Part VI

Tail Bounds and Applications

In this part of the book we delve deeply into understanding the tail of a random variable, namely the probability that the random variable exceeds some value. While we briefly touched on this topic in Section 5.9, in Chapter 18 we derive much more sophisticated tail bounds, including Chernoff bounds and Hoeffding bounds.

Tail bounds are important in providing guarantees on the probability of some bad outcome. In Chapters 19 and 20, we study some common applications of tail bounds.

First, in Chapter 19, we look at how tail bounds allow us to create confidence intervals on a statistical estimate. We also study a popular problem in theoretical computer science, called the balls-and-bins problem, where balls are distributed independently at random among bins, and we prove bounds on the bin occupancy.

Next, in Chapter 20, we turn to the problem of designing and evaluating hashing algorithms. Here we show how our tail bounds and the balls-and-bins analyses from Chapter 19 give us bounds on the number of items in a hash bucket and the probability of a hash collision.

Tail bounds are extremely important in the analysis of many randomized algorithms. Randomized algorithms are covered in depth in Part VII of the book.

18 Tail Bounds

Until now, we have typically talked about the mean, variance, or higher moments of a random variable (r.v.). In this chapter, we will be concerned with the **tail probability** of a r.v. X, specifically,

$$\mathbf{P}\{X \geq x\} \qquad \text{or} \qquad \mathbf{P}\{X > x\}.$$

The tail behavior is very important for offering quality of service (QoS) guarantees. For example, we might have to pay a penalty if the response time exceeds 1 second, and thus we want to know the fraction of jobs whose response time exceeds 1 second. Equivalently, we might want to be able to formulate a service level objective (SLO), like "99% of jobs should experience response time less than 1 second." There are many other examples of tail behavior in computer science. For example, router buffers in a network need to be provisioned so that the probability of overflow is low. Likewise, when designing a hash table, we care not only about keeping the expected number of items in a bucket low, but also about ensuring that no bucket has a *huge* number of items. All these examples require deriving tail behavior.

While the variance of a r.v. tells us something about its deviation from its mean, the tail of the r.v. gives us a lot more information. Unfortunately, it is often not easy to reason about the tail behavior of even very simple random variables. Consider, for example, $X \sim \text{Binomial}(n, p)$:

$$\mathbf{P}\{X \geq k\} = \sum_{i=k}^{n} \binom{n}{i} p^i (1-p)^{n-i}. \tag{18.1}$$

We do not have a closed-form representation of the tail probability in (18.1). Specifically, we don't have a sense of what this tail probability looks like as a simple function of k, n, and p. The tail probability in (18.1) comes up in many applications. Suppose, for example, that you are distributing n jobs among n machines by assigning each job to a random machine. In expectation each machine should get one job. You would like to know the probability that a particular machine gets $\geq k$ jobs. This probability is represented by (18.1) in the case where $p = \frac{1}{n}$.

As another example, consider $X \sim \text{Poisson}(\lambda)$. Here, X is representative of the number of arrivals to a website during 1 hour, where arrivals come from many different sources at an average total rate of λ arrivals per hour (see Chapter 12). To understand the probability that there are $\geq k$ arrivals during the hour, we need:

$$\mathbf{P}\{X \geq k\} = \sum_{i=k}^{\infty} e^{-\lambda} \frac{\lambda^i}{i!}. \tag{18.2}$$

Again, we do not have a closed-form expression for the tail probability in (18.2).

The purpose of this chapter is to investigate *upper bounds* on these tail probabilities. These upper bounds are generally called **tail bounds**. Sometimes the goal is to upper bound a tail probability of the form $\mathbf{P}\{X \geq k\}$. Other times, our goal is to upper bound the tail of the distance of a r.v. from its mean, i.e., we're trying to upper bound:

$$\mathbf{P}\{|X - \mu| \geq k\}, \qquad \text{where } \mu \equiv \mathbf{E}[X].$$

In this latter case, our tail bound is more specifically referred to as a **concentration bound** or **concentration inequality**, because we're looking at the concentration of X around its mean.

We will start by reviewing the Markov bound and the Chebyshev bound before moving on to the much more powerful Chernoff bound.

Note: This chapter and the next few will require knowing asymptotic notation well. Before you continue, you should review Section 1.6. You will need to understand the definitions of $O(n)$, $o(n)$, $o(1)$, $\Omega(n)$, $\omega(n)$ and their significance for high n.

18.1 Markov's Inequality

Theorem 18.1 (Markov's inequality) *Let X be a non-negative r.v., with finite mean $\mu = \mathbf{E}[X]$. Then, $\forall a > 0$,*

$$\mathbf{P}\{X \geq a\} \leq \frac{\mu}{a}.$$

Proof: This was proved earlier as Theorem 5.16. ■

Markov's bound is extremely weak.

