Chernoff bound

Yuval Filmus

December 7, 2017

Chernoff's bound (also known as Hoeffding's bound) is a fundamental large deviation bound which is extremely useful in theoretical computer science.

1. **Chernoff–Hoeffding bound.**[1] Let $X_1, \ldots, X_n$ be independent random variables supported on $[0,1]$ such that $\mathbb{E}[X_i] = \mu_i$. Define $X := X_1 + \cdots + X_n$ and $\mu = \mathbb{E}[X]/n$.

   (a) Let $\lambda > 0$ be a parameter to be determined. Show that $e^{\lambda t}$ is convex, and so $e^{\lambda t} \leq 1 + t(e^\lambda - 1) = (1-t) + te^\lambda$ for $t \in [0,1]$.

   (b) Deduce that $\mathbb{E}[e^{\lambda X_i}] \leq (1-\mu_i) + \mu_i e^\lambda$.

   (c) Show that $\mathbb{E}[e^{\lambda X}] = \prod_{i=1}^n \mathbb{E}[e^{\lambda X_i}]$, and deduce $\mathbb{E}[e^{\lambda X}] \leq \prod_{i=1}^n [(1-\mu_i) + \mu_i e^\lambda]$.

   (d) Use the arithmetic mean-geometric mean inequality to conclude that

   $$\mathbb{E}[e^{\lambda X}] \leq [(1-\mu) + \mu e^\lambda]^n.$$

   (e) Let $m = (\mu + t)n$. Use Markov's inequality to show that

   $$\Pr[X \geq (\mu+t)n] \leq \left(\frac{(1-\mu) + \mu e^\lambda}{e^{(\mu+t)\lambda}} \right)^n.$$

   (f) Using calculus, show that the right-hand side is minimized when

   $$e^\lambda = \frac{(1-\mu)(\mu+t)}{\mu(1-\mu-t)}.$$

   (g) Conclude that

   $$\Pr[X \geq (\mu+t)n] \leq \left[\left(\frac{\mu}{\mu+t} \right)^{\mu+t} \left(\frac{1-\mu}{1-\mu-t} \right)^{1-\mu-t} \right]^n.$$

   (h) Show that the right-hand side is $e^{-nD(\mu+t\|\mu)}$ (using the natural logarithm to define $D(\cdot\|\cdot)$), where we identify μ with a Bernoulli μ random variable.

   (i) Let $\delta(t) = D(\mu+t\|\mu)$. Show that $\delta(0) = \delta'(0) = 0$ and $\delta''(t) \geq 4$ for all $t \leq 1-\mu$. (Hint: if $x \in [0,1]$ then $x(1-x) \leq 1/4$.)

   (j) Deduce that for all $t \leq 1-\mu$, $\delta'(t) \geq 4t$ and so $\delta(t) \geq 2t^2$.

   (k) Conclude a form of the Chernoff–Hoeffding bound:

   $$\Pr[X \geq (\mu+t)n] \leq e^{-2t^2 n}.$$

[1] Our proof folows the monograph *Concentration of measure for the analysis of randomized algorithms* by Dubhashi and Panconesi.

Using more calculus, one can show that for $0 < \epsilon < 1$,

$$\Pr[|X - \mathbb{E}[X]| \geq \epsilon\,\mathbb{E}[X]] \leq 2e^{-(\epsilon^2/3)\,\mathbb{E}[X]}.$$

Another useful form of the bound is as follows. Let $X_1, \ldots, X_n$ be independent random variables, where X_i is supported on $[a_i, b_i]$. Define $X := X_1 + \cdots + X_n$. Then

$$\Pr[|X - \mathbb{E}[X]| \geq t] \leq 2\exp\left(-\frac{2t^2}{\sum_{i=1}^n (b_i - a_i)^2}\right).$$

2. **Randomness amplification.** Suppose that we are trying to discover an unknown bit b using independent samples of a random bit X which satisfies $\Pr[X = b] = 1/2 + \delta$. Let M result from taking the majority value of N random samples of X.
   (a) Suppose that $\delta = c/\sqrt{N}$. Show that $\Pr[M = b] \geq 1 - e^{-2c^2}$.
   (b) Suppose that δ is constant. Show that $\Pr[M \neq b]$ is exponentially small.

3. **Estimating an average.** Let $f\colon 2^U \to \mathbb{R}$ be a set function, that is, a function which accepts as input a subset S of U and outputs a real number. Suppose that $\mathcal{D}$ is a distribution over subsets of U, and we're interested in estimating $\mathbb{E}_{S\sim\mathcal{D}}[f(S)]$. The obvious way to estimate this quantity is to average $f(S)$ over N samples of $\mathcal{D}$, for large enough N. How many samples do we need to take in order to guarantee that with probability at least $1 - \delta$, the estimate deviates from the true value by at most $\epsilon \max_S |f(S)|$?

4. **Random sampling.** Let $p, \epsilon \in (0, 1)$ be arbitrary constants. Suppose that we sample a subset S of $\{1, \ldots, n\}$ by choosing each point with probability p independently. Let us say that S is (ϵ, δ)-*balanced* if $(p-\epsilon)|A| \leq |S \cap A| \leq (p+\epsilon)|A|$ for all $|A| \geq (1-\delta)n$. Show that there exists a constant $\delta > 0$ (depending on ϵ) such that with high probability (i.e., with probability tending to 1 as $n \to \infty$), S is (ϵ, δ)-balanced.

5. **McDiarmid's inequality.** Suppose that f is a function with n inputs such that changing the ith input can change the value of the function by at most c_i. McDiarmid's inequality (which follows from Azuma's inequality, a version of Chernoff's bound for martingales) states that if $X_1, \ldots, X_n$ are independent random variables then

$$\Pr[|f(X_1, \ldots, X_n) - \mathbb{E}[f(X_1, \ldots, X_n)]| \geq t] \leq 2\exp\left(-\frac{2t^2}{\sum_{i=1}^n c_i^2}\right).$$

   (a) Consider the experiment in which n balls are thrown uniformly at random into m bins, and denote by X the expected number of empty bins. Calculate $\mathbb{E}[X]$, and show that
   $$\Pr[|X - \mathbb{E}[X]| \geq t] \leq 2e^{-2t^2/n}.$$
   (b) Sample a random undirected graph on n vertices by adding each edge with probability $1/2$, and let X be the chromatic number of the resulting graph. Show that
   $$\Pr[|X - \mathbb{E}[X]| \geq t] \leq 2e^{-2t^2/n}.$$

Exercise:

- **Maximum degree in a random graph.** Sample a random undirected graph on n vertices by adding each edge with probability $1/2$. Show that there exists a constant $C > 0$ such that with high probability, all vertices have degree at most $n/2 + C\sqrt{n \log n}$. (Hint: Use the union bound.)