CHAPTER 1

A Probabilistic Averaging Technique

This chapter presents a powerful probabilistic proof technique which will be used, in combination with additional ideas, throughout this book. For a simple first application, consider the following problem.

Let $P = \{p_1, \dots, p_n\}$ be a set of n points in the plane and let $k \geq 0$ be an integer. The k-Delaunay graph of P, denoted by $G_k(P) = (P, E_k)$, is defined as follows:

> $\{p_i, p_j\} \in E_k$ if and only if there exists a closed disk in the plane containing $\{p_i, p_j\}$ and at most k points of $P \setminus \{p_i, p_j\}$.
>
> For each $\{p_i, p_j\} \in E_k$ fix any one such disk and denote it by D_{ij}.

Note that $E_0 \subseteq E_1 \subseteq \cdots \subseteq E_{n-2} = \binom{P}{2}$.

Our goal is to upper bound the number of edges in the k-Delaunay graph of P as a function of n and k. We will prove that $|E_k| = O(n(k+1))$[1].

First, observe that the 0-Delaunay graph of P is simply the Delaunay graph, which is planar and thus E_0 has size at most $3n$. Next, we upper bound $|E_k|$, for any integer $k \geq 1$, by the following argument.

> Let S be a random sample constructed by picking each point of P independently with probability $p = \frac{1}{k+1}$ and let $G_0(S)$ be the 0-Delaunay graph of S. We count the expected number of edges in $G_0(S)$ in two ways.
>
> **Upper bound:** As *any* 0-Delaunay graph on t vertices has at most $3t$ edges, the expected number of edges in $G_0(S)$ is
>
> $$\mathrm{E}\left[3|S|\right] = 3\,\mathrm{E}\left[|S|\right] = 3np = \frac{3n}{k+1}.$$
>
> **Lower bound:** For any $\{p_i, p_j\} \in E_k$, if both p_i and p_j are picked in S and none of the at most k other points of P lying in D_{ij} are picked in S, then $\{p_i, p_j\}$ is an edge in $G_0(S)$. As each point of P was picked independently, the probability that $\{p_i, p_j\}$ is an edge in $G_0(S)$ is at least
>
> $$p^2 \cdot (1-p)^k = \frac{1}{(k+1)^2} \cdot \left(1 - \frac{1}{k+1}\right)^k \geq \frac{1}{(k+1)^2} \cdot \frac{1}{e}, \tag{1.1}$$
>
> where the last step uses the fact that $\left(1 + \frac{1}{k}\right)^k \leq e$, and thus $\frac{1}{e} \leq \left(\frac{k}{k+1}\right)^k = \left(1 - \frac{1}{k+1}\right)^k$.

[1] The '+1' term is there just to take care of the case $k = 0$.

Using linearity of expectation, Equation (1.1) implies that the expected number of edges of E_k that are present in $G_0(S)$ is at least $|E_k| \cdot \frac{1}{(k+1)^2 e}$.

Combining the upper and lower bounds, we get

$$|E_k| \cdot \frac{1}{(k+1)^2 e} \;\le\; \text{expected number of edges in } G_0(S) \;=\; \frac{3n}{k+1},$$

implying that $|E_k| = O(n(k+1))$.

Before we move on to other applications of this technique, we make a few remarks.

- The use of random sampling in the above proof is a way to 'implement' a double-counting argument. Essentially we are summing up, over all edges e in $G_k(P)$, the number of subsets $S \subseteq P$ of size $\frac{n}{k+1}$ for which e is an edge in $G_0(S)$[2]. We counted this sum in two ways: iterating over edges of $G_k(P)$ gave a lower bound while iterating over subsets of P gave an upper bound. More precisely, define the set of pairs

  $$\mathcal{I} = \big\{\, (e, S) : e \in E_k,\ |S| = \lceil n/(k+1) \rceil,\ e \text{ is in } G_0(S) \big\}.$$

  Then the above double-counting argument gives

  $$|E_k| \cdot \binom{n-2-k}{\lceil n/(k+1) \rceil - 2} \;\le\; |\mathcal{I}| \;\le\; \binom{n}{\lceil n/(k+1) \rceil} \cdot 3\lceil n/(k+1) \rceil.$$

  Solving this for $|E_k|$ gives $|E_k| = O(n(k+1))$, as before.
- The utility of framing the argument probabilistically is that it beautifully captures the intuition behind the key idea: if there are 'too many' edges in $G_k(P)$, then in expectation more than $3|S|$ of these edges will 'filter through' to $G_0(S)$, for a random sample S. This contradicts the fact that for *any* S, $G_0(S)$ has at most $3|S|$ edges.
- The lower bound follows by considering, for each edge $\{p_i, p_j\}$ of $G_k(P)$, a specific event whose occurrence implies that $\{p_i, p_j\}$ appears as an edge in $G_0(S)$. This need not be the *only* event that could cause $\{p_i, p_j\}$ to be an edge in $G_0(S)$—e.g., there could be a disk other than D_{ij} containing p_i and p_j that happens to not contain any other point of S. Thus our lower bound is not necessarily tight.

  In fact, what we actually want to compute is a lower bound on the probability that there exists *some* disk containing $\{p_i, p_j\}$ and no other point of S. However the events for all possible disks containing $\{p_i, p_j\}$ are not independent, which makes computing this probability difficult. Fortunately, we do not lose much by considering *any one* such disk, and in fact the lower bound is optimal up to constant factors for certain point sets. In particular, the bound $|E_k| = O(n(k+1))$ is tight for the instance of n points lying on a line.
- The calculation, when carried out with probability $p \in (0,1)$ as a parameter, gives $|E_k| = O\left(\frac{n}{p(1-p)^k}\right)$. The value of p is then set to maximize the denominator. Roughly speaking, as the term $(1-p)^k = e^{-\Theta(pk)}$ decreases exponentially with p, it is best to set p so that $e^{-\Theta(pk)}$ is a constant—that is, $p = \Theta\left(\frac{1}{k}\right)$.

[2] A minor technical difference is that in the probabilistic version, the *expected size* is $\frac{n}{k+1}$.

As for the precise value of p that maximizes the denominator, since the derivative of $p\,(1-p)^k$ with respect to p is $(1-kp-p)\,(1-p)^{k-1}$, it can be verified that $p\,(1-p)^k$ is maximized at $p = \frac{1}{k+1}$.

This also makes sense intuitively: for each edge $\{p_i, p_j\} \in E_k$, the disk D_{ij} contains at most k other points of P and so picking each point with probability less than $\frac{1}{k}$ implies that, in expectation, D_{ij} will not contain any of these points.

- Other applications of this technique follow the same 'template'—pick a random sample and calculate the probability of some event due to it in two ways. The main technical work consists in finding good estimates for certain events; this is typically where a variety of other combinatorial and geometric ideas come into play.

1. Level Sets

> *Counting pairs is the oldest trick in combinatorics ... every time we count pairs, we learn something from it.*
>
> Gil Kalai

Our first application is a variant of the k-Delaunay graph problem. Given a finite set P of points in $\mathbb{R}^d$ and an integer $k \geq 1$, the objective is to upper bound the number of subsets of P of size at most k that are 'realizable' by geometric objects in $\mathbb{R}^d$. We first explain the problem for the case of disks in $\mathbb{R}^2$.

For a set P of n points in $\mathbb{R}^2$, define the set system

$$\mathcal{R}(P) = \big\{ D \cap P \,:\, D \text{ is a disk in } \mathbb{R}^2 \big\}.$$

We call $\mathcal{R}(P)$ the primal set system induced on P by disks. For any integer $k \geq 1$, let $\mathcal{R}_{=k}(P)$ be the sets of $\mathcal{R}(P)$ of size exactly k and let $\mathcal{R}_{\leq k}(P)$ be the sets of $\mathcal{R}(P)$ of size at most k. That is,

$$\mathcal{R}_{=k}(P) = \big\{ R \in \mathcal{R}(P) : |R| = k \big\} \quad \text{and} \quad \mathcal{R}_{\leq k}(P) = \big\{ R \in \mathcal{R}(P) : |R| \leq k \big\}.$$

The sets of $\mathcal{R}_{\leq k}(P)$ are called the ($\leq k$)-level sets, or simply ($\leq k$)-sets, of $\mathcal{R}(P)$.

