Probabilistic Inequalities: Review

This note reviews probabilistic inequalities that are frequently used in the analysis of randomized algorithms. Our focus will mainly be on Markov's inequality, Chebyshev's inequality, and Chernoff/Hoeffding bounds. We will also have a few examples to go along. Some of this material is taken from the two handouts posted on the course webpage [Ron06, LR06].

1 The Basics

Theorem 1 (Union bound). *Given n events $E_1, E_2, \ldots, E_n$ over the same sample space,*

$$\Pr[E_1 \cup E_2 \cup \cdots \cup E_n] \leq \Pr[E_1] + \Pr[E_2] + \cdots + \Pr[E_n].$$

Proof Idea. Prove using Venn diagrams for $n = 2$. Use induction for more events.

Theorem 2 (Linearity of Expectation). *For all random variables R_1 and R_2 and numbers $\alpha, \beta \in \mathbb{R}$, the expectation*

$$\mathbb{E}[\alpha R_1 + \beta R_2] = \alpha \cdot \mathbb{E}[R_1] + \beta \cdot \mathbb{E}[R_2].$$

2 Tail Bounds

It is usually easy to compute the expectation of the quantities of interest, while analyzing algorithms. But what we need in most cases are statements which say that, with high probability the actual value of the quantity of interest does not deviate very far from expectation. One can use tail bounds for such situations.

2.1 Markov's Inequality

Markov's inequality is one of the most basic tail bounds that applies to a wide range of random variables. To apply Markov's inequality, we require just the expectation of the random variable and the fact that it is non-negative.

Theorem 3 (Markov's Inequality). *If R is a non-negative random variable, then for all $x > 0$,*

$$Pr[R \geq x] \leq \frac{\mathbb{E}[R]}{x}.$$

Proof. This is a proof that is more general than what we saw in the class.

$$\begin{aligned}
\mathbb{E}[R] &= \mathbb{E}[R|R \geq x]\Pr[R \geq x] + \mathbb{E}[R|R \leq x]\Pr[R \leq x] && (\text{rewriting } \mathbb{E}[R]) \\
&\geq \mathbb{E}[R|R \geq x]\Pr[R \geq x] && (a + b \geq a \text{ when } a, b \geq 0) \\
&\geq x \cdot \Pr[R \geq x]
\end{aligned}$$

Rearranging the terms, $Pr[R \geq x] \leq \frac{\mathbb{E}[R]}{x}$. Note that this bound is interesting only for $x > \mathbb{E}[R]$. □

Corollary 4 (Alternative formulation of Markov's inequality). *If R is a non-negative random variable, then for all $c > 0$,*

$$\Pr[R \geq c \cdot \mathbb{E}[R]] \leq \frac{1}{c}.$$

Proof. Substitute $x = c \cdot \mathbb{E}[R]$ into the Markov's Inequality. Then

$$\Pr[R \geq c \cdot \mathbb{E}[R]] \leq \frac{\mathbb{E}[R]}{c \cdot \mathbb{E}[R]} = \frac{1}{c}. \qquad \square$$

Exercise 1.1. *If we pull some random person aside, what is the probability their weight R is at least 200 given that the average weight (expected value of R) is 100? What is the probability they weigh at least 300?*

Solution. Using Corollary 2, $\Pr[R \geq 200] \leq \frac{1}{2}$ and $\Pr[R \geq 300] \leq \frac{1}{3}$. $\square$

Exercise 1.2. Chinese appetizer problem: *There are n people eating appetizers in a Chinese restaurant, and there is a spinner on the table holding n appetizers, one for each person. Someone antisocial decides to spin it. What is the expected number of people who get their dishes back in front of them? What is the probability that all the people get their dish back?*

Solution. Let C denote the r.v. of the number of people who get their dish back. Using the definition of expectation and the fact that either everyone gets their dish (with probability $\frac{1}{n}$) or no one gets their dish (with probability $\frac{n-1}{n}$),

$$\mathbb{E}[C] = \Pr[\text{everyone gets their dish}] \cdot n + \Pr[\text{no one gets their dish}] \cdot 0 = \frac{1}{n} \cdot n = 1.$$