Question: Suppose we flip a fair coin n times. Using Markov's inequality, what is an upper bound on the probability of getting at least $\frac{3}{4}n$ heads?

Answer: Let X denote the number of heads. Then $X \sim \text{Binomial}\left(n, \frac{1}{2}\right)$.

$$\mathbf{P}\left\{X \geq \frac{3n}{4}\right\} \leq \frac{\mu}{\frac{3n}{4}} = \frac{\frac{n}{2}}{\frac{3n}{4}} = \frac{2}{3}. \tag{18.3}$$

This is clearly a terrible bound because it doesn't even involve n.

Question: Intuitively, as n gets higher, would you expect that the tail probability should get higher or lower?

Answer: Lower. As n gets higher, we would expect that we're unlikely to be so far from the mean.

The reason why Markov's inequality is so poor is that it only takes into account the mean of the r.v. Nevertheless, this is an important inequality because we will derive all our other inequalities from this one.

18.2 Chebyshev's Inequality

Chebyshev's inequality is a lot stronger than Markov's inequality because it takes into account the variability of the r.v. Chebyshev's inequality is derived by applying Markov's inequality to the deviation of a r.v. from its mean.

Theorem 18.2 (Chebyshev's inequality) *Let X be a r.v. with finite mean $\mu = \mathbf{E}[X]$ and finite variance $\mathbf{Var}(X)$. Then, $\forall a > 0$,*

$$\mathbf{P}\{|X - \mu| \geq a\} \leq \frac{\mathbf{Var}(X)}{a^2}.$$

Proof: This was proved earlier as Theorem 5.17. ■

Using the notation σ_X to denote the standard deviation of X, where $\sigma_X^2 = \mathbf{Var}(X)$, and using $C_X^2 = \frac{\mathbf{Var}(X)}{\mathbf{E}[X]^2}$ to denote the squared coefficient of variation of X, we obtain a few additional interpretations of Chebyshev's inequality:

$$\mathbf{P}\{|X - \mu| \geq a\sigma_X\} \leq \frac{1}{a^2} \tag{18.4}$$

$$\mathbf{P}\{|X - \mu| \geq a\mathbf{E}[X]\} \leq \frac{C_X^2}{a^2}. \tag{18.5}$$

Now let's go back to the coin flipping example.

Question: Suppose we flip a fair coin n times. Using Chebyshev's inequality, what is an upper bound on the probability of getting at least $\frac{3}{4}n$ heads?

Answer: Again letting X denote the number of heads:

$$\begin{aligned}
\mathbf{P}\left\{X \geq \frac{3n}{4}\right\} &= \mathbf{P}\left\{X - \frac{n}{2} \geq \frac{n}{4}\right\} \\
&= \frac{1}{2} \cdot \mathbf{P}\left\{\left|X - \frac{n}{2}\right| \geq \frac{n}{4}\right\} \\
&\leq \frac{1}{2} \cdot \frac{\mathbf{Var}(X)}{\left(\frac{n}{4}\right)^2} \\
&= \frac{1}{2} \cdot \frac{\frac{n}{4}}{\left(\frac{n}{4}\right)^2} \\
&= \frac{2}{n}.
\end{aligned} \tag{18.6}$$

Question: Where did the $\frac{1}{2}$ in the second line come from?

Answer: Since $X \sim \text{Binomial}\left(n, \frac{1}{2}\right)$, X is symmetric around $\frac{n}{2}$.

Assuming that $n > 3$, the $\frac{2}{n}$ bound in (18.6) is much tighter than the $\frac{2}{3}$ bound that we got from Markov's inequality. Furthermore, $\frac{2}{n}$ at least decreases with n.

18.3 Chernoff Bound

We derived the Chebyshev bound by squaring the r.v. $X - \mu$ and then applying Markov's inequality. To derive the Chernoff bound, we will first exponentiate the r.v. X and then apply Markov's inequality.

For any $t > 0$,

$$\begin{aligned}
\mathbf{P}\{X \geq a\} &= \mathbf{P}\{tX \geq ta\} \\
&= \mathbf{P}\left\{e^{tX} \geq e^{ta}\right\} \\
&\leq \frac{\mathbf{E}\left[e^{tX}\right]}{e^{ta}}.
\end{aligned} \tag{18.7}$$

Question: Why were we allowed to apply Markov's inequality?

Answer: For any X and any t, we know that e^{tX} is a non-negative r.v.

Since (18.7) is true for all t, it follows that:

$$\mathbf{P}\{X \geq a\} \leq \min_{t>0} \frac{\mathbf{E}\left[e^{tX}\right]}{e^{ta}}.$$

Bounds on specific distributions are obtained by choosing the appropriate value of t. Even if a minimizing t cannot be found, it is still true that any t provides a tail bound.