> Observe that $\mathcal{R}_{\leq 2}(P)$—the subsets of P of size at most two that are induced by disks—consists of $O(n)$ sets: the sets of size 1 in $\mathcal{R}_{\leq 2}(P)$ are the points of P and the sets of size 2 are precisely the edges of the Delaunay graph of P. At the other end, $\mathcal{R}_{\leq n}(P)$ is just $\mathcal{R}(P)$, with size $O\left(n^3\right)$.

Our first main result of this section implies both of the above two cases.

Lemma 1.2. *Let P be a set of n points in $\mathbb{R}^2$ and let $\mathcal{R}(P)$ be the primal set system induced on P by disks in the plane. Then for any integer $k \geq 1$,*

$$|\mathcal{R}_{\leq k}(P)| = O\left(nk^2\right).$$

To simplify the presentation, we will assume that $|P| \geq 3$, and that P is in general position; in particular, no three points lie on a line and no four points lie on a circle.

To prove Lemma 1.2, we will first count a slightly different structure called *canonical disks*, which are disks that are 'fixed' by points of P on their boundary.

Definition 1.3. A canonical disk spanned by $Q \subseteq \mathbb{R}^2$ is a disk whose boundary contains three points of Q.

Furthermore, a canonical disk D spanned by Q is called an empty canonical disk if the interior of D contains no point of Q.

Let $\mathcal{T}(P)$ be the set of all $\binom{n}{3}$ unordered triples of points of P. For a triple $\{p, q, r\} \in \mathcal{T}(P)$, let D_{pqr} be the unique *open* disk whose boundary contains $\{p, q, r\}$; we say that D_{pqr} is spanned by $\{p, q, r\}$. For an integer $k \geq 0$, define the level sets

$$\mathcal{T}_{\leq k}(P) = \Big\{ \{p, q, r\} \in \mathcal{T}(P) : |D_{pqr} \cap P| \leq k \Big\}.$$

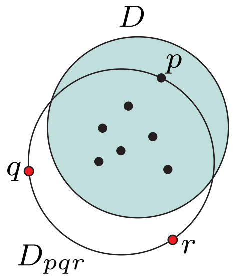
We first observe that the size of $\mathcal{R}_{\leq k}(P)$ is bounded, within a constant factor, by that of $\mathcal{T}_{\leq k}(P)$.

Claim 1.4. For any integer $k \geq 1$, $|\mathcal{R}_{\leq k}(P)| \leq 8 \cdot |\mathcal{T}_{\leq (k-1)}(P)|$.

Proof. Take any $R \in \mathcal{R}_{\leq k}(P)$ and let D be a disk realizing R; that is, $R = D \cap P$.

Now D can be scaled and translated—without any point of P 'crossing' the boundary of D—such that it contains three points of P, say $\{p, q, r\}$, on its boundary. Furthermore at least one of p, q or r belongs to R. See figure.

The interior of D_{pqr} contains at most $k-1$ points of P and so $\{p, q, r\} \in \mathcal{T}_{\leq (k-1)}(P)$. By slightly shifting and scaling D_{pqr}, for each of the 8 possible subsets of $\{p, q, r\}$, one can get a disk containing precisely that subset and all the points of P in the interior of D_{pqr}. One of these subsets is R, implying the claim.



□

We remark here that the constant 8 can be improved with a more careful argument (see discussion).

Now the proof of Lemma 1.2 follows from Claim 1.4 and the following statement.

Lemma 1.5. *Let P be a set of n points in $\mathbb{R}^2$ and let $k \geq 0$ be an integer. Then*

$$|\mathcal{T}_{\leq k}(P)| = O\left(n(k+1)^2\right).$$

Proof. First we establish the case $k = 0$.

Claim 1.6. For any $S \subseteq P$, $|\mathcal{T}_{\leq 0}(S)| \leq 2|S|$.

Proof. $\mathcal{T}_{\leq 0}(S)$ consists of unordered triples of S whose corresponding open disks do not contain any point of S in their interior. If the disk D_{pqr}, spanned by $p, q, r \in S$, contains no point of S in its interior, then by slightly shifting D_{pqr}, it follows that each of the three edges $\{p, q\}$, $\{q, r\}$ and $\{p, r\}$ belong to the Delaunay graph of S. In particular, the triangle with vertices $\{p, q, r\}$ is a face of the Delaunay graph of S. Thus $|\mathcal{T}_{\leq 0}(S)|$ is upper bounded by the number of faces in a planar graph on $|S|$ vertices, which is $2|S| - 4$. □

Now consider the case $k \geq 1$. Construct a random sample S by picking each point of P independently with probability $p = \frac{1}{k+1}$.

We count the expected size of $\mathcal{T}_{\leq 0}(S)$ in two ways.

Upper bound: From Claim 1.6,

$$\mathrm{E}\left[\,\left|\mathcal{T}_{\leq 0}(S)\right|\,\right] \leq \mathrm{E}\left[2\,|S|\right] = 2\,np.$$

Lower bound: The key is the following observation:

> a triple $\{p,q,r\} \in \mathcal{T}(P)$ is present in $\mathcal{T}_{\leq 0}(S)$ *if and only if* $\{p,q,r\} \subseteq S$ and none of the points in $D_{pqr} \cap P$ are picked in S.

As each point of P was picked independently, for any $\{p,q,r\} \in \mathcal{T}(P)$, we have

$$\Pr\left[\,\{p,q,r\} \in \mathcal{T}_{\leq 0}(S)\,\right] = p^3 \cdot (1-p)^{|D_{pqr} \cap P|}.$$

Therefore, by linearity of expectation,

$$\begin{aligned}
\mathrm{E}\left[\,\left|\mathcal{T}_{\leq 0}(S)\right|\,\right] &= \sum_{\{p,q,r\} \in \mathcal{T}(P)} \Pr\left[\,\{p,q,r\} \in \mathcal{T}_{\leq 0}(S)\,\right] \\
&\geq \sum_{\{p,q,r\} \in \mathcal{T}_{\leq k}(P)} \Pr\left[\,\{p,q,r\} \in \mathcal{T}_{\leq 0}(S)\,\right] \\
&= \sum_{\{p,q,r\} \in \mathcal{T}_{\leq k}(P)} p^3 \cdot (1-p)^{|D_{pqr} \cap P|} \\
&\geq \sum_{\{p,q,r\} \in \mathcal{T}_{\leq k}(P)} p^3 \cdot (1-p)^k = \left|\mathcal{T}_{\leq k}(P)\right| \cdot p^3 \cdot (1-p)^k.
\end{aligned}$$

Combining the upper and lower bounds,

$$\left|\mathcal{T}_{\leq k}(P)\right| \cdot p^3 \cdot (1-p)^k \;\leq\; \mathrm{E}\left[\,\left|\mathcal{T}_{\leq 0}(S)\right|\,\right] \;\leq\; 2\,np,$$

$$\text{and hence}\quad \left|\mathcal{T}_{\leq k}(P)\right| \leq \frac{2\,n}{p^2 \cdot (1-p)^k} = \frac{2\,n\,(k+1)^2}{\left(1 - \frac{1}{k+1}\right)^k} \leq 2e\,n\,(k+1)^2,$$

where the last step follows from the fact that $\left(1 - \frac{1}{k+1}\right)^k \geq \frac{1}{e}$. □

We next prove a similar statement for set systems where the elements are geometric objects in $\mathbb{R}^d$ and the sets are induced by points in $\mathbb{R}^d$. We consider the case of disks in the plane.