By Markov's inequality, $\Pr[C \geq n] \leq \frac{1}{n}$ and we know that $\Pr[C = n] = \frac{1}{n}$, so the bound is tight in this case. $\square$

Exercise 1.3. Hat check problem: *n people checked in their hats, and when they went to take their hats back the hatter went mad and started giving random people random hats. What is the expected number of people who get their hat back? What is the probability that everyone gets their hat back?*

Solution. Let C_i denote the random variable that is 1 if person i gets her hat back and 0 if she does not. So, the count of people who get their hats back is $C = \sum_{i=1}^{n} C_i$, and by linearity of expectation,

$$\mathbb{E}[C] = \mathbb{E}\left[\sum_{i=1}^{n} C_i\right] = \sum_{i=1}^{n} \mathbb{E}[C_i].$$

Since $\mathbb{E}[C_i] = \Pr[\text{wrong hat}] \cdot 0 + \Pr[\text{right hat}] \cdot 1$, and every person is equally likely to get any hat back, $\Pr[\text{right hat}] = \frac{1}{n}$ and $\mathbb{E}[C_i] = \frac{1}{n}$. Hence, $E[C] = 1$. So, by Markov's Inequality, $\Pr[C \geq n] \leq \frac{1}{n}$, but we know that $\Pr[C = n] = \frac{1}{n!}$, so the bound is extremely loose in this case. $\square$

The above examples illustrate the fact that the bound from Markov's Inequality can be either extremely loose or extremely tight, and without further information about a variable we cannot tell how tight the bound is.

Exercise 1.4. Why R must be positive: *Let's say R takes value 1 with probability .5, and value -1 with probability .5. What is the probability that R is at least 0?*

Solution. $\mathbb{E}[R]$ is clearly 0, and with Markov's inequality we would get that $\Pr[R \geq 1] \leq 0$, but we know that's not true. Markov's inequality does not apply in this case. $\square$

If we know that R has a lower bound, we can try to shift the variable instead. So, in the previous case, we could look at $R' = R + 1$. The following corollary uses this property.

Corollary 5 (Reverse Markov's inequality)**.** *If $R \leq u$ for some $u \in \mathbb{R}$, then for all $x < u$,*

$$\Pr[R \leq x] \leq \frac{u - \mathbb{E}[R]}{u - x}.$$

Proof. This can be proved by applying Markov's inequality to $R' = u - R$. Note that R' is a non-negative random variable, since $R \leq u$. $\square$

Exercise 1.5 (Quiz scores)**.** *Take R to be the score of a random student, and suppose all scores $\leq$ 100 and $\mathbb{E}[score] = 75$. What is the probability that a student gets a score of at most 50?*

Solution. Using the inequality from Corollary 4, $\Pr[R \leq 50] \leq \frac{100-75}{100-50} = \frac{1}{2}$.

2.2 Chebyshev's Inequality

Recall that the variance of a random variable R is

$$Var[R] = \mathbb{E}[(R - \mathbb{E}[R])^2] = \mathbb{E}[R^2] - \mathbb{E}[R]^2,$$

and its standard deviation is $\sigma[R] = \sqrt{Var[R]}$.

Theorem 6 (Chebyshev's Inequality)**.** *For all $x > 0$ and all random variables R,*

$$\Pr[|R - \mathbb{E}[R]| \geq x] \leq \frac{Var[R]}{x^2}.$$

Proof. It is easy to see that

$$\Pr[|R - \mathbb{E}[R]| \geq x] \leq \frac{Var[R]}{x^2} = \Pr[(R - \mathbb{E}[R])^2 \geq x] \leq \frac{Var[R]}{x^2}.$$

Applying Markov's inequality to $R' = (R - \mathbb{E}[R])^2$ is then enough to prove the theorem. $\square$

Corollary 7. $\Pr[|R - E[R]| \geq c \cdot \sigma[R]] \leq \frac{1}{c^2}$.