Theorem 18.3 (Chernoff bound) *Let X be a r.v. and a be a constant. Then*

$$\mathbf{P}\{X \geq a\} \leq \min_{t>0} \left\{ \frac{\mathbf{E}\left[e^{tX}\right]}{e^{ta}} \right\}. \tag{18.8}$$

Question: Why should we expect that the Chernoff bound is stronger than the Chebyshev bound?

Hint: The Chebyshev bound got its strength by invoking the second moment of the r.v. What moments of the r.v. does the Chernoff bound invoke?

Answer: Notice the $\mathbf{E}\left[e^{tX}\right]$ in the Chernoff bound expression. This is a type of moment-generating function. It looks very similar to the Laplace transform, $\mathbf{E}\left[e^{-sX}\right]$. In fact, the nth derivative of $\mathbf{E}\left[e^{tX}\right]$, when evaluated at $t = 0$, yields the nth moment of X. Hence $\mathbf{E}\left[e^{tX}\right]$ encapsulates *all* moments of X.

Question: What do we do if we want to upper bound the other side of the tail, $\mathbf{P}\{X \leq a\}$?

Hint: Think about using $t < 0$.

Answer: For any $t < 0$,

$$\begin{aligned}
\mathbf{P}\{X \leq a\} &= \mathbf{P}\{tX \geq ta\} \\
&= \mathbf{P}\left\{e^{tX} \geq e^{ta}\right\} \\
&\leq \frac{\mathbf{E}\left[e^{tX}\right]}{e^{ta}} \qquad \text{(by Markov's inequality).}
\end{aligned}$$

Hence,

$$\mathbf{P}\{X \leq a\} \leq \min_{t<0} \frac{\mathbf{E}\left[e^{tX}\right]}{e^{ta}}. \tag{18.9}$$

The Chernoff bound originated in this statistics paper [14], but it is widely used in theoretical computer science. We now consider several applications of the Chernoff bound to different distributions.

18.4 Chernoff Bound for Poisson Tail

We start by illustrating how the Chernoff bound can be used to bound the tail of X where $X \sim \text{Poisson}(\lambda)$, as in (18.2).

Let $X \sim \text{Poisson}(\lambda)$. For $t > 0$,

$$\begin{aligned}
\mathbf{E}\left[e^{tX}\right] &= \sum_{i=0}^{\infty} e^{ti} \cdot \frac{e^{-\lambda} \cdot \lambda^i}{i!} \\
&= e^{-\lambda} \cdot \sum_{i=0}^{\infty} \frac{(\lambda e^t)^i}{i!} \\
&= e^{-\lambda} \cdot e^{\lambda e^t} \qquad \text{by (1.11)} \\
&= e^{\lambda(e^t - 1)}.
\end{aligned}$$

Let $a > \lambda$. Using the above, we have:

$$\begin{aligned}
\mathbf{P}\{X \geq a\} &\leq \min_{t>0} \left\{ \frac{\mathbf{E}\left[e^{tX}\right]}{e^{ta}} \right\} \\
&= \min_{t>0} \left\{ \frac{e^{\lambda(e^t-1)}}{e^{ta}} \right\} \\
&= \min_{t>0} \left\{ e^{\lambda(e^t-1)-ta} \right\}.
\end{aligned}$$

It suffices to minimize the exponent of the above expression, $\lambda(e^t - 1) - ta$, which is minimized at

$$t = \ln\left(\frac{a}{\lambda}\right),$$

which is positive, since $a > \lambda$.

This yields

$$\begin{aligned}
\mathbf{P}\{X \geq a\} &\leq e^{\lambda(e^t-1)-ta}\Big|_{t=\ln\left(\frac{a}{\lambda}\right)} \\
&= e^{\lambda\left(\frac{a}{\lambda}-1\right)-a\ln\left(\frac{a}{\lambda}\right)} \\
&= e^{a-\lambda} \cdot \left(\frac{\lambda}{a}\right)^a. \qquad (18.10)
\end{aligned}$$

Question: What is a bound on the probability that X is at least twice its mean?

Answer: From (18.10), we have $\mathbf{P}\{X \geq 2\lambda\} \leq \left(\frac{e}{4}\right)^\lambda$.

18.5 Chernoff Bound for Binomial

Chernoff bounds are most commonly applied to a sum of independent random variables, as in the case of a Binomial. In this section, we derive the Chernoff bound on the tail of X where $X \sim \text{Binomial}(n, p)$. There are many generalizations and variants of this result; see, for example, Exercises 18.15 and 18.20.