Given a set $\mathcal{D} = \{D_1, \ldots, D_n\}$ of n distinct closed disks in $\mathbb{R}^2$, define the set system

$$\mathcal{R}(\mathcal{D}) = \left\{\mathcal{D}_p \;:\; p \in \mathbb{R}^2\right\}, \quad \text{where} \quad \mathcal{D}_p = \left\{D \in \mathcal{D} \;:\; D \ni p\right\}.$$

We call $\mathcal{R}(\mathcal{D})$ the dual set system induced on $\mathcal{D}$ by $\mathbb{R}^2$. Visually, each cell in the arrangement of $\mathcal{D}$ corresponds to a set in $\mathcal{R}(\mathcal{D})$ (note that different cells may correspond to the same subset).

For simplicity we will assume that $\mathcal{D}$ is in general position—in particular, the intersection of the boundaries of every pair of disks of $\mathcal{D}$ is either empty or consists of two distinct points and the intersection of the boundaries of any three disks of $\mathcal{D}$ is empty.

Our goal is to upper bound, for any integer $k \geq 1$, the size of $\mathcal{R}_{\leq k}(\mathcal{D})$. Our main result is the following.

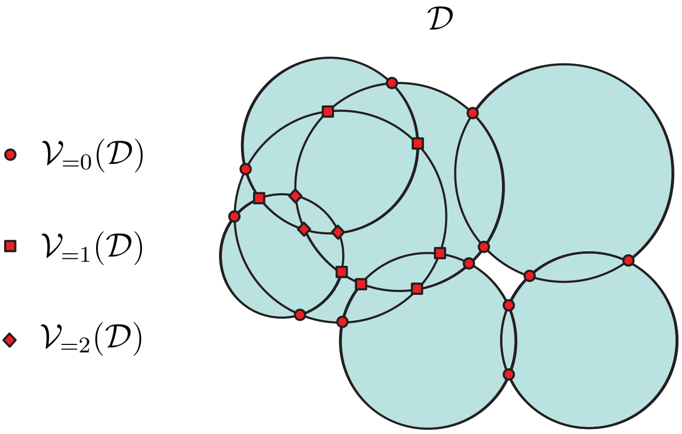
LEMMA 1.7. *Let $\mathcal{D}$ be a set of n closed disks in $\mathbb{R}^2$ and let $\mathcal{R}(\mathcal{D})$ be the dual set system induced on $\mathcal{D}$. Then for any integer $k \geq 1$,*

$$|\mathcal{R}_{\leq k}(\mathcal{D})| = O(nk).$$

As earlier, it suffices to consider canonical sets, defined as follows. Let $\mathcal{V}(\mathcal{D})$ be the set of at most $2\binom{n}{2}$ points in $\mathbb{R}^2$ that are the intersections of boundaries of the disks of $\mathcal{D}$. For any integer $k \geq 0$, define

$$\mathcal{V}_{\leq k}(\mathcal{D}) = \Big\{v \in \mathcal{V}(\mathcal{D}) : v \text{ is contained in the } \textit{interior} \text{ of at most } k \text{ disks of } \mathcal{D}\Big\}.$$

Similarly one can define $\mathcal{V}_{=k}(\mathcal{D})$. See figure.


$\mathcal{D}$
● $\mathcal{V}_{=0}(\mathcal{D})$
■ $\mathcal{V}_{=1}(\mathcal{D})$
◆ $\mathcal{V}_{=2}(\mathcal{D})$

The proof of the following claim is easy and left to the reader.

CLAIM 1.8. For any integer $k \geq 1$, $|\mathcal{R}_{\leq k}(\mathcal{D})| \leq 4 \cdot |\mathcal{V}_{\leq (k-1)}(\mathcal{D})| + |\mathcal{D}|$.

Now the proof of Lemma 1.7 follows from Claim 1.8 and the following statement.

LEMMA 1.9. *For any integer $k \geq 0$, $|\mathcal{V}_{\leq k}(\mathcal{D})| = O(n(k+1))$.*

PROOF. As before, we first upper bound the size of $\mathcal{V}_{\leq 0}(\mathcal{D})$ and then use the averaging technique to upper bound the size of $\mathcal{V}_{\leq k}(\mathcal{D})$ for $k \geq 1$.

CLAIM 1.10. For any $S \subseteq \mathcal{D}$, $|\mathcal{V}_{\leq 0}(S)| \leq 6\,|S|$.

PROOF. Any $v \in \mathcal{V}_{\leq 0}(S)$ is an intersection point between the boundary of two disks of S and is not contained in the interior of any disk of S. Let $G = (S, E)$ be a graph where there is an edge between two disks of S if and only if a common intersection point of their boundaries belongs to $\mathcal{V}_{\leq 0}(S)$. We now show that G is planar, and so $|\mathcal{V}_{\leq 0}(S)| \leq 2|E| \leq 2(3|S| - 6) \leq 6|S|$.
We claim that the following is a plane drawing of G: draw each edge $\{D_i, D_j\} \in E$ as a line segment between the centers of D_i and D_j. Consider any two edges $\{D_i, D_j\}$, $\{D_k, D_l\}$ and let q_{ij}, q_{kl} be the two corresponding points in $\mathcal{V}_{\leq 0}(S)$. Let l be the bisector of q_{ij} and q_{kl}. As both D_i, D_j contain q_{ij} and do not contain q_{kl}, their centers lie on the side of l containing q_{ij}. Similarly the centers of D_k and D_l lie on the side of l containing q_{kl}. Thus the line segments corresponding to the edges $\{D_i, D_j\}$ and $\{D_k, D_l\}$ cannot intersect. □

Now consider the case $k \geq 1$. Construct a random sample S by picking each disk of $\mathcal{D}$ independently with probability $p = \frac{1}{k+1}$.
We will count the expected size of $\mathcal{V}_{\leq 0}(S)$ in two ways.

Upper bound: From Claim 1.10,

$$\mathrm{E}\left[\left|\mathcal{V}_{\leq 0}(S)\right|\right] \leq \mathrm{E}\left[6\,|S|\right] = 6\,\mathrm{E}\left[|S|\right] = 6\,np.$$

Lower bound: This follows by considering the probability of each vertex in $\mathcal{V}_{\leq k}(\mathcal{D})$ ending up as a vertex of $\mathcal{V}_{\leq 0}(S)$. Let $v \in \mathcal{V}_{\leq k}(\mathcal{D})$ and let $D_i, D_j \in \mathcal{D}$ be the two disks such that v is an intersection point of the boundaries of D_i and D_j. Then

$v \in \mathcal{V}_{\leq 0}(S)$ if and only if $\{D_i, D_j\} \subseteq S$ and every disk of $\mathcal{D}$ containing v in its interior is not present in S.

As there are at most k such disks and each disk of $\mathcal{D}$ was picked independently,

$$\Pr\left[v \in \mathcal{V}_{\leq 0}(S)\right] \geq p^2\,(1-p)^k.$$

Therefore,

$$\begin{aligned}\mathrm{E}\left[\left|\mathcal{V}_{\leq 0}(S)\right|\right] &= \sum_{v \in \mathcal{V}(\mathcal{D})} \Pr\left[v \in \mathcal{V}_{\leq 0}(S)\right] \\ &\geq \sum_{v \in \mathcal{V}_{\leq k}(\mathcal{D})} \Pr\left[v \in \mathcal{V}_{\leq 0}(S)\right] \geq \left|\mathcal{V}_{\leq k}(\mathcal{D})\right| \cdot p^2 (1-p)^k.\end{aligned}$$

Combining the upper and lower bounds,

$$\begin{aligned}&\left|\mathcal{V}_{\leq k}(\mathcal{D})\right| \cdot p^2 (1-p)^k \;\leq\; \mathrm{E}\left[\left|\mathcal{V}_{\leq 0}(S)\right|\right] \;\leq\; 6\,np, \\ \text{and hence}\quad &\left|\mathcal{V}_{\leq k}(\mathcal{D})\right| \leq \frac{6\,n}{p\,(1-p)^k} = \frac{6\,n\,(k+1)}{\left(1-\frac{1}{k+1}\right)^k} \leq 6e\,n\,(k+1),\end{aligned}$$

where the last step used the fact that $\left(1-\frac{1}{k+1}\right)^k \geq \frac{1}{e}$. □

Primal and dual set systems can be defined more generally:

Definition 1.11. Given a set P of points in $\mathbb{R}^d$ and a (possibly infinite) family $\mathcal{R}$ of geometric objects in $\mathbb{R}^d$, the primal set system induced on P by $\mathcal{R}$ is

$$\left\{O \cap P \,:\, O \in \mathcal{R}\right\}.$$

Definition 1.12. Given a set $\mathcal{R}$ of geometric objects in $\mathbb{R}^d$, the dual set system induced on $\mathcal{R}$ by $\mathbb{R}^d$ is defined as

$$\left\{\mathcal{R}_p \colon p \in \mathbb{R}^d\right\}, \quad \text{where} \quad \mathcal{R}_p = \left\{R \in \mathcal{R} \colon R \ni p\right\}.$$

We now conclude with the case of primal and dual set systems induced by half-spaces in $\mathbb{R}^d$.

