1 Some applications

In the preceding lecture, we saw our first large-deviation inequality. Let $X_1, X_2, \ldots, X_n$ be a family of independent (but not necessarily identically distributed) $\{0,1\}$ random variables and $X = X_1 + X_2 + \cdots + X_n$. Let $p_i = \mathbb{E}[X_i]$ and define $\mu = \mathbb{E}[X] = p_1 + p_2 + \cdots + p_n$. We recall the following multiplicative form of the "Chernoff bound."

Theorem 1.1 (Chernoff bound, mutiplicative error)**.** *For every $\beta \geqslant 1$, it holds that*

$$\mathbb{P}\left(X \geqslant \beta\mu\right) \leqslant \left(\frac{e^{\beta-1}}{\beta^\beta}\right)^\mu, \tag{1.1}$$

and

$$\mathbb{P}\left(X \leqslant \frac{\mu}{\beta}\right) \leqslant \left(\frac{e^{1/\beta-1}}{\beta^\beta}\right)^\mu. \tag{1.2}$$

1.1 Balls in bins

Suppose we throw m balls uniformly at random into n bins. For $i = 1, \ldots, n$, let $X^{(i)}$ denote the number of balls that land in bin i. Let $Z := \max(X^{(1)}, \ldots, X^{(n)})$ denote the maximum load. Even to bound $\mathbb{E}[Z]$ seems tricky, and it is often the case that evaluating the expected maximum of a family of random variables requires understanding their tail behavior.

Let $X_j^{(i)}$ be the indicator random variable that is 1 if ball j lands in bin i. Then $\mathbb{E}[X_j^{(i)}] = 1/n$ and hence by linearity of expectation, $\mathbb{E}[X^{(i)}] = m/n$. Applying Theorem 1.1 yields, for any $i = 1, \ldots, n$,

$$\mathbb{P}\left[X^{(i)} \geqslant \beta\frac{m}{n}\right] \leqslant \left(\frac{e}{\beta}\right)^{\beta m/n}. \tag{1.3}$$

Let's consider two representative regimes.

Regime I: $m = n$**.** By choosing $\beta = \frac{c \log n}{\log\log n}$ for $c > 1$ large enough (as in the preceding lecture), (1.3) gives

$$\mathbb{P}\left[X^{(i)} \geqslant \beta\right] \leqslant n^{-2}$$

Now the deviation probability is small enough to apply a union bound:

$$\mathbb{P}[Z \geqslant \beta] \leqslant \sum_{i=1}^n \mathbb{P}[X^{(i)} \geqslant \beta] \leqslant n \cdot n^{-3} = \frac{1}{n}.$$

Thus with high probability, the maximum load is $O(\frac{\log n}{\log\log n})$.

Regime II: $m \geqslant cn\log n$**,** $c > 0$**.** In this case, we have $m/n \geqslant c\log n$, so applying (1.3) and a union bound gives

$$\mathbb{P}[Z \geqslant \beta] \leqslant n \cdot \beta^{-c\beta\log n}.$$

Now choosing $\beta \asymp 1/c$ gives $\mathbb{P}[Z \geqslant \beta] \leqslant \frac{1}{n}$, implying that the maximum load is only an $O(1)$ factor more than its expectation.

In comparing the two regimes, note that as the expected number of balls per bin rises, we actually get more concentration around the mean.

1.2 Randomized Quicksort

I don't particularly like this application since it requires some unnatural machinations to get independent random variables. In the next lecture, we will see that large-deviation inequalities hold for martingales, and this argument becomes more natural.

But a quick recap: Consider the numbers $\{1, 2, \ldots, n\}$. We construct a random rooted tree $\boldsymbol{T}$ where each node $v \in V(\boldsymbol{T})$ has an associated subset $S_v \subseteq \{1, \ldots, n\}$ defined as follows: For the root $r \in V(\boldsymbol{T})$, we have $S_r = \{1, \ldots, n\}$. Then inductively, for a node v with $|S_v| > 1$, we partition S_v uniformly at random into two sets S_v^L, S_v^R with $|S_v^L|, |S_v^R| \geqslant 1$, and we give v two children labeled by these sets. Thus $\boldsymbol{T}$ has precisely n leaves labeled by the singleton sets $\{1\}, \ldots, \{n\}$.

Let D_i denote the depth of the leaf labeled by $\{i\}$. The following claim is straightforward to verify inductively: The number of comparisons made by Quicksort is precisely $D_1 + D_2 + \cdots + D_n$. (Strictly speaking, this is only true because we are including the pivot in one of the two child lists, but a more clever implementation would only do better.)

Claim 1.2. *There is a constant $C \geqslant 1$ such that for any $i \in \{1, 2, \ldots, n\}$, it holds that*

$$\mathbb{P}[D_i \geqslant C \log n] \leqslant n^{-2}.$$

Taking a union bound gives

$$\mathbb{P}[\#\text{ comparisons} > Cn \log n] \leqslant \frac{1}{n},$$

i.e. Quicksort runs in $O(n \log n)$ time with high probability.

