Lecture 4

1 Deviation bounds.

Deviation bounds are bounds on the probability for a random variable being very different from its mean. In the balls in bins example, the mean is m/n, and we wish to bound the deviation from this mean.

The most basic deviation bound is the following.

Theorem 1
[Markov's inequality.] For a positive random variable X,

$$Pr[X > cE[X]] \leq 1/c.$$

Proof:

$$\begin{aligned}
E[X] &= \sum_a aPr[X = a] \\
&= \sum_{a \leq cE[X]} aPr[X = a] + \sum_{a > cE[x]} aPr[X = a] \\
&> \sum_{a > cE[X]} aPr[X = a] \\
&> cE[X] \sum_{a > cE[X]} Pr[X = a] \\
&= cE[X]Pr[X > cE[X]]
\end{aligned}$$

The first line is by definition of expectation. The next several are algebra. The last equality is the definition of probability of the event "$X > cE[X]$".

We now have

$$> \geq cE[X]Pr[X > cE[X]].$$

Dividing both sides by $cE[X]$ yields the theorerm. (Where do we use the fact that X is positive?)

□

Exercise: do not turn in. Show that for a random variable Y whose expectation is μ and whose maximum value is 2μ that the probability that the random variable is less than $\mu/4$ is at most $3/7$.

For our n balls into n bins, we get that the probability that there are more than k balls in bin 1 is at most $1/k$. This does not particularly help us prove an interesting upper bound in the *max* load, since we use the union bound on all the n bins.

A stronger deviation bound is called Chebyshev's inequality.

Theorem 2
[Chebyshev's inequality.] For any random variable X, and $\mu = E[X]$, and $\sigma^2 = E[(X-\mu)^2]$,

$$Pr[|X - E[X]| > t\sigma] \le 1/t^2.$$

Proof:
This inequality follows from Markov's inequality. Consider the positive random variable $Y = (X - E[X])^2$. Note that $E[Y] = \sigma^2$. We have

$$Pr[Y > t^2\sigma^2] \le 1/t^2,$$

by Markov's inequality.
But, $Pr[|X - E[X]| > t\sigma] = Pr[Y > t^2\sigma^2]$. Thus, the theorem holds.
□

The quantity $E[(X-E[X])^2]$ is called the variance of X, sometimes denoted by $Var[X]$. The square root of the variance is the standard deviation of X, usually denoted by the symbol σ. (By the way, the symbol μ often denotes $E[X]$.) It is easy to verify the equality about the variance is

$$Var[X] = E[(X - E[X])^2] = E[X^2] - E[X]^2. \tag{1}$$

To use Chebyshev, we need to compute the varaiance of our random variable. For our balls in bins experiments, the load on bin 1 is $X = \sum_i X_i$ where X_i is $0-1$ indicator variable for ball i choosing bin 1. Recall that $E[X] = 1$. Now,

$$X^2 = \sum_{i,j} X_i X_j = \sum_i X_i^2 + \sum_{i \ne j} X_i X_j$$

.

Noting that $X_i^2 = X_i$, and taking expectations, we get

$$E[X^2] = \sum_i E[X_i] + \sum_{i \ne j} E[X_i X_j].$$

Now, $\sum_i E[X_i] = E[X]$, and $E[X_i X_j] = Pr[X_i = 1 \text{ and } X_j = 1] = (1/n)^2$ since the events are independent. So, we have

$$E[X^2] = E[X] + n(n-1)/n^2,$$

and plugging into equation 1, we get

$$Var[X] = (n-1)/n.$$

This is upper bounded by 1. We can now use Chebyshev to bound the probability that the load is greater than $2\sqrt{n}$ by $1/4n$, which then allows us to say that the probability that any bin exceeds load $2\sqrt{n}$ is at most $1/4$.

This still is not so good as we showed $O(\log n/\log\log n)$ previously.

2 Chernoff Bounds

Now, we examine a bound that gets reasonably tight answers. The following theorem is ascribed to Chernoff and/or Hoeffding. There are numerous forms, of which this is one.

LEMMA 3
For a random variable, $X = \sum_i X_i$*, where* X_i *are* $0 - 1$ *random variables, and with mean* $\mu = E[X]$*, for* $\delta > 0$

$$Pr[X > (1+\delta)\mu] \leq \left(\frac{e^{\delta}}{(1+\delta)^{1+\delta}}\right)^{\mu}. \tag{2}$$

and for $1 > \delta > 0$*,*

$$Pr[X < (1-\delta)\mu] \leq \left(\frac{e^{-\delta}}{(1-\delta)^{1-\delta}}\right)^{\mu}. \tag{3}$$

Note the setup here only applies to random variables that come from the sum of random variables. It uses this crucially to arrive at this bound. This can be compared to the use of Chebyshev which requires the extra step of finding the standard deviation, where in the Chernoff bound something like this step is inherent in its proof.
PROOF:
We only give the proof for bounding the probability of the "upper tail" or equation 2. Let t be an arbitrary positive constant. Below, $p_i = Pr[X_i = 1] = E[X_i]$.