Let P be a set of n points in general position in $\mathbb{R}^d$ and $\mathcal{R}(P)$ the primal set system induced on P by downward-facing half-spaces—that is, considering the x_d-axis as vertical, the half-spaces which contain the point that is the 'minus infinity' of the x_d

axis. It can be shown that $|\mathcal{R}(P)| = O(n^d)$. In fact for points in general position there is a precise bound independent of the structure of P (stated without proof):

$$|\mathcal{R}(P)| = \sum_{i=0}^{d} \binom{n}{i}. \tag{1.13}$$

Now let $\mathcal{T}(P)$ be all the $\binom{n}{d}$ subsets of P of size d. For each $e \in \mathcal{T}(P)$, let h_e^+ be the unique downward-facing *open* half-space whose bounding hyperplane contains e. For an integer $k \geq 0$, define the level sets

$$\mathcal{T}_{\leq k}(P) = \left\{ e \in \mathcal{T}(P) : \left| h_e^+ \cap P \right| \leq k \right\}.$$

As earlier, the size of $\mathcal{R}_{\leq k}(P)$ can be upper bounded, within a multiplicative factor, by that of $\mathcal{T}_{\leq k}(P)$.

To bound $|\mathcal{T}_{\leq k}(P)|$ we again first need a bound on $|\mathcal{T}_{\leq 0}(P)|$. Observe that $|\mathcal{T}_{\leq 0}(P)|$ is simply the number of facets on the lower convex-hull of P. It is well-known, to those who know it, that the Upper Bound Theorem for convex polytopes implies that this is at most $2\sum_{i=0}^{\lfloor \frac{d}{2} \rfloor} \binom{n}{i}$ (see discussion). Now the probabilistic averaging technique of this chapter together with this 0-th level bound implies the following (stated without proof).

THEOREM 1.14. *Given a set P of n points in $\mathbb{R}^d$ and an integer $k \geq 0$,*

$$|\mathcal{T}_{\leq k}(P)| \leq 2\left(\frac{e}{\lceil d/2 \rceil}\right)^{\lceil \frac{d}{2} \rceil} \binom{n}{\lfloor \frac{d}{2} \rfloor} \left(k + \left\lceil \frac{d}{2} \right\rceil\right)^{\lceil \frac{d}{2} \rceil}.$$

The above is $O\left(n^{\lfloor d/2 \rfloor}(k+1)^{\lceil d/2 \rceil}\right)$ when the dimension d is considered a constant.

> For $d = 3$, Theorem 1.14 gives a bound of $O(nk^2)$—the same bound, within a multiplicative constant, as the one of Lemma 1.5. This is not a coincidence: there exists a mapping of points in $\mathbb{R}^2$ to $\mathbb{R}^3$, the so-called 'paraboloid lift', with the property that subsets realized by intersection with disks in $\mathbb{R}^2$ can be realized by intersection with half-spaces in $\mathbb{R}^3$. Thus Theorem 1.14 for $d = 3$ implies Lemma 1.5.

For later use, it will be convenient to state Theorem 1.14 in the dual setting.

DEFINITION 1.15. The level of a point $q \in \mathbb{R}^d$ with respect to a set $\mathcal{H}$ of hyperplanes in $\mathbb{R}^d$ is the number of hyperplanes of $\mathcal{H}$ lying strictly below q in the negative x_d direction; that is, the number of hyperplanes intersecting the ray

$$\left\{ q + \lambda\,(0, \ldots, 0, -1) : \lambda > 0 \right\}.$$

Given a set $\mathcal{H}$ of hyperplanes in $\mathbb{R}^d$ in general position, *a vertex* in the arrangement of $\mathcal{H}$ is a point lying in the intersection of some d hyperplanes of $\mathcal{H}$. Let $\mathcal{V}_{\leq k}(\mathcal{H})$ be the set of vertices of $\mathcal{H}$ of level at most k. Then by duality, Theorem 1.14 is equivalent to the following statement.

THEOREM 1.16. *Given a set $\mathcal{H}$ of n hyperplanes in $\mathbb{R}^d$ and an integer $k \geq 0$,*

$$|\mathcal{V}_{\leq k}(\mathcal{H})| \leq 2\left(\frac{e}{\lceil d/2 \rceil}\right)^{\lceil \frac{d}{2} \rceil} \binom{n}{\lfloor \frac{d}{2} \rfloor} \left(k + \left\lceil \frac{d}{2} \right\rceil\right)^{\lceil \frac{d}{2} \rceil}.$$

Bibliography and discussion. The influential probabilistic technique in this chapter was used in the seminal paper of Clarkson [**Cla87**], and then Clarkson and Shor [**CS89**], exactly for these problems of bounding the combinatorial complexity of configurations. Indeed, it is sometimes called the 'Clarkson-Shor technique' in the discrete and computational geometry literature. See [**APS08**, **Wag08**] for surveys on ($\leq k$)-sets for half-spaces and related set systems, where one can find information related to Theorem 1.14. Details on the Upper Bound Theorem for convex polytopes [**McM701a**] and related topics can be found in [**Zie95**]. See [**Mat99**, Section 3.1] for a discussion around Claim 1.4.

[APS08] P. K. Agarwal, J. Pach, and M. Sharir, *State of the union (of geometric objects)*, Surveys on discrete and computational geometry, Contemp. Math., vol. 453, Amer. Math. Soc., Providence, RI, 2008, pp. 9–48, DOI 10.1090/conm/453/08794. MR2405676

[CS89] K. L. Clarkson and P. W. Shor, *Applications of random sampling in computational geometry. II*, Discrete Comput. Geom. **4** (1989), no. 5, 387–421, DOI 10.1007/BF02187740. MR1014736

[Cla87] K. L. Clarkson, *New applications of random sampling in computational geometry*, Discrete Comput. Geom. **2** (1987), no. 2, 195–222, DOI 10.1007/BF02187879. MR884226

[Mat99] J. Matoušek. *Geometric Discrepancy: An Illustrated Guide.* Springer, 1999.

[McM701a] P. McMullen, *The maximum numbers of faces of a convex polytope*, Mathematika **17** (1970), 179–184, DOI 10.1112/S0025579300002850. MR283691

[Wag08] U. Wagner, *k-sets and k-facets*, Surveys on discrete and computational geometry, Contemp. Math., vol. 453, Amer. Math. Soc., Providence, RI, 2008, pp. 443–513, DOI 10.1090/conm/453/08810. MR2405692

[Zie95] G. M. Ziegler, *Lectures on polytopes*, Graduate Texts in Mathematics, vol. 152, Springer-Verlag, New York, 1995, DOI 10.1007/978-1-4613-8431-1. MR1311028

2. Concentration Bounds for Sums of Bernoulli Variables

> *I thought it was a rather trivial lemma, but many things are only trivial once you know them.*
>
> Herman Chernoff

We present an application of the probabilistic technique to computing tail bounds of some common probability distributions. That is, we would like to upper bound the probability that a random variable gets a value far from its expectation. This is a basic technical ingredient in nearly all the constructions and methods that will be seen later.