Proof. We can prove the corollary by substituting $x = c \cdot \sigma[R]$ in Chebyshev's inequality. $\square$

Exercise 1.6 (IQ)**.** *Take R to be the IQ of a random person you pull off the street. What is the probability that someone's IQ is at least 200 given that the average IQ is 100 and the standard deviation of IQ is 10?*

Solution. Let R denote the IQ of a random individual. We know that the IQ is at least 0. Applying Markov's inequality, we get that $\Pr[R \geq 200] \leq \frac{100}{200} = \frac{1}{2}$.

Applying Chebyshev's inequality, we get that

$$\begin{aligned}
\Pr[R \geq 200] &= \Pr[R - 100 \geq 100] \\
&\leq \Pr[|R - 100| \geq 100 \\
&= 10 \cdot \sigma[R]] \\
&\leq \frac{1}{10^2}
\end{aligned}$$

$\square$

Remark: Additionally, there exists a Chebyshev bound for one-sided analysis. But we will mainly use the two-sided version in this course.

2.3 Chernoff bound and Hoeffding bound

We will now see some of the most useful and widely used inequalities for getting tight bounds on the tails of sums of independent random variables. Depending on the way in which we define a tail, they are called Chernoff or Hoeffding bounds. These inequalities in their current form are taken from [Ron06].

Theorem 8 (Concentration Inequalities)**.** *Let $R_1, R_2, ...R_n$ be independent random variables such that $R_i \in [0,1]$ for all $i \in [n]$. Define $R = \frac{1}{m} \cdot \sum_{i=1}^{m} R_i$ and let $p = \mathbb{E}[R]$. Let $\gamma \in (0,1]$. Then the following concentration bounds hold for R.*

1. *Additive form (Hoeffding bounds)*
   - $\Pr[R \geq p + \gamma] \leq e^{-2\gamma^2 m}$.
   - $\Pr[R \geq p - \gamma] \leq e^{-2\gamma^2 m}$.
2. *Multiplicative form (Chernoff bounds)*
   - $\Pr[R \geq (1+\gamma) \cdot p] \leq e^{-\frac{\gamma^2 pm}{3}}$.
   - $\Pr[R \geq (1-\gamma) \cdot p] \leq e^{-\frac{\gamma^2 pm}{2}}$.

Exercise 1.7. *If 10 million people pick a 4-digit number (0000 till 9999) at random and only one of those numbers is a win, what is the probability that at least 1100 people win?*

Solution. Let R_i be the indicator random variable for person i guessing correctly. Therefore, $R_i = 1$ if the person i guesses correctly and is 0 otherwise. Define R as in the statement of Chernoff bound. It is easy to see that $\mathbb{E}[R_i] = 10^{-4}$ and $\mathbb{E}[R] = 10^{-4}$, where the latter is applying linearity of expectation to the definition of R. If we apply Markov's inequality to bound the required probability, we get the following.

$$\Pr[R > 1100/10^7] \leq 10/11$$

If we apply Chernoff bounds, we get the following.

$$\begin{aligned}
\Pr(R > 1100/10^7) &= \Pr[R > 1.1 \times 10^{-4}] \\
&= \Pr[R > (1+0.1) \cdot 10^{-4}] \\
&\leq e^{-\frac{0.01 \cdot 10^{-4} \cdot 10^7}{3}} \\
&\leq e^{-10/3} \\
&\leq \frac{1}{8}.
\end{aligned}$$

References

[LR06] Tom Leighton and Ronitt Rubinfeld. Large deviations. `http://www.cs.tau.ac.il/~ronitt/COURSES/F08/lec25.pdf`, 2006.

[Ron06] Dana Ron. Some useful probabilitic facts. `http://www.eng.tau.ac.il/~danar/TOP06/prob.pdf`, 2006.