1 Chernoff bounds

In the last lecture, we saw some applications of the Chernoff bound. In a moment, we’ll prove it. Consider a sum \( X = X_1 + \cdots + X_n \) of independent random variables. In order to prove exponential-type tail bounds, we need to use the full independence in a strong way. If, for instance, the random variables \( \{X_i\} \) are only pairwise independent, then it’s easy to construct examples where Chebyshev’s inequality is tight.

The key to exploiting independence is the fact that \( Z \) and \( Z' \) are independent, then \( \mathbb{E}[ZZ'] = \mathbb{E}[Z] \cdot \mathbb{E}[Z'] \). Of course, this extends to products of many independent random variables by induction: \( \mathbb{E}[Z_1 Z_2 \cdots Z_k] = \mathbb{E}[Z_1] \cdot \mathbb{E}[Z_2] \cdots \mathbb{E}[Z_k] \).

Laplace transforms. To get an expression involving products of every subset of \( k \) variables, we’ll use a trick known as the method of Laplace transforms. The basic idea is as follows: Fix a real number \( x \), and let \( t > 0 \) be a parameter. Using the Taylor expansion, we can write

\[
e^{tx} = \sum_{k=0}^{\infty} \frac{t^k x^k}{k!}.
\]

One should consider that no matter what the values of \( t \) and \( x \), the \( k! \) factor in the denominator eventually wins out since it grows faster than \( k^c \) for some constant \( c > 0 \).

Once \( k! \gg t^k x^k \), the contribution from the remaining terms go down quickly. We could truncate the sum at \( k \approx t|x| \) without losing much of its value. On the other hand, while \( k! \ll t^k x^k \), the numerator is growing quickly as \( k \) increases. Thus the bulk of the sum is concentrated near \( k \approx t|x| \). This means that by choosing an appropriate value of \( t \), we can get the sum to “concentrate” on powers \( x^k \) for \( k \) in a range around \( t|x| \).

If we apply this with \( x \) equal to our random variable \( X \), we get

\[
e^{tX} = \sum_{k=0}^{\infty} \frac{t^k (X_1 + \cdots + X_n)^k}{k!} = \sum_{k=0}^{\infty} \frac{t^k}{k!} \sum_{i_1, \ldots, i_k} X_{i_1} \cdots X_{i_k},
\]

where the latter sum is over all the ways of taking products of \( k \) factors. Taking expectations of both sides and using independence yields

\[
\mathbb{E}[e^{tX}] = \sum_{k=0}^{\infty} \frac{t^k}{k!} \sum_{i_1, \ldots, i_k} \mathbb{E}[X_{i_1}] \cdots \mathbb{E}[X_{i_k}].
\]

This equality uses independence of every subset of the variables \( \{X_i\} \). Moreover, as we just discussed, changing the value of \( t \) allows us to choose which values of \( k \) the sum is “focused on.” This gives us a way of exploring various higher moments of the random variables. The benefit of this method is that one can work entirely in the exponential setting without passing to the messy right-hand side.

**Theorem 1.1** (Chernoff bound, multiplicative error). Suppose that \( X = X_1 + X_2 + \cdots + X_n \) is a sum of independent \( \{0, 1\} \) random variables and \( \mu = \mathbb{E}[X] \). For every \( \beta \geq 1 \), we have

\[
P(X \geq \beta \mu) \leq \left( \frac{e^{\beta-1}}{\beta^\beta} \right)^\mu,
\]
and
\[
P(X \leq \frac{\mu}{\beta}) \leq \left(\frac{e^{1/\beta - 1}}{\beta^\beta}\right)^\mu.
\]

\textbf{Proof.} We will prove the first inequality. The second one is proved similarly. Let \(\lambda\) and \(t\) be positive numbers to be chosen later. By monotonicity of the function \(x \mapsto e^{tx}\), we have

\[
P[X \geq \lambda] = P[e^{tX} \geq e^{t\lambda}] \leq \frac{E[e^{tX}]}{e^{t\lambda}}
\]

where we have used Markov’s inequality applied to the random variable \(e^{tX}\).

Using independence, we have
\[
e^{tX} = \prod_{i=1}^{n} E[e^{tX_i}].
\]

Suppose now that \(p_i = E[X_i]\). Then we have
\[
E[e^{tX_i}] = p_ie^t + (1 - p_i) = 1 + p_i(e^t - 1) \leq e^{p_i(e^t - 1)},
\]
where we have used the inequality \(1 + x \leq e^x\) valid for all numbers \(x\). Plugging this in yields
\[
E[e^{tX}] \leq \prod_{i=1}^{n} e^{p_i(e^t - 1)} \leq e^{(e^t - 1)\sum_{i=1}^{n} p_i} = e^{(e^t - 1)\mu}.
\]

Finally, we choose \(t = \ln \beta\) and \(\lambda = \beta \mu\) and plug this bound in (1.1) to obtain
\[
P[X \geq \beta \mu] \leq \frac{e^{(\beta - 1)\mu}}{\beta^{\mu \ln \beta}} = \left(\frac{e^{\beta - 1}}{\beta^\beta}\right)^\mu.
\]

By a similar method (but optimizing the choice of \(t\) differently), one can also prove the bounds:

For every \(t > 0\):
\[
P[X \geq \mu + tn] \leq e^{-2t^2n}
\]
\[
P[X \leq \mu - tn] \leq e^{-2t^2n}.
\]

2