The setting is the following.

> Let $I = \{1, 2, \ldots, n\}$ be a set of n elements from which we will pick a random sample. We aim to pick np elements of I, for a given parameter $p \in [0, 1]$.

The 0-1 valued random variable X_i will be used to indicate whether $i \in I$ is picked in our random sample. Our goal is to estimate the probability that the sum of these n variables, $X = \sum_{i=1}^{n} X_i$, falls far from its expectation $\mathrm{E}[X]$. More precisely, for any $\delta \geq 0$, we are interested in bounding $\Pr\big[X \geq (1+\delta)\,\mathrm{E}[X]\big]$ and $\Pr\big[X \leq (1-\delta)\,\mathrm{E}[X]\big]$.

In fact, we consider the more general case where for a fixed set $J \subseteq I$ with $X_J = \sum_{j \in J} X_j$, we are interested in upper bounds on

$$\Pr\big[X_J \geq (1+\delta) \cdot \mathrm{E}[X_J]\big] \qquad \text{and} \qquad \Pr\big[X_J \leq (1-\delta) \cdot \mathrm{E}[X_J]\big].$$

There are several natural ways to pick a random sample from I. Two basic ones, given a parameter p, are the following.

Binomial distribution: Pick each element of I *independently* with probability p. That is, let $X_1, \ldots, X_n$ be n independent 0-1 random variables where

$$X_i = \begin{cases} 1 & \text{with probability } p, \\ 0 & \text{otherwise.} \end{cases}$$

For any $J \subseteq I$, we have

$$\mathrm{E}[X_J] = \mathrm{E}\left[\sum_{j \in J} X_j\right] = \sum_{j \in J} \mathrm{E}[X_j] = \sum_{j \in J} \Pr[X_j = 1] = |J|\,p.$$

One can write the exact equation for the tail bounds using the fact that the value of each X_i was set independently:

$$\Pr\big[X_J \geq (1+\delta) \cdot |J|\,p\big] = \sum_{i=\lceil (1+\delta)\cdot|J|p\rceil}^{|J|} \Pr[X_J = i] \tag{1.17}$$

$$= \sum_{i=\lceil (1+\delta)\cdot|J|p\rceil}^{|J|} \binom{|J|}{i} p^i (1-p)^{|J|-i}.$$

As there is no closed-form formula for this, several methods have been proposed to estimate the right-hand side of the above expression (see discussion).

The fact that $\{X_1, \ldots, X_n\}$ are independent has two advantages: first it makes calculations easier and second, for any $J \subseteq I$ the induced probability distribution on X_J remains the same (that is, each element of J is picked independently with probability p). On the other hand, the number of elements X is not fixed and is a random variable with expectation np.

Sampling without replacement: A second natural way to sample is to choose, out of all $\binom{n}{np}$ np-sized subsets of I, one uniformly at random (assume that np is an integer). This then sets the values of $X_1, \ldots, X_n$, with $\sum_i X_i$ being equal to np. Note that for any $J \subseteq I$, $\mathrm{E}[X_J] = |J|\, p$ since for any i,

$$\Pr[X_i = 1] = \frac{\binom{n-1}{np-1}}{\binom{n}{np}} = \frac{(n-1)!}{(np-1)!(n-np)!} \cdot \frac{(np)!(n-np)!}{n!} = \frac{np}{n} = p.$$

More generally, for any $J \subseteq I$, letting $t = |J|$, the probability that $X_j = 1$ for all $j \in J$, can be upper bounded as

$$\begin{aligned}(1.18)\qquad \Pr\left[\left(\prod_{j\in J} X_j\right) = 1\right] &= \frac{\binom{n-t}{np-t}}{\binom{n}{np}} = \frac{(n-t)!}{(np-t)!} \cdot \frac{(np)!}{n!} \\ &= \frac{np}{n}\,\frac{np-1}{n-1} \cdots \frac{np-t+1}{n-t+1} \le p^t,\end{aligned}$$

since each term $\frac{np-i}{n-i} \le p$ for $p \le 1$. Similarly, the probability that $X_j = 0$ for all $j \in J$, can be upper bounded as

(1.19)

$$\begin{aligned}\Pr\left[\left(\prod_{j\in J} (1 - X_j)\right) = 1\right] &= \frac{\binom{n-t}{np}}{\binom{n}{np}} = \frac{(n-t)!}{(n-t-np)!} \cdot \frac{(n-np)!}{n!} \\ &= \frac{n-np}{n}\,\frac{n-np-1}{n-1} \cdots \frac{n-np-t+1}{n-t+1} \le (1-p)^t,\end{aligned}$$

since each term $\frac{n-np-i}{n-i} \le (1-p)$ for $i \ge 0$.
We can again write the precise equation for the tail bounds for any $J \subseteq I$:

$$\begin{aligned}\Pr\left[X_J \ge (1+\delta) \cdot |J|\, p\right] &= \sum_{i=\lceil (1+\delta)\cdot |J| p\rceil}^{|J|} \Pr[X_J = i] \\ &= \sum_{i=\lceil (1+\delta)\cdot |J| p\rceil}^{|J|} \frac{\binom{|J|}{i} \cdot \binom{n-|J|}{np-i}}{\binom{n}{np}}.\end{aligned}$$

The advantage of this distribution is that $X = np$ always; however the variables $\{X_1, \ldots, X_n\}$ are no longer independent. Consequently, for a $J \subseteq I$, the induced probability distribution on X_J is *not* the one where a $(|J|p)$-sized subset of J is chosen uniformly at random from the set of all $(|J|p)$-sized subsets of J.

The variables $X_1, \ldots, X_n$ are an example of *negatively associated* random variables. We note that in this case the tail bounds are even better—that is, more sharply concentrated around the expectation—than for binomial distribution. Intuitively, for any $i, j \in I$, the fact that $X_i = 1$ makes it *less* likely that $X_j = 1$ and the fact that $X_i = 0$ makes it *more* likely that $X_j = 1$.

Formally, if $p \in (0,1)$,

$$\Pr\left[X_j = 1 \mid X_i = 1\right] = \frac{\binom{n-2}{np-2}}{\binom{n-1}{np-1}} = \frac{np-1}{n-1} < p, \qquad \text{and}$$

$$\Pr\left[X_j = 1 \mid X_i = 0\right] = \frac{\binom{n-2}{np-1}}{\binom{n-1}{np}} = \frac{np}{n-1} > p.$$

Our main theorem, a multiplicative version of a tail bound for negatively associated random variables, is the following.

THEOREM 1.20. *Let $X_1, \ldots, X_n$ be n indicator random variables and let $\delta > 0$ be a given parameter. Set $X = \sum_{i=1}^n X_i$.*

(1) *Let $p_1, \ldots, p_n$ be reals in $[0,1]$, $0 < \sum_{i=1}^n p_i < n$, such that*

$$\text{for any } I' \subseteq [n] \qquad \Pr\left[\left(\prod_{i \in I'} X_i\right) = 1\right] \le \prod_{i \in I'} p_i.$$

Let $\tilde{p} = \frac{\sum_i p_i}{n}$. Then

$$\Pr\left[X \ge (1+\delta)\, n\tilde{p}\right] \le \left(\frac{\left(1 - \frac{\tilde{p}\delta}{1-\tilde{p}}\right)^{(1+\delta)\tilde{p}-1}}{(1+\delta)^{(1+\delta)\tilde{p}}}\right)^n. \tag{1.21}$$

The above expression can be simplified to give

$$\Pr\left[X \ge (1+\delta)\, n\tilde{p}\right] \le e^{-\frac{\delta^2}{2+\delta} n\tilde{p}}.$$

(2) *Let $r_1, \ldots, r_n$ be reals in $[0,1]$, $0 < \sum_{i=1}^n r_i < n$, such that*

$$\text{for any } I' \subseteq [n] \qquad \Pr\left[\left(\prod_{i \in I'} (1 - X_i)\right) = 1\right] \le \prod_{i \in I'} (1 - r_i).$$