Fix an element $i \in \{1, \ldots, n\}$. Let $S_0, S_1, \ldots, S_{D_i} = \{i\}$ be the labels of the nodes occuring from the root down to the leaf labeled $\{i\}$ in $\boldsymbol{T}$. Define $S_j := \{i\}$ for $j > D_i$. Then an elementary calculation gives

$$\mathbb{P}\left[|S_{j+1}| \leqslant \frac{1}{2}|S_j|\right] \geqslant \frac{1}{2} \qquad \forall j \geqslant 0.$$

Define:

$$Y_j = \begin{cases} 1 & \text{if } |S_{j+1}| \leqslant \frac{1}{2}|S_j| \text{ or } S_{j+1} = S_j \\ 0 & \text{otherwise.} \end{cases}$$

The next claim is straightforward to verify.

Claim 1.3. *For every $j \geqslant 0$, we have $\mathbb{P}[Y_j = 1] \geqslant \frac{1}{2}$. Moreover, if $\sum_{j=0}^{M-1} Y_j \geqslant \log_2 n$, then $D_i \leqslant M$.*

If $\{Y_j\}$ were independent random variables, then we could apply (1.2) to conclude that

$$\mathbb{P}\left[Y_0 + Y_1 + \cdots + Y_{M-1} \leqslant \frac{M}{2\beta}\right] \leqslant \left(\frac{e^{1/\beta}}{\beta}\right)^{\beta M/2}$$

Choosing $M = \Theta(\log n)$ and $\beta = \Theta(1)$ would then yield Claim 1.2.

Unfortunately, these random variables are not independent, as clearly $Y_j = 1$ for $j \geqslant D_i$, for instance. One solution is to use a hack: We can couple the $\{Y_j\}$ random variables to *independent* random variables $\{\tilde{Y}_j\}$ such that $\tilde{Y}_j = 1 \implies Y_j = 1$ and $\mathbb{P}[\tilde{Y}_j = 1] = 1/2$. Then we can legitimately apply the Chernoff bound to the family $\{\tilde{Y}_j\}$ and reach the same conclusion.

This is easy to do by defining

$$\tilde{Y}_j = Z_j Y_j,$$

where $\{Z_j\}$ is a collection of $\{0,1\}$ random variables such that $Z_j = 1$ with probability $1/(2\,\mathbb{P}[Y_j = 1])$ (so that Z_j is independent of Y_j conditioned on S_j). Note that this definition makes sense since $\mathbb{P}[Y_j = 1] \geqslant 1/2$.

Now we have $\mathbb{P}[\tilde{Y}_j = 1] = 1/2$ for all $j \geqslant 0$, and moreover, the random variables $\{\tilde{Y}_j\}$ are independent:

$$\mathbb{P}\left[\tilde{Y}_j = 1 \mid \tilde{Y}_0, \ldots, \tilde{Y}_{j-1}, \tilde{Y}_{j+1}, \ldots\right] = \frac{1}{2} \qquad \forall j \geqslant 0\,.$$

Even this preceding fact is slightly tricky to verify and the whole argument doesn't reflect one's natural intuition: Independence shouldn't matter as long as we have probability at least 1/2 to reduce the size of S_{j+1} conditioned on S_j with $|S_j| > 1$. That is the purview of *martingale theory*, and we will cover large-deviation inequalities for martingales in the next lecture.

1.3 Negative correlation

Say that a collection $\{X_1, \ldots, X_n\}$ of random variables are *negatively correlated* if it holds that for any subset $S \subseteq [n]$:

$$\mathbb{E}\left[\prod_{i=1}^{n} X_i\right] \leqslant \prod_{i=1}^{n} \mathbb{E}[X_i]\,. \tag{1.4}$$

Note that if $\{X_1, \ldots, X_n\}$ are independent, then this holds with equality.

Examples. We will state some examples of negatively correlated families (without proof).

1. **Loads of the bins.** The variables $\{X^{(i)} : i = 1, \ldots, n\}$ from Section 1.1 are negatively correlated.

   This stands to reason: Telling you that some bins have unusually large (resp., small) load makes the expected load of the remaining bins smaller (resp., larger).

2. **Random permutations.** If $\{X_1, \ldots, X_n\} = \{1, 2, \ldots, n\}$, then the family $\{X_i\}$ is negatively correlated.

   The intuition here is the same as the previous example.

3. **Random spanning trees.** Suppose $G = (V, E)$ is an undirected graph and $\boldsymbol{T}$ is a uniformly random spanning tree of G. For $e \in E$, let X_e denote the indicator random variable that is 1 precisely when e is an edge of $\boldsymbol{T}$. Then the family $\{X_e : e \in E\}$ is negatively correlated.