Theorem 18.4 (Pretty Chernoff bound for Binomial) *Let random variable $X \sim$ Binomial(n, p), where $\mu = \mathbf{E}[X] = np$. Then, for any $\delta > 0$,*

$$\mathbf{P}\{X - np \geq \delta\} \leq e^{-2\delta^2/n} \tag{18.11}$$

$$\mathbf{P}\{X - np \leq -\delta\} \leq e^{-2\delta^2/n}. \tag{18.12}$$

Observe that the bounds in Theorem 18.4 decrease with higher δ, as expected.

Question: One would likewise expect that the bounds in Theorem 18.4 decrease with higher n. Is this true?

Answer: This is a bit subtle:

- If δ is $\Theta(n)$, like the $\delta = \frac{n}{4}$ that we saw earlier, then the bound is of the form $e^{-\Theta(n)}$, which does in fact decrease with n, as we would expect. This is the strongest case of the bound. *This is the appropriate regime for using the pretty Chernoff bound.*
- If δ is $\Theta(\sqrt{n})$, then the bound appears to be constant in n. This makes sense because now we're looking at the probability of deviating from the mean by some number of standard deviations (again assuming p is a constant), which should become independent of n for high n and should just converge to a constant by the Central Limit Theorem (CLT).
- If δ is a constant, like 10, then the bound sadly grows with n. This is because the variance of Binomial(n, p) is $np(1-p)$, which grows with higher n (assuming that p is a constant), so the probability of exceeding a constant δ increases as n gets bigger. This is the weakest case of the bound.

We will prove Theorem 18.4 in Section 18.7. But first we consider an example of its use.

Question: Suppose we flip a fair coin n times. Using the Chernoff bound, what is an upper bound on the probability of getting at least $\frac{3}{4}n$ heads?

Answer: Again letting X denote the number of heads:

$$\begin{aligned}\mathbf{P}\left\{X \geq \frac{3n}{4}\right\} &= \mathbf{P}\left\{X - \frac{n}{2} \geq \frac{n}{4}\right\} \\ &\leq e^{-2(n/4)^2/n} \\ &= e^{-n/8}. \end{aligned} \tag{18.13}$$

The bound in (18.13) goes to zero exponentially fast in n and is much tighter than the bound of $\frac{2}{n}$ that we obtained in (18.6) via Chebyshev's inequality.

18.6 Comparing the Different Bounds and Approximations

At this point, it is useful to step back and compare the bounds that we've seen (Markov, Chebyshev, Chernoff) with both the exact answer and the *approximation* given by CLT (Theorem 9.8). We focus on our usual question.

Question: What is the exact answer for the probability of getting at least $\frac{3}{4}n$ heads with a fair coin?

Answer:

$$\sum_{i=\frac{3n}{4}}^{n} \binom{n}{i} \cdot \left(\frac{1}{2}\right)^{i} \cdot \left(1 - \frac{1}{2}\right)^{n-i} = 2^{-n} \cdot \sum_{i=\frac{3n}{4}}^{n} \binom{n}{i}. \tag{18.14}$$

The exact answer has no closed form (which is why we've been looking for bounds), but we will evaluate it soon numerically so that we can see how it compares with the bounds that we've already computed.

CLT offers an *approximate* solution for the problem. Notice that all of our coin flips are independent, with probability $p = \frac{1}{2}$. If the number of these coin flips, n, is large, then the total number of heads, X, converges to a Normal distribution by the CLT.

Question: What is the mean and standard deviation of this Normal?

Answer: $\mathbf{E}[X] = \frac{n}{2}$. Since $X \sim \text{Binomial}(n, \frac{1}{2})$, we know $\mathbf{Var}(X) = \frac{n}{4}$, so $\sigma_X = \sqrt{\frac{n}{4}}$.

We now apply the CLT approximation by first formulating our question in terms of a standard Normal:

$$\begin{aligned}\mathbf{P}\left\{X \geq \frac{3n}{4}\right\} &= \mathbf{P}\left\{X - \frac{n}{2} \geq \frac{n}{4}\right\} \\ &= \mathbf{P}\left\{\frac{X - \frac{n}{2}}{\sqrt{\frac{n}{4}}} \geq \frac{\frac{n}{4}}{\sqrt{\frac{n}{4}}}\right\} \\ &= \mathbf{P}\left\{\frac{X - \frac{n}{2}}{\sqrt{\frac{n}{4}}} \geq \sqrt{\frac{n}{4}}\right\} \\ &= \mathbf{P}\left\{\text{Normal}(0,1) \geq \sqrt{\frac{n}{4}}\right\} \\ &= 1 - \Phi\left(\sqrt{\frac{n}{4}}\right).\end{aligned}$$

Figure 18.1 compares the different approximations and bounds that we've seen, along with the exact result. As you can see, the Markov and Chebyshev bounds are both worthless for this example (we didn't even plot the Markov bound). The Chernoff bound is reasonable. The Normal approximation from the CLT is not a bound, but it's a really good approximation, particularly when n is high.

