CMSC 858S: Randomized Algorithms

Fall 2001

Handout 3: Facts related to the Chernoff-Hoeffding bounds

1 The bounds

As we saw in class, the Chernoff-Hoeffding bounds help upper-bound the probability that a sum of bounded and independent random variables deviates much from its mean. Suppose $X = \sum_{i=1}^{n} X_i$, where the X_i are *independent* random variables, each taking values in the interval $[0, 1]$. (In other words, each X_i is bounded; the most common case in randomized algorithms is where each X_i takes on values in $\{0, 1\}$.) Then, for $\delta > 0$, the bounds give

$$\Pr[X \geq \mu(1+\delta)] \leq F^+(\mu, \delta) \doteq \left(\frac{e^{\delta}}{(1+\delta)^{1+\delta}}\right)^{\mu}. \tag{1}$$

For the "lower tail" with $0 < \delta \leq 1$, we get

$$\begin{aligned} \Pr[X \leq \mu(1-\delta)] &\leq \left(\frac{e^{-\delta}}{(1-\delta)^{1-\delta}}\right)^{\mu} && (2) \\ &\leq F^-(\mu, \delta) \doteq e^{-\mu\delta^2/2}. && (3) \end{aligned}$$

The following simple upper bounds for $F^+(\mu, \delta)$ are useful:

$$\text{for } \delta \leq 1, \quad F^+(\mu, \delta) \leq e^{-\delta^2\mu/3}; \qquad \text{for } \delta \geq 1, \quad F^+(\mu, \delta) \leq e^{-(1+\delta)\ln(1+\delta)\mu/5}. \tag{4}$$

It is useful to be conversant with these bounds; it is especially convenient to remember that: $F^+(\mu, \delta)$ (i) decays exponentially in $\mu\delta^2$ for "small" δ ($\delta \leq 1$), and (ii) decays exponentially in $\mu(1+\delta)\ln(1+\delta)$ for larger δ ($\delta > 1$). Also, some of the constants such as 3 and 5 in the exponents above, can be improved.

As discussed in class and in the book, we often have to solve the following inverse problem: given μ and ϵ, find a "good enough" value $\Delta^+(\mu, \epsilon)$ such that $F^+(\mu, \Delta^+(\mu, \epsilon)) \leq \epsilon$. By "good enough", we mean a value that is close to (i.e., not much smaller than) the largest real δ such that $F^+(\mu, \delta) \leq \epsilon$. Similarly, we often want a value $\Delta^-(\mu, \epsilon)$ such that $F^-(\mu, \Delta^-(\mu, \epsilon)) \leq \epsilon$.

From the definition of $F^-(\mu, \delta)$, a natural choice for $\Delta^-(\mu, \epsilon)$ is seen to be $\sqrt{2\ln(1/\epsilon)/\mu}$. In general, if ϵ is so small that $\ln(1/\epsilon) \gg \mu$, there may be no possible value for $\Delta^-(\mu, \epsilon)$.

We must do a little more work to find a good choice for $\Delta^+(\mu, \epsilon)$, since, as seen above, the behavior of $F^+(\mu, \delta)$ depends on whether δ is "small" or "large". Using (4), it is possible to show that the following is a suitable choice:

$$\begin{aligned} \Delta^+(\mu, \epsilon) &= \sqrt{3\ln(1/\epsilon)/\mu} \ \text{ if } \mu \geq 3\ln(1/\epsilon); && (5) \\ &= 10 \cdot \frac{\ln(1/\epsilon)}{\mu \cdot \ln(\ln(1/\epsilon)/\mu)} \ \text{ if } \mu < 3\ln(1/\epsilon). && (6) \end{aligned}$$

Knowing these types of bounds for Δ^+ and Δ^- is of much use in the design and analysis of randomized algorithms.

2 Special cases

As mentioned before, many constants seen above, such as the 10 in the definition of the first case for $\Delta^+(\mu, \epsilon)$, are by no means tight. Getting the "right" constants is important in some situations, typically where the parameter δ is either "very small" or (relatively) "very large". We briefly discuss these situations now.

The bound $F^-(\mu, \delta)$ on $\Pr[X \leq \mu(1-\delta)]$ is quite good when δ is "small" (close to 0). However, if $\delta \to 1$, one can often take advantage of the fact that $\left(\frac{e^{-\delta}}{(1-\delta)^{1-\delta}}\right)^{\mu} \to e^{-\mu}$ as $\delta \to 1$.