Let $\tilde{r} = \frac{\sum_i r_i}{n}$. Then

$$\Pr\left[X \le (1-\delta)\, n\tilde{r}\right] \le \left(\frac{\left(1 + \frac{\delta\tilde{r}}{1-\tilde{r}}\right)^{-1+\tilde{r}-\delta\tilde{r}}}{(1-\delta)^{\tilde{r}(1-\delta)}}\right)^n. \tag{1.22}$$

The above expression can be simplified to give

$$\Pr\left[X \le (1-\delta)\, n\tilde{r}\right] \le e^{-\frac{\delta^2}{2} n\tilde{r}}.$$

We remark that the two preconditions of the above theorem imply that for any variable X_i, we have

$$\Pr\left[X_i = 1\right] \le p_i, \quad \text{and}$$
$$\Pr\left[(1 - X_i) = 1\right] \le 1 - r_i \quad \text{or equivalently,} \qquad \Pr\left[X_i = 1\right] \ge r_i.$$

Thus the variables p_i and r_i are upper and lower bounds on the probability that $X_i = 1$, and therefore $n\tilde{r} \le \mathrm{E}\left[\sum_i X_i\right] \le n\tilde{p}$.

The above theorem applies to *both* the two earlier distributions—binomial and sampling without replacement—as it avoids using independence of the X_i's and instead uses an upper bound on $\Pr\left[\left(\prod_{i\in J} X_i\right) = 1\right]$ and $\Pr\left[\left(\prod_{i\in J}(1-X_i)\right) = 1\right]$ for every $J \subseteq I$.

Binomial distribution: Theorem 1.20 and independence implies the following.

COROLLARY 1.23. *Let $I = \{1, \dots, n\}$ and $p_1, \dots, p_n \in [0,1]$ be given parameters. Let $R \subseteq I$ be a random sample constructed by picking each $i \in I$ independently with probability p_i. Then for a fixed $J \subseteq I$ and $\delta > 0$,*

$$\begin{aligned}\Pr\Big[\,|J\cap R| \geq (1+\delta)\,\mathrm{E}\left[|J\cap R|\right]\Big] &= \Pr\Big[\,|J\cap R| \geq (1+\delta)\sum_{j\in J} p_j\Big] \\ &\leq e^{-\frac{\delta^2}{2+\delta}\sum_{j\in J} p_j}, \\ \Pr\Big[\,|J\cap R| \leq (1-\delta)\,\mathrm{E}\left[|J\cap R|\right]\Big] &= \Pr\Big[\,|J\cap R| \leq (1-\delta)\sum_{j\in J} p_j\Big] \\ &\leq e^{-\frac{\delta^2}{2}\sum_{j\in J} p_j}.\end{aligned}$$

In particular,

$$\Pr\left[\,|J\cap R| \geq (1+\delta)\sum_{j\in J} p_j \quad \bigcup \quad |J\cap R| \leq (1-\delta)\sum_{j\in J} p_j\right] \leq 2\,e^{-\frac{\delta^2}{2+\delta}\sum_{j\in J} p_j}.$$

Sampling without replacement: Theorem 1.20 together with Equations (1.18) and (1.19) implies the following.

COROLLARY 1.24. *Let $I = \{1, \dots, n\}$ and $t \in [n]$ be a given parameter. Let $R \subseteq I$ be a random sample of size t chosen uniformly from all $\binom{n}{t}$ t-sized subsets of I. Then for any fixed $J \subseteq I$ and $\delta > 0$,*

$$\begin{aligned}\Pr\left[|J\cap R| \geq (1+\delta)\,|J|\frac{t}{n}\right] &\leq e^{-\frac{\delta^2}{2+\delta}|J|\frac{t}{n}}, \\ \Pr\left[|J\cap R| \leq (1-\delta)\,|J|\frac{t}{n}\right] &\leq e^{-\frac{\delta^2}{2}|J|\frac{t}{n}}.\end{aligned}$$

In particular,

$$\Pr\left[\,|J\cap R| \geq (1+\delta)\,|J|\frac{t}{n} \quad \bigcup \quad |J\cap R| \leq (1-\delta)\,|J|\frac{t}{n}\right] \leq 2\,e^{-\frac{\delta^2}{2+\delta}|J|\frac{t}{n}}.$$

We remark here that these bounds are tight within constant factors in the exponent for certain ranges of δ. Here is one lower bound (stated without proof; see discussion).

THEOREM 1.25. *Let $I = \{1, \dots, n\}$ and $p \in \left(0, \frac{1}{2}\right]$. Let $R \subseteq I$ be a random sample constructed by picking each $i \in I$ independently with probability p. Then for $\delta \in \left[\sqrt{\frac{3}{np}}, \frac{1}{2}\right]$,*

$$\begin{aligned}\Pr\Big[\,|R| \geq (1+\delta)\,np\Big] &\geq e^{-9\delta^2 np}, \\ \Pr\Big[\,|R| \leq (1-\delta)\,np\Big] &\geq e^{-9\delta^2 np}.\end{aligned}$$

Overview of ideas. The proof of Theorem 1.20 will use our probabilistic averaging technique. That is, we will take a random sample of $\{1, 2, \ldots, n\}$ and calculate the probability of a carefully chosen event due to it in two ways.

At first glance, it might seem odd to estimate the probability of a random event—in our case the tail bounds on X—by taking *another* random sample! However it is a mistake to confuse these two separate probability distributions, with very different purposes—one is part of the input problem and the other is part of the averaging proof technique.

Perhaps a more modular way to think about this is to consider the quantity we are bounding—$\Pr\left[X \geq (1+\delta)\, n\tilde{p}\right]$—*combinatorially*: the support of the probability distribution consists of 2^n binary strings corresponding to all possible assignments of the 0-1 variables $X_1, \ldots, X_n$. Each string $s \in \{0,1\}^n$ has some probability, say $w(s)$, of being chosen.

> Let $|s|$ denote the number of 1's in s. The precise value of $w(s)$ depends on the probability distribution. For example, when $p_1 = \cdots = p_n = p$ and where np is an integer, $w(s) = p^{|s|}(1-p)^{n-|s|}$ for the binomial distribution. Similarly $w(s) = 1/\binom{n}{np}$ if $|s| = np$, and 0 otherwise, for the sampling without replacement distribution.

Then our goal is to upper bound the combinatorial quantity

$$\sum_{\substack{s \in \{0,1\}^n \\ |s| \geq (1+\delta) n\tilde{p}}} w(s).$$

Seen this way, it is similar to the earlier use of the probabilistic averaging technique to upper bound the sizes of level sets, with one difference being that earlier we were bounding the cardinality instead of a weighted sum.

As a warm-up, we first prove the following weaker bound, called *Markov's inequality*, under the conditions of Theorem 1.20:

$$\Pr\left[X \geq (1+\delta)\, n\tilde{p}\right] \leq \frac{1}{1+\delta}. \tag{1.26}$$

While Markov's inequality has an even simpler direct proof (furthermore, Markov's inequality holds for *any* positive random variable X for which $\mathrm{E}[X]$ exists, with $\mathrm{E}[X]$ replacing $n\tilde{p}$ in the stated bound. That is, X need not be the sum of n indicator variables), the following proof is an easy natural application of the probabilistic averaging technique and gives insight into the proof of Theorem 1.20.

> Let S be a random sample of the index set $I = \{1, 2, \ldots, n\}$ where each index is picked independently with probability q. Note that S is independent of the X_i variables.
>
> We count the following quantity in two ways:
>
> $$\mathrm{E}\left[|S_1|\right], \quad \text{where} \quad S_1 = \{i \in S \colon X_i = 1\}.$$
>
> That is, the expected number of indices $i \in S$ for which $X_i = 1$.