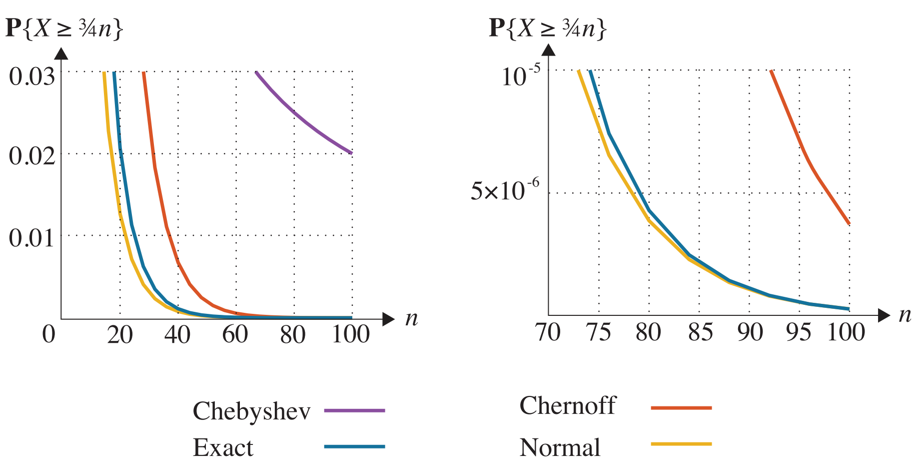

Figure 18.1 *Evaluation of* $\mathbf{P}\left\{X \geq \frac{3n}{4}\right\}$ *via Chebyshev, Chernoff, and Normal (CLT), where* $X \sim \text{Binomial}(n, 0.5)$. *Both graphs show the same comparison, but under different ranges. The first graph, with range* $1 \leq n \leq 100$, *shows that the Chebyshev bound is poor; the Chernoff bound is better; the Normal approximation from the CLT is very good. The second graph, with range* $n > 70$, *shows that, for higher* n, *Chebyshev is so bad that it doesn't even appear on the graph, and even the Chernoff bound doesn't look so great. Notice that the Normal approximation gets better and better with higher* n.

18.7 Proof of Chernoff Bound for Binomial: Theorem 18.4

The proof of Theorem 18.4 relies on Lemma 18.5. In the exercises, we will not in general have such a cute lemma to simplify our analysis, so the bounds that we will be able to prove will not always look as cute.

Lemma 18.5 *For any $t > 0$ and $0 < p < 1$ and $q = 1 - p$, we have that:*

$$pe^{tq} + qe^{-tp} \leq e^{t^2/8}.$$

Proof: The proof only uses calculus and is deferred to Section 18.10. ■

Proof: [Theorem 18.4] We will prove (18.11). The proof of (18.12) is left as an exercise. It will help to view $X = \sum_{i=1}^{n} X_i$ where $X_i \sim \text{Bernoulli}(p)$.

For any $t > 0$,

$$\begin{aligned}
\mathbf{P}\{X - np \geq \delta\} &= \mathbf{P}\{t(X - np) \geq t\delta\} \\
&= \mathbf{P}\left\{e^{t(X-np)} \geq e^{t\delta}\right\} \\
&\leq e^{-t\delta} \cdot \mathbf{E}\left[e^{t(X-np)}\right] \\
&= e^{-t\delta} \cdot \mathbf{E}\left[e^{t((X_1-p)+(X_2-p)+\cdots+(X_n-p))}\right] \\
&= e^{-t\delta} \cdot \prod_{i=1}^{n} \mathbf{E}\left[e^{t(X_i-p)}\right] \qquad \text{(because } X_i\text{'s are independent)} \\
&= e^{-t\delta} \cdot \prod_{i=1}^{n} \left(p \cdot e^{t(1-p)} + (1-p) \cdot e^{-tp}\right) \\
&\leq e^{-t\delta} \cdot \prod_{i=1}^{n} \left(e^{t^2/8}\right) \qquad \text{(by Lemma 18.5)} \\
&= e^{-t\delta + nt^2/8}.
\end{aligned} \tag{18.15}$$

We now want to find the $t > 0$ that minimizes this bound. It suffices to minimize the exponent in (18.15):

$$\begin{aligned}
\frac{d}{dt}\left(-t\delta + nt^2/8\right) &= -\delta + \frac{2nt}{8} \\
\frac{d^2}{dt^2}\left(-t\delta + nt^2/8\right) &= \frac{2n}{8} > 0.
\end{aligned}$$