   Suppose I tell you that $X_{e_1} = \cdots = X_{e_k} = 1$ for some edges $e_1, \ldots, e_k \in E$. Intuitively, one can contract the connected components in the graph spanned by $\{e_1, \ldots, e_k\}$ and consider a uniformly random spanning tree on the rest. Now the problem of connecting everything together has become easier, and thus $\mathbb{P}[X_e]$ decreases for $e \notin \{e_1, \ldots, e_k\}$. Certainly if e connects two vertices that are already connected in the graph with edges $\{e_1, \ldots, e_k\}$, then $\mathbb{P}[X_e = 1 \mid X_{e_1} = \cdots = X_{e_k} = 1] = 0$.

   Actually proving negative correlation for this family is non-trivial.

It turns out that the Chernoff bounds Theorem 1.1 hold if we consider negatively correlated random variables. It is still an area of active research to determine good notions for "negative dependence" in general settings. In particular, the notion of of negative correlation above is unsuitable for many settings, especially because it can be hard to verify and does not satisfy natural closure properties (making it difficult to derive new negatively correlated families from old ones).

Theorem 1.4 (Chernoff for negatively correlated random variables). *If, in the statement of Theorem 1.1, we only assume that $X_1, \ldots, X_n$ are negatively correlated $\{0,1\}$ random variables (instead of independent), then the conclusion still holds.*

Proof. To see this, note that the one place we used independence in the proof of Theorem 1.1 is in the calculation: When $X = X_1 + \cdots + X_n$,

$$\mathbb{E}\left[e^{tX}\right] = \mathbb{E}\left[\prod_{i=1}^{n} e^{tX_i}\right] = \prod_{i=1}^{n} \mathbb{E}\left[e^{tX_i}\right].$$

Let us see that the inequality $\mathbb{E}\left[e^{tX}\right] \leqslant \prod_{i=1}^{n} \mathbb{E}\left[e^{tX_i}\right]$ still holds when $\{X_1, \ldots, X_n\}$ are only assumed to be negatively correlated.

To this end, let $\{\tilde{X}_1, \ldots, \tilde{X}_n\}$ be *independent* $\{0,1\}$ random variables with $\mathbb{E}[\tilde{X}_i] = \mathbb{E}[X_i]$ for each $i = 1, \ldots, n$, and define $\tilde{X} := \tilde{X}_1 + \cdots + \tilde{X}_n$. For any nonnegative integer k,

$$\mathbb{E}[X^k] = \sum_{\alpha} \mathbb{E}[X_1^{\alpha_1} X_2^{\alpha_2} \cdots X_n^{\alpha_n}],$$

where the sum is over all $\alpha \in \mathbb{R}^n$ with $\alpha_i \geqslant 0$ and $\sum_i \alpha_i = k$. Using the negative correlation property, this gives

$$\mathbb{E}[X^k] \leqslant \sum_{\alpha} \mathbb{E}[X_1^{\alpha_1}]\,\mathbb{E}[X_2^{\alpha_2}] \cdots \mathbb{E}[X_n^{\alpha_n}] = \sum_{\alpha} \mathbb{E}[\tilde{X}_1^{\alpha_1}]\,\mathbb{E}[\tilde{X}_2^{\alpha_2}] \cdots \mathbb{E}[\tilde{X}_n^{\alpha_n}],$$

where the last line follows because X_i and $\tilde{X}_i$ have the same distribution for every i. Finally, note that by independence,

$$\sum_{\alpha} \mathbb{E}[\tilde{X}_1^{\alpha_1}]\,\mathbb{E}[\tilde{X}_2^{\alpha_2}] \cdots \mathbb{E}[\tilde{X}_n^{\alpha_n}] = \sum_{\alpha} \mathbb{E}[\tilde{X}_1^{\alpha_1} \tilde{X}_2^{\alpha_2} \cdots \tilde{X}_n^{\alpha_n}] = \mathbb{E}[\tilde{X}^k].$$

We conclude that for every integer $k \geqslant 0$,

$$\mathbb{E}[X^k] \leqslant \mathbb{E}[\tilde{X}^k]. \tag{1.5}$$

Using the Taylor expansion

$$e^{tX} = 1 + tX + \frac{t^2 X^2}{2} + \frac{t^3 X^3}{6} + \cdots,$$

and applying (1.5) to each term gives

$$\mathbb{E}[e^{tX}] \leqslant \mathbb{E}[e^{t\tilde{X}}] = \prod_{i=1}^{n} \mathbb{E}[e^{t\tilde{X}_i}] = \prod_{i=1}^{n} \mathbb{E}[e^{tX_i}],$$

yielding our desired inequality. Now the proof of the Chernoff bound can proceed exactly as in the preceding lecture. □