Disclaimer: *These notes have not been subjected to the usual scrutiny for formal publications. They are to be used only for the class.*

Outline:

1. The additive vs. multiplicative Chernoff bounds
2. A few other forms of Chernoff/Hoeffding bounds

1 The additive vs. multiplicative Chernoff bounds

Note that for homework 9 problem 2, you are asked to bound a random variable using *additive* Chernoff bounds. For problem 4, instead you are asked to mimic the proof of the *multiplicative* Chernoff bound. You may wonder. What's the difference?

In the lecture, we covered the *additive* Chernoff bound. However, in section 4.1 of the Motwani and Raghavan book, the Chernoff bound appeared in its *multiplicative* form. Put it simply: the additive Chernoff bound has the form of

$$\Pr[X > \lambda] < \ldots \text{ or } \Pr[X > \lambda + \mu] < \ldots$$

whereas the multiplicative Chernoff bound has the form

$$\Pr[X > (1+\delta)\mu] < \ldots$$

Here μ is the expected value (the mean) of the random variable X. The value λ is the absolute error from the mean, whereas δ denotes the *relative* error from the mean.

The proofs for both bounds look very similar. In fact, they both have the following three ingredients:

1. Study the random variable e^{tX} instead of X, where t is a dummy variable that disappears at the end (recall that t is being minimized over at the end).
2. The expectation of e^{tX} can be factored into the expectation of the independent random variables e^{tX_i}.
3. Relax the product of expected values and optimize over t.

These three steps are the main message to take away from the proofs (the same thing is outlined in section 4.1).

However, if you follow the two proofs more closely, esp. about where the relaxations (the "$\leq$" or the "$<$") occur. You will see that they differ in step 3, that is, the ways they relax the product of the expected values are different. For the proof in the lecture (namely, the proof in the book "Probabilistic Method" by Alon and Spencer), each term in the product was relaxed via the following inequality:

$$E[e^{\lambda X_i}] = (e^{\lambda} + e^{-\lambda})/2 = \cosh(\lambda) \leq e^{\lambda^2}/2$$

Contrast this with the inequality used in Motwani and Raghavan:

$$1 + x < e^x \qquad \text{with } x = p_i(e^t - 1)$$

The two different relaxations give rise to the two different forms of the bound. For problem 2, you should invoke the "Alon-Spencer" form of the bound. For problem 4, you should try to mimic the proof of the "Motwani-Raghavan" form of the bound.

2 A few other forms of Chernoff/Hoeffding bounds

Here are yet again two well-known forms of the bound. The first one appeared as the original form of the bound. The second one is quite handy in practice.

Theorem 1 ([C52], [H63]) *Let* $X_1, X_2, \ldots, X_n$ *be* independent *0-1 random variables with* $\Pr[X_i = 1] = p_i$. *Denote* $X = \sum_{i=1}^{n} X_i$, $\mu = E[X]$ *and* $p = \mu/n$. *Then, for* $0 \leq \lambda < n - \mu$,

$$\Pr[X \geq \mu + \lambda] \leq \exp(nH_p(p + \frac{\lambda}{n}))$$
$$\Pr[X \leq \mu - \lambda] \leq \exp(nH_{1-p}(1 - p + \frac{\lambda}{n}))$$

where $H_p(x) \equiv x \ln(\frac{p}{x}) + (1-x)\ln(\frac{1-p}{1-x})$ *is the* relative entropy *of* x *with respect to* p.

The proof is similar to the previous proofs. It also has the three ingredients but distinguishes itself in the way of relaxing the product of expected values, here Jensen's inequality is invoked in place of the inequality used in the other two proofs.

Theorem 2 ([AV79]) *Assume the same set up as above, then*

$$\Pr[X \leq (1-\beta)\mu] \leq \exp(\frac{-\beta^2\mu}{2}) \qquad \text{for all } 0 < \beta \leq 1$$
$$\Pr[X \geq (1+\beta)\mu] \leq \begin{cases} \exp(\frac{-\beta^2\mu}{2+\beta}) & \text{for all } \beta > 1 \\ \exp(\frac{-\beta^2\mu}{3}) & \text{for all } 0 < \beta \leq 1 \end{cases}$$

It is a corollary of the original form of the bound, therefore is weaker. But this form of the bound is easier to memorize and more handy to use than the previous original theorem.

References

[C52] H. CHERNOFF, “Asymptotic efficiency for tests based on the sum of observations”, Ann. Math. Stat., Vol 23, 1952, pp. 493-507

[H63] W. HOEFFDING, “Probability for sums of bounded random variables”, J. of the American Statistical Association, No. 58, 1963, pp. 13-30

[AV79] D. ANGLUIN and L.G. VALIANT,”Fast probabilistic algorithms for Hamiltonian circuits and matchings”, J. of Computer and System Sciences, No. 19, 1979, pp. 155-193