As for $F^+(\mu, \delta)$, it can be shown that

$$F^+(\mu, \delta) \leq e^{-\mu\delta^2/2 + \mu\delta^3/6}. \tag{7}$$

Thus, if δ is close to 0, the dominant term in the exponent is $-\mu\delta^2/2$. At the other extreme where δ is "large", note that

$$F^+(\mu, \delta) = e^{-\mu(1+\delta)\ln(1+\delta)\cdot\left[1 - \frac{\delta}{(1+\delta)\ln(1+\delta)}\right]}; \tag{8}$$

as δ grows large, the dominant term in the exponent is $-\mu(1+\delta)\ln(1+\delta)$. These two observations lead to improved bounds on $\Delta^+(\mu, \epsilon)$ for the cases of $\mu \gg \ln(1/\epsilon)$ and $\mu \ll \ln(1/\epsilon)$ respectively. For convenience, let M denote the term $\ln(1/\epsilon)/\mu$. Then, for some function $f_1(M)$ that tends to 0 as $M \to 0$, we can set

$$\Delta^+(\mu, \epsilon) = \sqrt{(2 + f_1(M)) \cdot M} \quad \text{if } \mu \gg \ln(1/\epsilon).$$

And, for some function $f_2(M)$ that tends to 0 as $M \to \infty$, we can set

$$\Delta^+(\mu, \epsilon) = (1 + f_2(M))M/\ln M \quad \text{if } \mu \ll \ln(1/\epsilon).$$

3 A convenient upper-tail bound when $\delta \gg 1$

When $\delta \gg 1$, the following union bound-based approach sometimes gives a cleaner-looking bound. Given reals $z_1, z_2, \ldots, z_n$ and a positive integer $k \leq n$, define

$$S_k(z_1, z_2, \ldots, z_n) = \sum_{i_1 < i_2 < \cdots < i_k} z_{i_1} z_{i_2} \cdots z_{i_k}.$$

If $z_1, z_2, \ldots, z_n$ are constrained to be non-negative reals that add up to some value y, verify that $S_k(z_1, z_2, \ldots, z_n)$ attains a maximum when all the z_i are equal—i.e., equal to y/n. (One way to do this is as follows. We proceed by induction on n; the case of $n = 1$ is trivial. So suppose $n \geq 2$. If $k = n$, we use the arithmetic mean-geometric mean inequality, which states that under the above constraints on the z_i, $z_1 z_2 \cdots z_n$ is maximized when all the z_i are equal. Next suppose $k < n$. In this case, make suitable use of the fact

$$S_k(z_1, z_2, \ldots, z_n) = S_k(z_1, z_2, \ldots, z_{n-1}) + z_n \cdot S_{k-1}(z_1, z_2, \ldots, z_{n-1}),$$

and of the induction hypothesis.) Thus we have, for any collection of non-negative z_i, that

$$\begin{aligned} S_k(z_1, z_2, \ldots, z_n) &\leq \binom{n}{k} \cdot \left(\frac{z_1 + z_2 + \cdots + z_n}{n}\right)^k \\ &\leq (n^k/k!) \cdot \left(\frac{z_1 + z_2 + \cdots + z_n}{n}\right)^k \\ &\leq \frac{(z_1 + z_2 + \cdots + z_n)^k}{k!}. \end{aligned} \tag{9}$$

Suppose X is as defined in the beginning of Section 1, with the further property that each X_i takes on values in $\{0,1\}$. Let $p_i = \Pr[X_i = 1]$; thus, $\mu = \sum_i p_i$. The following is a useful way to upper-bound $\Pr[X \geq a]$ if a is an integer that is much greater than μ. (However, the following discussion allows a to be an arbitrary positive integer.)

We have

$$\begin{aligned}
\Pr[X \geq a] &= \Pr[\exists i_1 < i_2 < \cdots < i_a : \ X_{i_1} = X_{i_2} = \cdots = X_{i_a} = 1] \\
&\leq \sum_{i_1 < i_2 < \cdots < i_a} \Pr[X_{i_1} = X_{i_2} = \cdots = X_{i_a} = 1] \ \ \text{(union bound)} \\
&= \sum_{i_1 < i_2 < \cdots < i_a} p_{i_1} p_{i_2} \cdots p_{i_a} \\
&= S_a(p_1, p_2, \ldots, p_n) \\
&\leq \mu^a / a! \ \ \text{(by (9))}.
\end{aligned}$$

Thus we get the inequality that we used in class:

$$\Pr[X \geq a] \leq \mu^a / a!.$$