Hence the minimum is obtained by finding that $t > 0$ which satisfies:

$$-\delta + \frac{2nt}{8} = 0.$$

So

$$t = \frac{4\delta}{n},$$

which is positive, as desired. Substituting this value of t into (18.15), we have:

$$\begin{aligned}
\mathbf{P}\{X - np \geq \delta\} &\leq e^{-\frac{4\delta}{n}\cdot\delta + n\left(\frac{4\delta}{n}\right)^2/8} \\
&= e^{-\frac{4\delta^2}{n} + \frac{2\delta^2}{n}} \\
&= e^{-\frac{2\delta^2}{n}}.
\end{aligned}$$

■

18.8 A (Sometimes) Stronger Chernoff Bound for Binomial

The Chernoff bound that we derived in Theorem 18.4 was very pretty. However, it's not always as strong (tight) as possible. We now introduce another bound for the Binomial. In addition to sometimes being a lot stronger, this new bound holds for a **more general definition of a Binomial**, where the coins can have different probabilities. Specifically, imagine that we are again interested in the sum of n coin flips (call this X), but this time the ith coin has probability p_i of coming up heads.

Theorem 18.6 (Sometimes stronger Chernoff bound for Binomial) *Define $X = \sum_{i=1}^{n} X_i$ where the X_i's are independent with $X_i \sim \text{Bernoulli}(p_i)$ and $\mu = \mathbf{E}[X] = \sum_{i=1}^{n} p_i$. Then, $\forall \epsilon > 0$,*

$$\mathbf{P}\{X \geq (1+\epsilon)\mu\} < \left(\frac{e^{\epsilon}}{(1+\epsilon)^{(1+\epsilon)}}\right)^{\mu}. \tag{18.16}$$

Furthermore, when $0 < \epsilon < 1$,

$$\mathbf{P}\{X \leq (1-\epsilon)\mu\} \leq \left(\frac{e^{-\epsilon}}{(1-\epsilon)^{(1-\epsilon)}}\right)^{\mu}. \tag{18.17}$$

Proof: The proof is given in Exercises 18.20 and 18.21. ■

To interpret the bound in Theorem 18.6, it helps to consider the inner expression:

$$f(\epsilon) = \frac{e^{\epsilon}}{(1+\epsilon)^{(1+\epsilon)}}. \tag{18.18}$$

Figure 18.2 shows a plot of this expression as a function of ϵ.

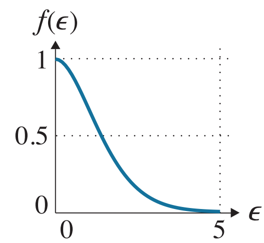

Figure 18.2 *Plot of expression (18.18). Higher ϵ leads to tighter bound.*

We make two observations: First, $\forall \epsilon > 0$, $f(\epsilon) < 1$. This implies that the bound in Theorem 18.6 is exponentially decreasing, as desired. Second, $f(\epsilon)$ decreases very quickly with higher ϵ. This too makes sense, since the Binomial should be concentrated around its mean. *The bound in Theorem 18.6 is particularly strong when ϵ is high.*

It is important to spend some time comparing the pretty bound for the Binomial in Theorem 18.4 with the (sometimes) stronger bound in Theorem 18.6. The following questions will help.

Question: Which is the better bound in the case where $p_i = p = \frac{1}{2}$, and where we are interested in the probability of at least $\frac{3n}{4}$ heads in n flips?

Answer: By Theorem 18.4, where $\delta = \frac{n}{4}$,

$$\mathbf{P}\left\{X \geq \frac{3n}{4}\right\} = \mathbf{P}\left\{X - \frac{n}{2} \geq \frac{n}{4}\right\} \leq e^{-\frac{n}{8}}.$$

By Theorem 18.6, where $\epsilon = \frac{1}{2}$,

$$\begin{aligned}\mathbf{P}\left\{X \geq \frac{3n}{4}\right\} &= \mathbf{P}\left\{X \geq \left(1 + \frac{1}{2}\right) \cdot \frac{n}{2}\right\} \\ &\leq \left(\frac{e^{.5}}{(1.5)^{1.5}}\right)^{\frac{n}{2}} \\ &\approx (0.89)^{\frac{n}{2}} \\ &\approx (1.54)^{-\frac{n}{8}}.\end{aligned}$$

Thus, Theorem 18.4 produces a *tighter* bound than Theorem 18.6 in this case, although both bounds are reasonable. Observe that it should be unsurprising that Theorem 18.6 is not so great because ϵ is only 0.5 here, which is not a good value for Theorem 18.6 (see Figure 18.2).

