_i be the number of edges incident to vertex i and set all λ_{ij} values to be zero. Then the dual separation problem reduces to solving a MAX CUT problem which is NP-hard. From the equivalence of separation and optimization, computing $f^{++}(\boldsymbol{x})$ is in turn NP-hard.

A.4. Proof of Lemma 4

The polytope $\mathcal{F}_2$ is given by the values of the vector $(f(\emptyset), f(1), f(2), f(1,2))$ satisfying the following linear constraints:

$$f(1) + f(2) \geq 1, 0 \leq f(1) \leq 1, 0 \leq f(2) \leq 1, f(\emptyset) = 0, f(1,2) = 1.$$

The polytope $\mathcal{F}_3$ is given by the values of the vector $(f(\emptyset),\ f(1),\ f(2),\ f(3),\ f(1,2),\ f(1,3),$ $f(2,3),\ f(1,2,3))$ satisfying the following linear constraints:

$$\begin{aligned}
&f(1) + f(2) \geq f(1,2), f(1) + f(3) \geq f(1,3), f(2) + f(3) \geq f(2,3),\\
&f(1,2) + f(1,3) \geq f(1) + 1, f(1,2) + f(2,3) \geq f(2) + 1,\\
&f(1,3) + f(2,3) \geq f(3) + 1, f(1) \leq f(1,2), f(2) \leq f(1,2),\\
&f(1) \leq f(1,3), f(3) \leq f(1,3), f(2) \leq f(2,3), f(3) \leq f(2,3),\\
&f(1,2) \leq 1, f(1,3) \leq 1, f(2,3) \leq 1, f(1) \geq 0, f(2) \geq 0, f(3) \geq 0,\\
&f(\emptyset) = 0, f(1,2,3) = 1.
\end{aligned}$$

One can find all the extreme points of the polytope $\mathcal{F}_n$ by solving for every set of 2^n linearly independent active constraints. If the corresponding solution satisfies the remaining linear constraints describing $\mathcal{F}_n$, it is an extreme point of the polytope. For $n = 2$ and 3 it is easy to do this by enumeration. In practice this can also be verified by using the open source software *polymake* (Assarf et al., 2017) which finds all the extreme points. For the polytopes $\mathcal{F}_3^1$, $\mathcal{F}_3^2$ and $\mathcal{F}_3^3$, we simply add new inequalities to the set $\mathcal{F}_3$. It is straightforward to see that the extreme points E_6, E_5, E_4 violate some of these additional conditions and are thus excluded from $\mathcal{E}(\mathcal{F}_2^1),\ \mathcal{E}(\mathcal{F}_2^2),\ \mathcal{E}(\mathcal{F}_2^3)$ respectively.

A.5. Proofs of inequalities

Inequality I_1 holds for values of α and β satisfying $1 \geq \alpha, \beta \geq 0$ and $\alpha + \beta \leq 1$ since:

$$\alpha + \beta - 4\alpha\beta = \underbrace{(1 - \alpha - \beta)}_{\geq 0}\underbrace{(\alpha + \beta)}_{\geq 0} + (\alpha - \beta)^2 \geq 0.$$

Inequality I_2 holds for values of α and β satisfying $1 \geq \alpha, \beta \geq 0$ and $\alpha + \beta \geq 1$ since:

$$4\alpha + 4\beta - 4\alpha\beta - 3 = \underbrace{(3 - \alpha - \beta)}_{\geq 0}\underbrace{(\alpha + \beta - 1)}_{\geq 0} + (\alpha - \beta)^2 \geq 0.$$

A.6. Completing the proof of Theorem 6

For regions R_2, R_3 and R_4 (moderate probabilities), we further partition of the set of functions as $\mathcal{F}_3 = \mathcal{F}_3^1 \cup \mathcal{F}_3^2 \cup \mathcal{F}_3^3$ where these sets are defined in Lemma 4. We verify $\delta = 4\underline{f}^{++}(\boldsymbol{x}) - 3\overline{f}^{+}(\boldsymbol{x}) \geq 0$ over each region by using the pairwise independent distribution θ_1^{++} to evaluate $\underline{f}^{++}(\boldsymbol{x})$ and appropriate dual feasible solutions of the linear program (2) to evaluate $\overline{f}^{+}(\boldsymbol{x})$. We use the following dual solutions:

For $\mathcal{F}_3^1$:
(D_1) $(\lambda_0, \lambda_1, \lambda_2, \lambda_3) = (f(2) - f(2|3), f(1) - f(2) + f(2|3), f(2|3), f(3|2))$,
(D_2) $(\lambda_0, \lambda_1, \lambda_2, \lambda_3) = (f(1) - f(1|3), f(1|3), f(2|3), f(3|1))$,
For $\mathcal{F}_3^2$:
(D_3) $(\lambda_0, \lambda_1, \lambda_2, \lambda_3) = (f(1) - f(1|3), f(1|3), f(2) - f(1) + f(1|3), f(3|1))$,
(D_4) $(\lambda_0, \lambda_1, \lambda_2, \lambda_3) = (f(2) - f(2|3), f(1|3), f(2|3), f(3|2))$,
For $\mathcal{F}_3^3$:
(D_5) $(\lambda_0, \lambda_1, \lambda_2, \lambda_3) = (f(1) - f(1|2), f(1|2), f(2|1), f(3) - f(1) + f(1|2))$,
(D_6) $(\lambda_0, \lambda_1, \lambda_2, \lambda_3) = (f(1,2) - f(1|3) - f(2|3), f(1|3), f(2|3), f(1|3) + f(2,3) - f(1,2))$,
(D_7) $(\lambda_0, \lambda_2, \lambda_2, \lambda_3) = (f(2) - f(2|3), f(1|2), f(2|3), f(3|2))$.

It is easy to verify that these solutions are dual feasible for the specified F_3^i. For each region, we make use of a combination of the dual feasible solutions shown in Table 5. We now complete the

Table 5: Dual feasible solutions selected for each region.

REGION	DUAL FEASIBLE SOLUTION
R_2	D_1 IN $\mathcal{F}_3^1$, D_3 IN $\mathcal{F}_3^2$, D_5 IN $\mathcal{F}_3^3$
R_3	D_1 IN $\mathcal{F}_3^1$, D_4 IN $\mathcal{F}_3^2$, D_7 IN $\mathcal{F}_3^3$
R_4	D_2 IN $\mathcal{F}_3^1$, D_4 IN $\mathcal{F}_3^2$, D_6 IN $\mathcal{F}_3^3$

proof:

1. Region R_2: For $E_1 = (0,1,0,0,1,1,0,1)$, we need to check for each of the three cases with $f \in \mathcal{F}_3^1$, $f \in \mathcal{F}_3^2$ and $f \in \mathcal{F}_3^3$ by using the corresponding dual feasible solution from Table 5. In each case, we get:

$$\begin{aligned}\delta &= 4(\theta_1^{++}(1) + \theta_1^{++}(1,2) + \theta_1^{++}(1,3) + \theta_1^{++}(1,2,3)) - 3x_1 \\ &= x_1 \\ &\geq 0.\end{aligned}$$

Similar computations for E_2 and E_3 for each of the three cases gives $\delta = x_2$ and $\delta = x_3$ which are nonnegative. For $E_4 = (0,1,1,0,1,1,1,1)$, we need to check only for two cases where $f \in \mathcal{F}_3^1$ and $f \in \mathcal{F}_3^2$ by using the corresponding dual feasible solution from Table 5. In each case, we obtain:

$$\begin{aligned}\delta &= 4(\theta_1^{++}(1) + \theta_1^{++}(2) + \theta_1^{++}(1,2) + \theta_1^{++}(1,3) + \theta_1^{++}(2,3) + \theta_1^{++}(1,2,3)) - 3(x_1 + x_2) \\ &= x_1 + x_2 - 4x_1x_2 \\ &\geq 0 \text{ [using } I_1 \text{ since } x_1 + x_2 \leq 1 \text{ in } R_2].\end{aligned}$$

