CS 744: Pseudorandomness Generators

Lecture 10: Chernoff bound

January 25, 2015

Suppose we have n random variables, say $X_1, \ldots, X_n$, what can we conclude about their sum $S = X_1 + \cdots + X_n$?

This question has implications to our basic idea of boosting the correctness probability of BPP algorithms. For simplicity, let us assume that the BPP algorithm we are interested in, outputs either 0 or 1. One basic technique is to repeat the output on the same instance x of size n, and output the majority.

The majority function is somewhat difficult to handle mathematically, so we will change our viewpoint, and ask what is the *sum* of the outputs? If the sum is greater than $n/2$, then the majority is 1, and if the sum is less than $n/2$, then the majority is 0. So the probabibility that the majority is wrong by a signifant amount is the probability that S is significantly less than $n/2$.

Assume that the random variables are bounded, without loss of generality, let each $X_i \in [-1, 1]$. If the random variables are sufficiently independent, the sum S instead of having the maximum magnitude of n, will typically concentrated in a range of values of size $O(\sqrt{n})$. Thus most of the samples of the n random variables will lead to a sum in the range $ES + O(\sqrt{n})$. This phenomenon is known as *concentration of measure*. The intuitive reason is that independent random variables are unlikely to all go in one direction, which is necessary for a large magnitude sum.

Moreover, with a greater amount of independence, the concentration is *sub-gaussian* — $Ce^{-c\lambda^2}$, for some $C, c > 0$. With high probability, the sum is about $ES + O((\sqrt{\log n})\sigma)$, where σ is the standard deviation of S. With *overwhelming* probability, the sum is in the range $ES + O((\log^{1/2+\varepsilon} n)\sigma)$. Assuming that the random S has mean 0 and variance 1, this is about $O(\log^{1/2+\varepsilon} n)$.

This tighter bound has obvious consequences to error reduction of BPP algorithms - this means that the error can be made around $\frac{1}{2^n}$ in polynomially many repetitions.

1 Chernoff Bound

Lemma 1. *[Hoeffding's lemma] Let X be a random variable taking values in $[a, b]$. Then for any $t > 0$,*

$$E[e^{tX}] \leq e^{tEX}(1 + O(t^2\ Var(X)\ e^{O(t(b-a))})).$$

Proof. We assume that $EX = 0$ and $b - a = 1$. This implies that $X = O(1)$. We have

$$e^{tX} \;=\; \sum_{i=0}^{\infty} \frac{(tX)^i}{i!} \;=\; 1 + tX + O(t^2 X^2 e^{O(t)}).$$

Taking expectations on both sides, we have

$$\begin{aligned} E[e^{tX}] &= 1 + tEX + O(t^2\ E[X^2]\ e^{O(t)}) \\ &= 1 + O(t^2\ E[X^2]\ e^{O(t)}) \\ &= 1 + O(t^2\ \mathrm{Var}(X)\ e^{O(t)}), \end{aligned}$$

proving the claim. □

Corollary 2. *Let X be a random variable taking values in $[a, b]$. Then for any $t > 0$,*

$$E[e^{tX}] \leq e^{tEX}\ e^{O(t^2(b-a)^2)}.$$

Proof. By the Theorem above,

$$E[e^{tX}] \leq e^{tEX}(1 + O(t^2\ \mathrm{Var}(X)\ e^{O(t(b-a))})).$$

Since $t^2\ \mathrm{Var}(X) \leq t^2(b-a)^2$, we have

$$E[e^{tX}] \leq e^{tEX}(1 + O(t^2\ (b-a)^2\ e^{O(t(b-a))})).$$

Since $x^2 e^x = e^{2\log x + x} = e^{O(x^2)}$, we have

$$E[e^{tX}] \leq e^{tEX} e^{O(t^2(b-a)^2)},$$

thus establishing the claim. □

This can be used to establish the Chernoff bound.

Theorem 3. *Let $X_1, \ldots, X_n$ be independent random variables with $|X_i| < K$, mean μ_i and variance σ_i^2. Then for any $\lambda > 0$, one has, for some $C, c > 0$,*

$$P(|S - \mu| \;\geq\; \lambda\sigma) \quad\leq\quad C \max(e^{-c\lambda^2}, e^{-c\lambda\sigma/K}),$$

where $\mu = \sum_{i=1}^n \mu_i$ and $\sigma^2 = \sum_{i=1}^n \sigma_i^2$.

Proof. Let $t \in [0, 1]$. Let $Y_i = \frac{X_i - \mu_i}{K}$. Then $EY_i = 0$, and $|Y_i| \leq 1$. Let $T = \sum_{i=1}^n Y_i$. By the symmetry of Y_i around 0, it suffices to establish the following.

$$P(T > \lambda\sigma) \leq C \max\left(e^{-c\lambda^2}, e^{-c\lambda\sigma}\right).$$

To do this, consider the exponential moment generating function of T:

$$E[e^{tT}], \quad \text{where } t \in [0,1].$$

We have

$$E[e^{tT}] \;=\; E\left[e^{t\sum_{i=1}^{n} Y_i}\right] = \prod_{i=1}^{n} E[e^{tY_i}],$$

where the last equality follows by the mutual independence of Y_is. Now, we use the Hoeffding bound on each individual Y_i,

$$E[e^{tY_i}] \;\leq\; e^{O(t^2\sigma_i^2)},$$

to get

$$E[e^{tT}] \;\leq\; e^{O(t^2\sigma^2)}.$$

Now,

$$e^{t\lambda\sigma} P(T > \lambda\sigma) \leq E[e^{tT}],$$

yielding

$$P(T > \lambda\sigma) \leq e^{O(t^2\sigma^2 - t\lambda\sigma)}.$$

Optimize on $t \in [0,1]$ to get the required result. □