Question: Which is the better bound, in the case where $p_i = p = \frac{1}{n}$, and where we are interested in the probability that $X \geq 21$?

Answer: By Theorem 18.4, with $\delta = 20$, we have:

$$\mathbf{P}\{X \geq 21\} = \mathbf{P}\{X - 1 \geq 20\} \leq e^{-2\cdot(20)^2/n} = e^{-\frac{800}{n}} \to 1 \text{ as } n \to \infty.$$

The issue here is that, although δ is high, it does not increase with n, and Theorem 18.4 is only really strong when δ is $\Theta(n)$.

By contrast, by Theorem 18.6, with $\epsilon = 20$, we have:

$$\mathbf{P}\{X \geq 21\} = \mathbf{P}\{X \geq (1+20)\cdot 1\} \leq \frac{e^{20}}{21^{21}} \approx 8.3 \cdot 10^{-20}.$$

So Theorem 18.6 yields a far stronger bound for large n (although it is weaker when n is small). Note that $\epsilon = 20$ here, which is in the ideal range for Theorem 18.6, as shown in Figure 18.2.

The above shows clearly that one has to be careful in choosing a good (tight) Chernoff bound for one's application.

18.9 Other Tail Bounds

There are many other tail bounds in the literature, which either generalize the Chernoff bound, or consider a more specialized case, or a little of both. One important bound is the Hoeffding bound:

Theorem 18.7 (Hoeffding's inequality) *Let $X_1, X_2, \ldots, X_n$ be independent random variables satisfying $a_i \leq X_i \leq b_i$ for all i where $a_i \leq b_i$ are real numbers. Let*

$$X = \sum_{i=1}^{n} X_i.$$

Then,

$$\mathbf{P}\{X - \mathbf{E}[X] \geq \delta\} \leq \exp\left(-\frac{2\delta^2}{\sum_{i=1}^n (b_i - a_i)^2}\right) \tag{18.19}$$

$$\mathbf{P}\{X - \mathbf{E}[X] \leq -\delta\} \leq \exp\left(-\frac{2\delta^2}{\sum_{i=1}^n (b_i - a_i)^2}\right). \tag{18.20}$$

Proof: The proof of Hoeffding's inequality is left to Exercise 18.24. It is similar to the Chernoff bound proofs, but relies on a convexity argument. ■

Question: For Hoeffding's bound, do the X_i's need to be identically distributed?

Answer: Interestingly, the answer is no. The X_i's need to be independent, but they can each follow a different distribution, and in fact have their own lower and upper bounds. This makes the Hoeffding bound very general!

Notice that the format of the bounds in Theorem 18.7 is very similar to that in Theorem 18.4. The difference is that the n in the denominator of the exponent in Theorem 18.4 is now replaced by $\sum_{i=1}^{n}(b_i - a_i)^2$. Notice that the Hoeffding bound becomes smaller for higher δ, and becomes larger as $b_i - a_i$ increases.

18.10 Appendix: Proof of Lemma 18.5

This appendix contains the technical details needed to prove Lemma 18.5. We start with a basic identity from calculus:

Lemma 18.8 *If $g(0) = h(0)$ and $g'(k) \leq h'(k)$ for all $k \geq 0$, then $g(t) \leq h(t)$ for all $t \geq 0$.*

Proof:

$$\begin{aligned} h(t) - g(t) &= (h(t) - g(t)) - (h(0) - g(0)) \\ &= \int_0^t (h'(k) - g'(k))dk \\ &\geq 0, \qquad \text{because } h'(k) - g'(k) \geq 0. \end{aligned}$$ ■

Lemma 18.9 *For all $0 \leq p \leq 1$ and $t \geq 0$,*

$$pe^{t(1-p)} + (1-p)\,e^{-tp} \leq e^{\frac{t^2}{8}}. \tag{18.21}$$

Proof: Multiplying both sides of (18.21) by e^{tp} yields

$$f(t) \equiv pe^t + 1 - p \leq e^{tp + \frac{t^2}{8}}. \tag{18.22}$$

Now taking the natural log of both sides of (18.22) yields

$$g(t) \equiv \ln(f(t)) = \ln\left(pe^t + (1-p)\right) \leq \frac{t^2}{8} + tp \equiv h(t). \tag{18.23}$$

It suffices to show that $g(t) \leq h(t)$, $\forall t \geq 0$, as defined in (18.23).

Note that $f'(t) = f''(t) = pe^t$ and $0 \leq f'(t) \leq f(t)$, so $0 \leq \frac{f'(t)}{f(t)} \leq 1$.