The key here is that we do not need to verify any inequality at E_4 for functions in $\mathcal{F}_3^3$ since it is not an extreme point for this polytope (see Lemma 4). For E_5, we obtain for both cases $f \in \mathcal{F}_3^1$ and $f \in \mathcal{F}_3^3$, $\delta = x_1 + x_3 - 4x_1x_3$ which is nonnegative since $x_1 + x_3 \leq 1$ in region R_2. Again we do not need to verify the inequality at E_5 for $\mathcal{F}_3^2$. For E_6, for both the cases $f \in \mathcal{F}_3^2$ and $f \in \mathcal{F}_3^3$, we get $\delta = x_2 + x_3 - 4x_2x_3$ which is nonnegative since $x_2 + x_3 \leq 1$ in region R_2. Here do not need to verify the inequality at E_6 for $\mathcal{F}_3^1$. For E_7 and each of the three cases with $f \in \mathcal{F}_3^1$, $f \in \mathcal{F}_3^2$ and $f \in \mathcal{F}_3^3$ by using the appropriate dual feasible solutions we get:

$$\begin{aligned}
\delta &= 4(\theta_1^{++}(1) + \theta_1^{++}(2) + \theta_1^{++}(3) + \theta_1^{++}(1,2) + \theta_1^{++}(1,3) + \theta_1^{++}(2,3) + \theta_1^{++}(1,2,3)) - 3 \\
&= 4(x_1 + x_2 + x_3) - 4x_3(x_1 + x_2) - 3 \\
&\geq 0 \text{ [using } I_2 \text{ since } x_1 + x_2 + x_3 \geq 1 \text{ in } R_2].
\end{aligned}$$

For E_8, in each of the three cases we have:

$$\begin{aligned}
\delta &= 4(\tfrac{1}{2}\theta_1^{++}(1) + \tfrac{1}{2}\theta_1^{++}(2) + \tfrac{1}{2}\theta_1^{++}(3) + \theta_1^{++}(1,2) + \theta_1^{++}(1,3) + \theta_1^{++}(2,3) + \theta_1^{++}(1,2,3)) \\
&\quad -3(\tfrac{1}{2}x_1 + \tfrac{1}{2}x_2 + \tfrac{1}{2}x_3) \\
&= \tfrac{1}{2}(4(x_1 + x_2 + x_3 - x_1x_2) - 3(x_1 + x_2 + x_3)) \\
&= \tfrac{1}{2}(x_1 + x_2 + x_3 - 4x_1x_2) \\
&\geq 0 \text{ [using } I_1 \text{ since } x_1 + x_2 \leq 1 \text{ in } R_2 \text{ and } x_3 \geq 0].
\end{aligned}$$

2. Region R_3: For E_1, E_2 and E_3 in each of the three cases with $f \in \mathcal{F}_3^1$, $f \in \mathcal{F}_3^2$ and $f \in \mathcal{F}_3^3$ by using the dual feasible solutions from Table 5, we get $\delta = x_1$, $\delta = x_2$ and $\delta = x_3$ which are nonnegative. For $E_4 = (0,1,1,0,1,1,1,1)$, we have for $f \in \mathcal{F}_3^1$ and $f \in \mathcal{F}_3^2$ by using the specified dual solutions $\delta = x_1 + x_2 - 4x_1x_2$ which is nonnegative since $x_1 + x_2 \leq 1$ in R_3. Again we do not need to verify the inequality at E_4 for $\mathcal{F}_3^3$. At E_5, for both $f \in \mathcal{F}_3^1$ and $f \in \mathcal{F}_3^3$, we get $\delta = x_1 + x_3 - 4x_1x_3$ which is nonnegative using I_1 since $x_1 + x_3 \leq 1$ in region R_3. At E_6, for both $f \in \mathcal{F}_3^2$ and $f \in \mathcal{F}_3^3$, we get from the dual solutions,:

$$\begin{aligned}
\delta &= 4(\theta_1^{++}(2) + \theta_1^{++}(3) + \theta_1^{++}(1,2) + \theta_1^{++}(1,3) + \theta_1^{++}(2,3) + \theta_1^{++}(1,2,3)) - 3 \\
&= 4x_2 + 4x_3 - 4x_2x_3 - 3 \\
&\geq 0 \text{ [using } I_2 \text{ since } x_2 + x_3 \geq 1 \text{ in } R_3].
\end{aligned}$$

We do not need to verify the inequality at E_6 for $\mathcal{F}_3^1$. For E_7 and E_8, we get $\delta = 4(x_1 + x_2 + x_3) - 4x_3(x_1 + x_2) - 3$ and $\delta = 1/2(x_1 + x_2 + x_3 - 4x_1x_2)$ which are nonnegative in R_3 since $x_1 + x_2 + x_3 \geq 1$ and $x_1 + x_2 \leq 1$ (use I_2 and I_1).

