The Chernoff Bound

April 2022

Let's look at the famous Chernoff bound with an application to computational complexity theory.

Theorem. *Let* $X_1, \ldots, X_n$ *be independent Bernoulli random variables with expected value* $p_1, \ldots, p_n$*. Let* $\mu = \sum_{i=1}^{n} p_i$*. Then, for any* $0 < \delta < 1$

$$\Pr\left[\sum_{i=1}^{n} X_i \geq (1+\delta)\mu\right] \leq \left(\frac{e^{\delta}}{(1+\delta)^{1+\delta}}\right)^{\mu} \tag{1}$$

and

$$\Pr\left[\sum_{i=1}^{n} X_i \leq (1-\delta)\mu\right] \leq \left(\frac{e^{-\delta}}{(1-\delta)^{1-\delta}}\right)^{\mu}. \tag{2}$$

Proof. Let $t > 0$ be arbitrary. Note that $\sum_{i=1}^{n} X_i \geq (1+\delta)\mu$ if and only if $\exp\left(\sum_i tX_i\right) \geq \exp(t(1+\delta)\mu)$. Taking the expected value of the left-hand side, we get

$$\begin{aligned}
\mathbb{E}\left[\prod_{i=1}^{n} \exp(tX_i)\right] &= \prod_{i=1}^{n} \mathbb{E}[\exp(tX_i)] \quad \text{by independence} \\
&= \prod_{i=1}^{n} (1 - p_i + p_i e^t) \quad \text{since } X_i \text{ are Bernoulli} \\
&= \prod_{i=1}^{n} (1 + p_i(e^t - 1)) \\
&\leq \prod_{i=1}^{n} \exp(p_i(e^t - 1)) \\
&= \exp\left(\sum_{i=1}^{n} p_i(e^t - 1)\right) = \exp(\mu(e^t - 1)).
\end{aligned}$$

Apply Markov's inequality to get

$$\Pr\left[\prod_{i=1}^{n} \exp(tX_i) \geq \exp(t(1+\delta)\mu)\right] \leq \frac{\exp(\mu(e^t - 1))}{\exp(t(1+\delta)\mu)}.$$

This is true for any $t > 0$, but using calculus we can find that $t = \log(1+\delta)$ gives the minimum of the right-hand side. Thus, we get

$$\Pr\left[\sum_{i=1}^{n} X_i \geq (1+\delta)\mu\right] \leq \left(\frac{e^{\delta}}{(1+\delta)^{1+\delta}}\right)^{\mu}$$

by substituting $t = \log(1+\delta)$. Similar strategy allows us to conclude the second inequality. □

A more useful form of the theorem is given by the following corollary.

Corollary. *With the same setup as above,*

$$\Pr\left[\sum_{i=1}^{n} X_i \geq (1+\delta)\mu\right] \leq \exp\left(\frac{-\delta^2\mu}{3}\right)$$

and

$$\Pr\left[\sum_{i=1}^{n} X_i \le (1-\delta)\mu\right] \le \exp\left(\frac{-\delta^2\mu}{2}\right).$$

Proof. Write

$$\left(\frac{e^{\delta}}{(1+\delta)^{1+\delta}}\right)^{\mu} = e^{[\delta-(1+\delta)\log(1+\delta)]\mu}, \tag{3}$$

and

$$\left(\frac{e^{-\delta}}{(1-\delta)^{1-\delta}}\right)^{\mu} = e^{[-\delta-(1-\delta)\log(1-\delta)]\mu}. \tag{4}$$

For $0 \le \delta < 1$, we have the identity $\log(1-\delta) \ge \frac{-\delta+\delta^2/2}{1-\delta}$. This can be proved by observing that both sides agree at $\delta = 0$, and the derivative of the right-hand side is always smaller than that of the left. Hence,

$$(1-\delta)\log(1-\delta) \ge -\delta + \delta^2/2.$$

Substituting this into (3), we see that (2) implies

$$\Pr\left[\sum_{i=1}^{n} X_i \le (1-\delta)\mu\right] \le \exp\left(\frac{-\delta^2\mu}{2}\right).$$

The same method as above shows the identity $\log(1+\delta) \ge \frac{\delta}{1+\delta/2}$. Plugging this into (1) and (4), we get

$$\Pr\left[\sum_{i=1}^{n} X_i \ge (1+\delta)\mu\right] \le \exp\left(\frac{-\delta^2\mu}{2+\delta}\right).$$