$$\begin{aligned} Pr[X > (1+\delta)\mu] &\leq Pr[e^{tX} > e^{(1+\delta)t\mu}] \\ &\leq \frac{E[e^{tX}]}{e^{(1+\delta)t\mu}} \end{aligned}$$

The last line follows from Markov's inequality. The first step is odd, but somewhat standard since $f(t) = E[e^{tX}]$ is the "moment generating function." That is, the k-moment, $E[X^k]$, is the k-derivative of f evaluated at 0, $f^k(0)$.
We proceed by finding an upper bound for $E[e^{tX}] = E[e^{t\sum_i X_i}]$.

$$\begin{aligned} E[e^{t\sum_i X_i}] = \Pi_i E[e^{tX_i}] &\leq \Pi_i(p_i e^t + (1-p_i)e^0) \\ &\leq \Pi_i(1 + p_i(e^t - 1)) \\ &\leq \Pi_i e^{p_i(e^t-1)} \\ &\leq e^{\sum_i p_i(e^t-1)} \\ &\leq e^{(e^t-1)\sum_i p_i} = e^{(e^t-1)\mu} \end{aligned}$$

The first line follows from the fact that the X_i are independent. (That is, $E[XY] = E[X]E[Y]$ for independent random variables X and Y.) The third line follows from the fact that for positive x, $(1+x) < e^x$.
Plugging back in we get.

$$Pr[X > (1+\delta)\mu] \le \left(\frac{e^{(e^t-1)}}{e^{(1+\delta)t}}\right)^{\mu}$$

Choosing $t = \ln(1+\delta)$, we get equation 2.
□

The Lemma above gives us an upper bound on the probability of the *upper and lower tails* of the distribution of X. Note, that the case of equation 2 can be expressed in a simpler form, which is more convenient for our needs:

$$\Pr\big[X \ge (1+\delta)\mu\big] \le \begin{cases} e^{-\delta^2\mu/3} & \text{if } \delta \le 1, \\ e^{-\delta^2\mu/4} & \text{if } \delta \le 2e-1, \\ 2^{-\delta\mu} & \text{if } \delta > 2e-1. \end{cases} \tag{4}$$

Question 1: Prove one of these inequalities using the lemma above (You may be within a constant factor on the exponent if you like or on the range of δ. For example, $e^{-\delta^2\mu/6}$ versus $e^{\delta^2\mu/3}$.)

We can use equation 4 to show an upper bound on the maximum load with m balls in n bins, with high probability. As we would like to bound the probability by

$$\Pr\Big[X \ge (1+\delta)\left(\frac{m}{n}\right)\Big] \le \frac{1}{n^2},$$

we need to determine the appropriate δ values for which this may hold. For the simple case where $m = n$, assuming $\delta \ge 2e$ we get

$$\Pr\big[X \ge (1+\delta)\big] \le 2^{-\delta}$$

hence, for $\delta \ge 2\log_2 n$, it holds that

$$\Pr\big[X \ge (1+\delta)\big] \le 2^{-2\log_2 n} = \frac{1}{n^2}$$

Obviously, a deviation of $2\log n$ is not too satisfying... However, we can achieve a better bound with high probability, for the general case and with bigger values of m: opting for the first case of equation 4, we require that

$$e^{-\delta^2\mu/3} \le \frac{1}{n^2}$$

therefore

$$\begin{aligned} -\frac{\delta^2\mu}{3} &\le -2\ln n \\ \delta^2 &\ge \frac{6\ln n}{\mu} = \frac{6n\ln n}{m} \\ \delta &\ge \sqrt{\frac{6n\ln n}{m}} \quad . \end{aligned}$$

Since it is assumed that $\delta \leq 1$, we get that m must be greater than $6n \ln n$. Then, denoting $c = \frac{m}{6n \ln n} \geq 1$, the maximum load is

$$\left(1 + \frac{1}{\sqrt{c}}\right)\left(\frac{m}{n}\right)$$

with high probability.

Question 2: Say a poll gives support of 50% for a candidate and used 1000 uniformly random samples (with repetition.) Use the Chernoff bound above to upper bound the probability that the poll turns out this way if her support is actually 40% or less. Remember that the expectation of a random variable is what the pollster is estimating; that is, it is presumed that the support is not random it is a fixed fraction of the population. (Note: you may use, if you justify it, the fact that the probability of the poll returning an estimate a value of v or greater when the underlying support is s is strictly less than when it is s' if $s < s'$.)

This is an example of an "inverse problem", we wish to derive the underlying parameter, not assume a parameter and argue about behavior. Thus, we have this complicated way of saying it; that is, the probability we get result y if the original parameter is less than x is at most z.