3. Region R_4: For E_1, E_2 and E_3 in each of the three cases with $f \in \mathcal{F}_3^1$, $f \in \mathcal{F}_3^2$ or $f \in \mathcal{F}_3^3$ by using the dual feasible solutions from Table 5, we get $\delta = x_1$, $\delta = x_2$ and $\delta = x_3$ which are nonnegative. For E_4, we have for $f \in \mathcal{F}_3^1$ and $f \in \mathcal{F}_3^2$, $\delta = x_1 + x_2 - 4x_1x_2 \geq 0$ using I_1 since $x_1 + x_2 \leq 1$ in this region. At E_5, for both $f \in \mathcal{F}_3^1$ and $f \in \mathcal{F}_3^3$, we get:

$$\begin{aligned}
\delta &= 4(\theta_1^{++}(1) + \theta_1^{++}(3) + \theta_1^{++}(1,2) + \theta_1^{++}(1,3) + \theta_1^{++}(2,3) + \theta_1^{++}(1,2,3)) - 3 \\
&= 4x_1 + 4x_3 - 4x_1x_3 - 3 \\
&\geq 0 \text{ [using } I_2 \text{ since } x_1 + x_3 \geq 1 \text{ in } R_4].
\end{aligned}$$

At E_6, for both $f \in \mathcal{F}_3^2$ and $f \in \mathcal{F}_3^3$ by using the appropriate dual feasible solutions, we get:

$$\begin{aligned}
\delta &= 4(\theta_1^{++}(2) + \theta_1^{++}(3) + \theta_1^{++}(1,2) + \theta_1^{++}(1,3) + \theta_1^{++}(2,3) + \theta_1^{++}(1,2,3)) - 3 \\
&= 4x_2 + 4x_3 - 4x_2x_3 - 3 \\
&\geq 0 \text{ [using } I_2 \text{ since } x_2 + x_3 \geq 1 \text{ in } R_4].
\end{aligned}$$

For E_7 and E_8, we get $\delta = 4(x_1+x_2+x_3) - 4x_3(x_1+x_2) - 3$ and $\delta = 1/2(x_1+x_2+x_3-4x_1x_2)$ which are nonnegative in R_4 since $x_1 + x_2 + x_3 \geq 1$ and $x_1 + x_2 \leq 1$ (use I_2 and I_1).

A.7. Proof of Lemma 7

Proof For each $i \in [m]$, $f_i^+(\boldsymbol{x}_i) = \max_{\theta_i \in \Theta_i} \mathbb{E}_{\theta_i}[f_i(S)]$ where Θ_i is the set of joint distributions of the random input with given marginal probability vector $\boldsymbol{x}_i$. To compute $f^+(\boldsymbol{x})$, we need to evaluate $\max_{\theta \in \Theta} \mathbb{E}_\theta[\sum_{i=1}^m f_i(S)]$ where Θ is the set of joint distributions of the entire random input with the given marginal probability vector $\boldsymbol{x} = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_m)$. Note that since Θ is specified by the marginals of the partitioned distributions, the maximum values can be computed individually for each f_i and summed up. Hence:

$$f^+(\boldsymbol{x}) = \sum_{i=1}^m \max_{\theta \in \Theta_i} \mathbb{E}_\theta[f_i(S)] = \sum_{i=1}^m f_i^+(\boldsymbol{x}_i).$$

Similarly for the upper pairwise independent extension, we have $f^{++}(\boldsymbol{x}) = \sum_{i=1}^m f_i^{++}(\boldsymbol{x}_i)$. Furthermore, $f_i^+(\boldsymbol{x}_i)/f_i^{++}(\boldsymbol{x}_i) \leq 4/3$ for all $\boldsymbol{x}_i \in [0,1]^{n_i}$ and $n_i \leq 3$ from Theorems 5 and 6. Together this gives:

$$\frac{f^+(\boldsymbol{x})}{f^{++}(\boldsymbol{x})} = \frac{\sum_{i=1}^m f_i^+(\boldsymbol{x}_i)}{\sum_{i=1}^m f_i^{++}(\boldsymbol{x}_i)} \leq \frac{4}{3}.$$