Upper bound: Using linearity of expectation and the fact that we have $\Pr[X_i = 1] \le p_i$,

$$\mathrm{E}[|S_1|] = \sum_{i=1}^{n} \Pr[i \in S \text{ and } X_i = 1] \le \sum_{i=1}^{n} p_i q = n\tilde{p}\, q.$$

The last step used the fact that S and X are independent.

Lower bound: Consider the elements of the event space for the variable $X = X_1 + \cdots + X_n$ for which $X \ge (1+\delta)\, n\tilde{p}$. Note that for each event $\{X_1, \ldots, X_n\}$ with $X = k$, the expected number of indices $i \in S$ with $X_i = 1$ is precisely kq. Thus we have

$$\mathrm{E}\Big[\ |S_1| \ \Big| \ X \ge (1+\delta)\, n\tilde{p}\ \Big] \ge (1+\delta)\, n\tilde{p}\, q.$$

Summing up over all events,

$$\begin{aligned}\mathrm{E}[|S_1|] =& \mathrm{E}\left[|S_1| \mid X \ge (1+\delta)n\tilde{p}\right] \cdot \Pr[X \ge (1+\delta)n\tilde{p}] \ + \\ & \mathrm{E}\left[|S_1| \mid X < (1+\delta)n\tilde{p}\right] \cdot \Pr[X < (1+\delta)n\tilde{p}] \\ \ge& \mathrm{E}\left[|S_1| \mid X \ge (1+\delta)n\tilde{p}\right] \cdot \Pr[X \ge (1+\delta)n\tilde{p}] \\ \ge& (1+\delta)\, n\tilde{p}\, q \cdot \Pr[X \ge (1+\delta)\, n\tilde{p}] .\end{aligned}$$

Putting the upper and lower bounds together,

$$(1+\delta)\, n\tilde{p}\, q \cdot \ \Pr[X \ge (1+\delta)\, n\tilde{p}] \ \le \ \mathrm{E}[|S_1|] \ \le \ n\tilde{p}\, q,$$

$$\text{and hence} \qquad \Pr[X \ge (1+\delta)\, n\tilde{p}] \le \frac{1}{1+\delta}.$$

An astute reader will notice that the proof above is needlessly complicated, as the parameter q does not play any role: the dependence on q is linear in both the upper and lower bounds and thus cancels out. Setting $S = I$ (i.e., $q = 1$) gives the standard proof of Markov's inequality. This will not remain the case for the proof of the main theorem, to which we turn to next.

We now prove our main theorem.

Proof of Theorem 1.20. As before, let S be a random sample where each element in $\{1, 2, \ldots, n\}$ is picked independently with probability q.

We count the following quantity in two ways:

$\Pr\Big[\prod_{i \in S} X_i = 1\Big]$.

That is, the probability that for *each* index $i \in S$, $X_i = 1$.

Note that this probability is over both the choice of S and the choice of X. Furthermore S and X are independent.

Upper bound. It will be instructive to consider it in three, progressively more general, scenarios:

- **each $X_i = 1$ independently with probability $\boldsymbol{p_i}$:** Then we have

$$\begin{aligned}\Pr\left[\prod_{i\in S} X_i = 1\right] &= \prod_{i=1}^{n}\Big(1 - \Pr\left[i \in S \text{ and } X_i = 0\right]\Big)\\ &= \prod_{i=1}^{n}\left(1 - q\left(1 - p_i\right)\right) \le \left(\frac{\sum_{i=1}^{n}\left(1 - q\left(1 - p_i\right)\right)}{n}\right)^{n}\\ &= \left(\frac{n - nq + q\sum_{i=1}^{n} p_i}{n}\right)^{n} = \left(q\tilde{p} + 1 - q\right)^{n},\end{aligned}$$

where the third step uses the inequality of arithmetic and geometric means, that $\prod_{i=1}^{n} a_i \le \left(\frac{\sum_{i=1}^{n} a_i}{n}\right)^{n}$ for any non-negative reals $a_1, \ldots, a_n$.

- $\boldsymbol{p_1 = \cdots = p_n = p}$**:** Then $\tilde{p} = p$ and so

$$\begin{aligned}\Pr\left[\prod_{i\in S} X_i = 1\right] &= \sum_{Q\subseteq[n]} \Pr\Big[S = Q \text{ and } \prod_{i\in Q} X_i = 1\Big]\\ &= \sum_{Q\subseteq[n]} \Pr\left[S = Q\right]\cdot\Pr\left[\prod_{i\in Q} X_i = 1\right] \quad \Big(S,\ X \text{ are independent}\Big)\\ &\le \sum_{i=0}^{n}\binom{n}{i} q^{i}(1-q)^{n-i}\cdot p^{i} \quad \Big(\text{by input assumption}\Big)\\ &= (qp + 1 - q)^{n} \quad \Big(\text{by the binomial theorem}\Big).\end{aligned}$$

- **the general case:**

$$\begin{aligned}\Pr\left[\prod_{i\in S} X_i = 1\right] &= \sum_{Q\subseteq[n]} \Pr\left[S = Q\right]\cdot\Pr\left[\prod_{i\in Q} X_i = 1\right]\\ &\le \sum_{Q\subseteq[n]} q^{|Q|}(1-q)^{n-|Q|}\cdot\prod_{i\in Q} p_i \quad \Big(\text{by input assumption}\Big)\\ &= (1-q)^{n}\sum_{Q\subseteq[n]}\prod_{i\in Q}\frac{qp_i}{1-q} = (1-q)^{n}\prod_{i=1}^{n}\left(1 + \frac{qp_i}{1-q}\right),\end{aligned}$$

where the last step uses the fact that $\prod_{i=1}^{n}(1 + a_i) = \sum_{Q\subseteq[n]}\prod_{i\in Q} a_i$ (each term in the L.H.S. of this expression, when opened up, corresponds to a choice of either 1 or a from each of the n product terms). Continuing,

$$\begin{aligned}&= (1-q)^{n}\prod_{i=1}^{n}\left(\frac{qp_i + 1 - q}{1-q}\right) = \prod_{i=1}^{n}\left(qp_i + 1 - q\right)\\ &\le \left(q\tilde{p} + 1 - q\right)^{n} \quad \Big(\text{as earlier}\Big).\end{aligned}$$

Lower bound. Consider the elements of the event space of $X = X_1 + \cdots + X_n$ for which $X \ge (1+\delta)\, n\tilde{p}$. Note that for each instance of $\{X_1, \ldots, X_n\}$ with $X = k$, the probability that for *each* index $i \in S$ we have $X_i = 1$ is exactly $(1-q)^{n-k}$. In our case $k \geq (1+\delta)\, n\tilde{p}$ and since $(1-q)^{n-k}$ is mononotically increasing with k, we have

$$\Pr\left[\prod_{i\in S} X_i = 1 \;\middle|\; X \geq (1+\delta)\, n\tilde{p}\right] \geq (1-q)^{n-(1+\delta)n\tilde{p}}.$$

Summing up over all events,

$$\begin{aligned}
\Pr\left[\prod_{i\in S} X_i = 1\right] =& \Pr\left[\prod_{i\in S} X_i = 1 \;\middle|\; X \geq (1+\delta)\, n\tilde{p}\right] \cdot \Pr[X \geq (1+\delta)n\tilde{p}] \;\; + \\
& \Pr\left[\prod_{i\in S} X_i = 1 \;\middle|\; X < (1+\delta)\, n\tilde{p}\right] \cdot \Pr[X < (1+\delta)n\tilde{p}] \\
\geq& \Pr\left[\prod_{i\in S} X_i = 1 \;\middle|\; X \geq (1+\delta)\, n\tilde{p}\right] \cdot \Pr[X \geq (1+\delta)n\tilde{p}] \\
\geq& (1-q)^{n-(1+\delta)n\tilde{p}} \cdot \Pr\left[X \geq (1+\delta)\, n\tilde{p}\right].
\end{aligned}$$

