Concentration Inequalities

EECS 126 (UC Berkeley)

Fall 2021

Suppose we have a collection of scalar random variables $X_1, \dots, X_n$. We may often wish to analyze the distribution of their sum (or equivalently, their average)

$$S_n = X_1 + \dots + X_n.$$

It turns out that, assuming that our X_i have a sufficient amount of regularity and independence, which will be quantified throughout this section, the probability of being close to the mean will sharply concentrate in a relatively narrow range.

1 The Moment Method, Chernoff/Hoeffding Bounds

The first moment method should be a familiar application of Markov's inequality,

$$\mathbb{P}(|S_n| \geq \lambda) \leq \frac{1}{\lambda} \sum_{i=1}^{n} \mathbb{E}|X_i|, \tag{1}$$

as should the second moment method, an application of Chebyshev's inequality,

$$\mathbb{P}(|S_n| \geq \lambda) \leq \frac{1}{\lambda^2} \sum_{i=1}^{n} \operatorname{Var}(X_i), \tag{2}$$

where we have assumed that the X_i are pairwise independent.

Exercise 1.1. Come up with examples of random variables $X_1, \dots, X_n$ in which (1) and (2) are tight.

We can play a similar game with k-th moments, by assuming k-wise independence. We'd have to do some combinatorial bookkeeping with the terms in

$$\mathbb{E}|S_n|^k = \sum_{1 \leq i_1, \dots, i_k \leq n} \mathbb{E} X_{i_1} \dots X_{i_k},$$

and after some algebra involving Stirling's formula, we can arrive at the large deviation bound

$$\mathbb{P}(|S_n| \geq \lambda\sqrt{n}) \leq 2 \left(\frac{\sqrt{ek/2}}{\lambda} \right)^k. \tag{3}$$

But instead of dwelling on this, we can often obtain much better bounds using exponential moments, namely by considering the moment generating function $\mathbb{E}e^{tS_n}$. The following is a bound for a single random variable which will be useful.

Notes by Albert Zhang

Lemma 1.2 (Hoeffding's lemma)
If X is a scalar random variable taking values in $[a,b]$, then for any $t>0$,

$$\mathbb{E}e^{tX} \le e^{t\mathbb{E}X}\left(1+O\left(t^2\mathrm{Var}(X)e^{O(t(b-a))}\right)\right), \tag{4}$$

and in particular,

$$\mathbb{E}e^{tX} \le e^{t\mathbb{E}X}e^{O(t^2\mathrm{Var}(X))} \le e^{t\mathbb{E}X}e^{O(t^2(b-a)^2)}. \tag{5}$$

Proof. Note that we can subtract the mean from X, a, b and assume that $\mathbb{E}X=0$. Furthermore, by normalizing X we can assume that $b-a=1$. Then $X=O(1)$, and we have the Taylor expansion

$$\begin{aligned} e^{tX} &= 1+tX+(tX)^2\left(\frac{1}{2}+\frac{tX}{3!}+\cdots\right) \\ &= 1+tX+O\left(t^2X^2e^{O(t)}\right). \end{aligned}$$

Taking expectations gives us

$$\mathbb{E}e^{tX} = 1+O\left(t^2\mathrm{Var}(X)e^{O(t)}\right),$$

proving (4). To get the other bound, note that $\mathrm{Var}(X)\le(b-a)^2$, and consider the function

$$f(x)=\frac{1+x^2e^x}{e^{x^2}},$$

which can be shown to be bounded. With $x=t(b-a)$, this gives us (5). □

Using some calculus, we can sharpen Hoeffding's lemma to the following explicit bound:

Theorem 1.3 (Chernoff bound)
Let $X_1,\ldots,X_n$ be independent scalar random variables with $|X_i|\le K$ almost surely, with means μ_i and variances σ_i^2. Then for any $\lambda>0$, we have

$$\mathbb{P}(|S_n-\mu|\ge\lambda\sigma)\le C\max\left(e^{-c\lambda^2}, e^{-c\lambda\sigma/K}\right), \tag{6}$$

where $C, c>0$ are constants, $\mu := \sum_{i=1}^n \mu_i$, and $\sigma^2 := \sum_{i=1}^n \sigma_i^2$.

Proof. We may assume that $\mu_i=0$ and $K=1$. It then suffices to prove the upper tail bound

$$\mathbb{P}(S_n\ge\lambda\sigma)\le C\max\left(e^{-c\lambda^2}, e^{-c\lambda\sigma}\right).$$

Note that by independence,

$$\mathbb{E}e^{tS_n}=\prod_{i=1}^n \mathbb{E}e^{tX_i}.$$

By (5) and the fact that $|X|\le 1$, we have

$$\mathbb{E}e^{tX_i}\le e^{O(t^2\sigma_i^2)},$$

and together this gives us

$$\mathbb{E}e^{tS_n} \leq e^{O(t^2\sigma^2)}.$$

By Markov's inequality, one has

$$\mathbb{P}(S_n \geq \lambda\sigma) \leq e^{O(t^2\sigma^2)-t\lambda\sigma}.$$

Optimizing over t subject to the constraint $t \in [0,1]$ gives us (6). □

Exercise 1.4. By letting t take values in some larger interval than $[0,1]$, show that the term $e^{-c\lambda\sigma/K}$ in the Chernoff bound can be replaced with $(\lambda K/\sigma)^{-c\lambda\sigma/K}$, which is better for when $\lambda K \gg \sigma$.

Corollary 1.5 (Hoeffding bound)

Let $X_1, \ldots, X_n$ be independent random variables taking values in intervals $[a_i, b_i]$, respectively. Then

$$\mathbb{P}(|S_n - \mathbb{E}S_n| \geq \lambda\sigma) \leq Ce^{-c\lambda^2},$$

where $C, c > 0$ are constants and $\sigma^2 := \sum_{i=1}^n |b_i - a_i|^2$.

Proof. This follows from Chernoff's bound and the assumption that $\mathrm{Var}(S_n) = \sum_{i=1}^n |b_i - a_i|^2$. □