■

A.8. Proof of Theorem 10

Proof With identical probabilities $x_i = x$ for $i \in [n]$, the conditions in Theorem 9 reduce to $x \leq \min(1/4, 1/(n-1))$. With $x \leq 1/(n-1)$, the proof belows shows that it is possible to do away with the $1/4$ bound. When $n = 2, 3, 4$ with $x \leq 1/4$ or $n \geq 5$ with $x \leq 1/(n-1)$, the result follows directly from Theorem 9 since $x \leq \min(1/4, 1/(n-1))$. For $n = 2, 3$, the earlier results in Section 3.2 subsume the identical marginal case and thus we only need to prove the result for $n = 4$ with $1/4 < x \leq 1/3$. We can use the pairwise independent distribution from Table 4 to compute $\underline{f}^{++}(\boldsymbol{x})$ for $n = 4$. The difference here is that due to the identical nature of the marginal probabilities, it is easy to explicitly specify the distribution (see Table 6). This gives the

Table 6: Feasible pairwise independent distribution for $n = 4$ with identical marginals $x \leq 1/3$

S	$\theta(S)$
$\emptyset$	$(1-3x)(1-x)$
$\{1\}$	$x(1-x)$
$\{2\}$	$x(1-x)$
$\{3\}$	$x(1-x)$
$\{4\}$	$x(1-x)$
$\{1,2,3,4\}$	x^2

lower bound $\underline{f}^{++}(\boldsymbol{x}) = x(1-x)\sum_{i=1}^4 f(i) + x^2$. Let $a = \sum_{i=1}^4 f(i)$. Note that by our definition of a monotone submodular functions, we have $4 \geq a \geq 1$. Hence $1/4 \leq 1/a \leq 1$ and we can partition the identical probability space $(1/4, 1/3]$ into two regions $R_1 = \{x : 1/4 < x \leq 1/a\}$ and $R_2 = \{x : 1/a < x \leq 1/3\}$. If $1/a > 1/3$, R_2 is empty while R_1 subsumes this condition and it is thus sufficient to focus on the case when $1/4 \leq a < 1/3$. Next, to compute $\overline{f}^+(\boldsymbol{x})$, we consider the dual feasible solution $\lambda_0 = 1, \lambda_i = 0$ for $i = 1, 2, 3, 4$ with $\overline{f}^+(\boldsymbol{x}) = 1$. Along with the dual solution $\lambda_0 = 0, \lambda_i = 1$ for $i = 1, 2, 3, 4$, we get an upper bound $\overline{f}^+(\boldsymbol{x}) = \min(ax, 1)$, so that exactly one of the two dual solutions will be chosen in each region R_1, R_2. Note that both the dual solutions are feasible for all $f \in \mathcal{F}_4$ and hence it is sufficient to verify the inequality $4\underline{f}^{++}(\boldsymbol{x}) - 3\overline{f}^+(\boldsymbol{x}) = 4ax(1-x) + 4x^2 - 3\min(ax, 1)$ for all $x \in R_1 \cup R_2$. For $x \in R_1$, $4\underline{f}^{++}(\boldsymbol{x}) - 3\overline{f}^+(\boldsymbol{x}) = 4ax(1-x) + 4x^2 - 3ax = ax(1-4x) + 4x^2 \geq 1 - 4x + 4x^2 = (1-2x)^2 \geq 0$ where we used the fact that $1 - 4x \leq 0$ and $ax \leq 1$. For $x \in R_2$, $ax > 1$ and $4\underline{f}^{++}(\boldsymbol{x}) - 3\overline{f}^+(\boldsymbol{x}) = 4ax(1-x) + 4x^2 - 3\min(ax, 1) = 4ax(1-x) + 4x^2 - 3 \geq 4(1-x) + 4x^2 - 3 = (1-2x)^2 \geq 0$ and the result follows. ■