Since $\delta < 1$, we have

$$\Pr\left[\sum_{i=1}^{n} X_i \ge (1+\delta)\mu\right] \le \exp\left(\frac{-\delta^2\mu}{3}\right).$$

□

0.1 An application to computational complexity theory

The Chernoff bound intuitively says that if we have a coin that has 0.357 probability of landing heads, then, with exponentially high probability in the number of times we toss it, the ratio of heads is in 0.357 ± 0.02. Of course, 0.357 and 0.02 are arbitrary.

Theorem (Error reduction for **BPP**, 7.10 of Arora and Barak)**.** *Let $L \subseteq \{0,1\}^*$ be a language and suppose that there exists a poly-time probabilistic TM M such that for every $x \in \{0,1\}^*$, $\Pr[M(x) = L(x)] \ge 1/2+|x|^{-c}$. Then, for every constant $d > 0$, there is a poly-time probabilistic TM M' such that for every $x \in \{0,1\}^*$, $\Pr[M(x) = L(x)] \ge 1-2^{-|x|^d}$.*

Proof. The idea is to simply call M many times and take the majority of the bits returned by M. More precisely, M' calls M for a total of k times to get bits $b_1, \dots, b_k$. Here, k is an unknown number we need to find out. Then, M' returns the majority of $b_1, \dots, b_k$. Define the random variables X_i with $X_i = 1$ if $b_i = L(x)$ and $X_i = 0$ otherwise. Note that if more than $k/2$ of X_i is 1, then M' is correct. The expected value of X_i is $p = 1/2 + |x|^{-c}$. So the expected value of $X = X_1 + \ldots + X_k$ is pk. We wish to use the Chernoff bound

$$\Pr\left[X \le (1-\delta)pk\right] \le \exp(-\delta^2\mu/2)$$

for some suitable δ. Note that for this to give us what we need, we must have $(1-\delta)p \ge 1/2$. Plugging in the value of p, we see that we can take $\delta = |x|^{-c}/2$.

Finally, we want k such that $\exp(-\delta^2\mu/2) \le 2^{-|x|^d}$. Solving, we see that $k = 16\log(2)|x|^{d+2c}$ is a solution. □

This result is crucial since it explains how the success rate of 2/3 in the definition of **BPP** can simply be replaced by any constant $> 1/2$, or even *a shrinking* $1/2 + |x|^{-c}$.

Another application is about randomized reduction to `3Sat`. It turns out the success rate of reduction can be improved to arbitrarily high as well.

Theorem. *Let* $L \subseteq \{0,1\}^*$ *be a language and suppose that there exists a poly-time probabilistic TM* M *such that for every* $x \in \{0,1\}^*$, *if* $x \in L$, *then* $\Pr[M(x) \in 3\mathit{Sat}] \geq 1/2 + |x|^{-c}$, *otherwise* $\Pr[M(x) \notin 3\mathit{Sat}] \geq 1/2 + |x|^{-c}$. *Then, for every constant* $d > 0$, *there is a poly-time probabilistic TM* M' *such that for every* $x \in \{0,1\}^*$, *if* $x \in L$, *then* $\Pr[M(x) \in 3\mathit{Sat}] \geq 1 - 2^{-|x|^d}$, *otherwise* $\Pr[M(x) \notin 3\mathit{Sat}] \geq 1 - 2^{-|x|^d}$.

Proof. The calculations in this proof are exactly the same as the last one. On input $|x|$, the machine M' runs M k times with fresh randomness each time to get boolean formulas $\phi_1(y_1), \ldots, \phi_k(y_k)$. Each ϕ_i has length polynomial in $|x|$, so do the y_i's. Consider the formula $Maj_{i=1}^{k} \phi_i(y_i)$. This is satisfiable if and only if more than half of the ϕ_i's are satisfiable. This is guaranteed to happen with exponentially high probability if x is in L, otherwise it happens with exponentially low probability. □

This theorem leads to a short proof to $\mathbf{BP} \cdot \mathbf{NP} \subseteq \mathbf{NP}/poly$, which is analogous to the proof of $\mathbf{BPP} \subseteq \mathbf{P}/poly$.