Combining the upper and lower bounds,

$$\begin{aligned}
(1-q)^{n-(1+\delta)n\tilde{p}} \cdot \Pr\left[X \geq (1+\delta)\, n\tilde{p}\right] \;\; \leq \;\; \Pr\left[\prod_{i\in S} X_i = 1\right] \;\; \leq \;\; (1-q\,(1-\tilde{p}))^n \\
\Longrightarrow \Pr\left[X \geq (1+\delta)\, n\tilde{p}\right] \leq \left(\frac{1-q\,(1-\tilde{p})}{(1-q)^{1-(1+\delta)\tilde{p}}}\right)^n.
\end{aligned}$$

To minimize the R.H.S. of the above expression[3], we set $q = \frac{\delta}{(1-\tilde{p})(1+\delta)}$. Then

$$\begin{aligned}
\Pr\left[X \geq (1+\delta)\, n\tilde{p}\right] &\leq \left(\frac{1-\frac{\delta}{(1-\tilde{p})(1+\delta)}(1-\tilde{p})}{\left(1-\frac{\delta}{(1-\tilde{p})(1+\delta)}\right)^{1-(1+\delta)\tilde{p}}}\right)^n \\
&= \left(\frac{\frac{1}{1+\delta}}{\left(\frac{1-\tilde{p}-\tilde{p}\delta}{(1-\tilde{p})(1+\delta)}\right)^{1-(1+\delta)\tilde{p}}}\right)^n = \left(\frac{\left(\frac{1-\tilde{p}-\tilde{p}\delta}{1-\tilde{p}}\right)^{(1+\delta)\tilde{p}-1}}{(1+\delta)^{(1+\delta)\tilde{p}}}\right)^n \\
&= \left(\frac{\left(1-\frac{\tilde{p}\delta}{1-\tilde{p}}\right)^{(1+\delta)\tilde{p}-1}}{(1+\delta)^{(1+\delta)\tilde{p}}}\right)^n,
\end{aligned}$$

getting the required bound.

The other direction—an upper bound on the probability that the number of 1's in X is at most $(1-\delta)n\tilde{r}$—is equivalent to upper bounding the probability that the number of 0's in X is at least

$$n - (1-\delta)\, n\tilde{r} = \left(\frac{1-(1-\delta)\,\tilde{r}}{(1-\tilde{r})}\right) n\;(1-\tilde{r}) = \left(1+\frac{\delta\tilde{r}}{1-\tilde{r}}\right) n\,(1-\tilde{r})\,.$$

[3] The partial derivative w.r.t. q is $\frac{\left(n\tilde{p}\right)\left((1+\delta)(\tilde{p}-1)q+\delta\right)\left(((\tilde{p}-1)q+1)(1-q)^{\delta\tilde{p}+\tilde{p}-1}\right)^n}{(q-1)((\tilde{p}-1)q+1)}$.

Set $Y_i = 1 - X_i$ for $i = 1, \ldots, n$, and let $Y = \sum_i Y_i$. That is, $Y = n - X$ is a random variable denoting the number of 0's in X. Then

$$\Pr\left[X \leq (1-\delta)\, n\tilde{r}\right] = \Pr\left[Y \geq \left(1 + \frac{\delta\tilde{r}}{1-\tilde{r}}\right) n\,(1-\tilde{r})\right].$$

Thus we can apply the previous bound on the variable Y_i's, now with probabilities $(1 - r_i)$ instead of p_i. We have $\frac{\sum_{i=1}^{n}(1-r_i)}{n} = (1 - \tilde{r})$ and so from Equation (1.21) with $\delta' = \frac{\delta\tilde{r}}{1-\tilde{r}}$,

$$\begin{aligned}
\Pr\left[Y \geq (1+\delta')\, n\,(1-\tilde{r})\right] &\leq \left(\frac{\left(1 - \frac{(1-\tilde{r})\delta'}{1-(1-\tilde{r})}\right)^{(1+\delta')(1-\tilde{r})-1}}{(1+\delta')^{(1+\delta')(1-\tilde{r})}}\right)^{n} \\
&= \left(\frac{\left(1 - \frac{\delta\tilde{r}}{\tilde{r}}\right)^{(1+\frac{\delta\tilde{r}}{1-\tilde{r}})(1-\tilde{r})-1}}{\left(1 + \frac{\delta\tilde{r}}{1-\tilde{r}}\right)^{(1+\frac{\delta\tilde{r}}{1-\tilde{r}})(1-\tilde{r})}}\right)^{n} \\
&= \left(\frac{(1-\delta)^{-\tilde{r}(1-\delta)}}{\left(1 + \frac{\delta\tilde{r}}{1-\tilde{r}}\right)^{1-\tilde{r}+\delta\tilde{r}}}\right)^{n} = \left(\frac{\left(1 + \frac{\delta\tilde{r}}{1-\tilde{r}}\right)^{-1+\tilde{r}-\delta\tilde{r}}}{(1-\delta)^{\tilde{r}(1-\delta)}}\right)^{n},
\end{aligned}$$

getting the required bound.

The simplifications of these expressions are covered in many places and we refer the reader to existing literature on this (see discussion). □

Bibliography and discussion. The proof given of these tail bounds is from [**IK10**], while Theorem 1.25 is from [**KY15**]. A nice exposition of several proofs of tail bounds similar to the one presented here (Chernoff's bound, Bernstein's inequality, Hoeffding's extension) together with the details of simplification of the expressions in Theorem 1.20 can be found in [**Mul18**] (see also [**DP09**]). A discussion on the differences between sampling with and without replacement can be found in [**FK15**, Section 21.5]. A discussion on the asymmetry between the upper and lower tail bounds in Theorem 1.20 can be found in [**AS16**, Appendix A]. Some approximations for sums of binomial coefficients (such as those of Equation (1.17)) can be found in [**GKP94**, Chapter 5] (see also [**Spi19**]). Many other concentration inequalities can be found in the text [**BLM13**].

[AS16] N. Alon and J. H. Spencer, *The probabilistic method*, 4th ed., Wiley Series in Discrete Mathematics and Optimization, John Wiley & Sons, Inc., Hoboken, NJ, 2016. MR3524748

[BLM13] S. Boucheron, G. Lugosi, and P. Massart, *Concentration inequalities*, Oxford University Press, Oxford, 2013. A nonasymptotic theory of independence; With a foreword by Michel Ledoux, DOI 10.1093/acprof:oso/9780199535255.001.0001. MR3185193

[DP09] D. P. Dubhashi and A. Panconesi, *Concentration of measure for the analysis of randomized algorithms*, Cambridge University Press, Cambridge, 2009, DOI 10.1017/CBO9780511581274. MR2547432

[FK15] A. Frieze and M. Karoński, *Introduction to random graphs*, Cambridge University Press, Cambridge, 2016, DOI 10.1017/CBO9781316339831. MR3675279

[GKP94] R. L. Graham, D. E. Knuth, and O. Patashnik, *Concrete mathematics*, 2nd ed., Addison-Wesley Publishing Company, Reading, MA, 1994. A foundation for computer science. MR1397498

[IK10] R. Impagliazzo and V. Kabanets, *Constructive proofs of concentration bounds*, Approximation, randomization, and combinatorial optimization, Lecture Notes in Comput. Sci., vol. 6302, Springer, Berlin, 2010, pp. 617–631, DOI 10.1007/978-3-642-15369-3_46. MR2755867

[KY15] P. Klein and N. E. Young, *On the number of iterations for Dantzig-Wolfe optimization and packing-covering approximation algorithms*, SIAM J. Comput. **44** (2015), no. 4, 1154–1172, DOI 10.1137/12087222X. MR3390154

[Mul18] W. Mulzer, *Five proofs of Chernoff's bound with applications*, Bull. Eur. Assoc. Theor. Comput. Sci. EATCS **124** (2018), 59–76. MR3793013

[Spi19] M. Z. Spivey, *The art of proving binomial identities*, Discrete Mathematics and its Applications (Boca Raton), CRC Press, Boca Raton, FL, 2019, DOI 10.1201/9781351215824. MR3931743