Since $g(t) = \ln(f(t))$, we have that $g'(t) = \frac{f'(t)}{f(t)}$. Furthermore, using the fact that $f''(t) = f'(t)$, we have

$$g''(t) = \frac{f(t)\, f''(t) - f'^2(t)}{f^2(t)} = \left(1 - \frac{f'(t)}{f(t)}\right) \cdot \frac{f'(t)}{f(t)} \leq \frac{1}{4}. \quad (18.24)$$

The last step involving the $\frac{1}{4}$ comes from the fact that, for all x, the quantity $(1-x)(x)$ is maximized at $x = \frac{1}{2}$.

Since $g(0) = h(0)$, by Lemma 18.8 it suffices to show that $g'(t) \leq h'(t)$ for all $t \geq 0$.

Since $g'(0) = h'(0)$, by Lemma 18.8 it suffices to show that $g''(t) \leq h''(t)$ for all $t \geq 0$. But this latter statement is true because, by (18.24),

$$g''(t) \leq \frac{1}{4} = h''(t). \qquad \blacksquare$$

18.11 Exercises

18.1 **Chebyshev bound**
A coin has probability $p = \frac{1}{3}$ of coming up heads on each flip. You flip the coin n times. Let X denote the number of heads you get. Use Chebyshev's inequality to upper bound the quantity: $\mathbf{P}\left\{X \geq \frac{1}{2}n\right\}$.

18.2 **Test scores: easy bounds**
Suppose I know only that the mean test score is 40%.
(a) What can I say about the fraction of the class with test score > 80%?
(b) Suppose I'm given further information that the standard deviation of test scores is 10%. What can I now say about the fraction of the class with test score > 80%?

18.3 **Reverse Markov inequality**
Let Y be a non-negative r.v. which is never greater than value b. Let $0 < a < b$. Prove:

$$\mathbf{P}\{Y \leq a\} \leq \frac{\mathbf{E}[b - Y]}{b - a}.$$

18.4 **The distribution of the average**
There are $n = 25$ students in my class. Their scores are independent

because they don't talk to each other, ever! Each student's score is well modeled by a r.v. (not necessarily Normal) with mean 40% and standard deviation of 10% (it's a hard class). Approximately what's the chance that the class average, A_n, exceeds 50%?

(a) What does Chebyshev's inequality tell us about $\mathbf{P}\{A_n > 50\%\}$?

(b) For large n, what does the CLT tell us about $\mathbf{P}\{A_n > 50\%\}$?

18.5 **Sunny Sundays**

Sundays are sunny with probability $\frac{7}{10}$, while all other days are, independently, only sunny with probability $\frac{7}{40}$. Upper bound the probability that in a sequence of n days (where n is a multiple of 7), at least half of the days are sunny. You'll want a bound that is exponentially decreasing in n.

18.6 **Kurtosis bound**

Let X be a r.v. and $a > 0$ be some constant. Define

$$\text{Kurt}(X) \equiv \mathbf{E}\left[(X - \mathbf{E}[X])^4\right].$$

The Chebyshev bound gives an upper bound on $\mathbf{P}\{|X - \mathbf{E}[X]| > a\}$ in terms of $\mathbf{Var}(X)$. Derive an upper bound on $\mathbf{P}\{|X - \mathbf{E}[X]| > a\}$ in terms of Kurt(X).

18.7 **Coupon collecting**

There are n distinct coupon types that you would like to collect. Each day you are sent a random coupon from among the n types. Let X denote the number of days needed to collect all n distinct coupons, given that coupons are chosen randomly with replacement. The following identity is useful in answering some of the questions below:

$$\sum_{i=1}^{\infty} \frac{1}{i^2} = \frac{\pi^2}{6}.$$

(a) What is $\mathbf{E}[X]$? What does this approach for high n? Write your answer using $\Theta(\cdot)$.

(b) Derive $\mathbf{Var}(X)$. What does this approach for high n? Write your answer using $\Theta(\cdot)$.

(c) Derive an asymptotic upper bound on $\mathbf{P}\{X \geq 2n \ln n\}$ for large n using Markov's inequality.

(d) Derive an asymptotic upper bound on $\mathbf{P}\{X \geq 2n \ln n\}$ for large n using Chebyshev's inequality. Express your answer using $\Theta(\cdot)$.

Note: For $\mathbf{E}[X]$ in (c) and (d), use the asymptotic mean from part (a).

18.8 **Getting a job**

Jiacheng has independent probability 50% of being hired by each company at which he interviews. Suppose Jiacheng interviews at 20 companies. What is the probability that Jiacheng doesn't get a job?

(a) Use the Chernoff bound in Theorem 18.4 to upper bound the probability that Jiacheng doesn't get a job.
(b) Now use the Chernoff bound in Theorem 18.6 to upper bound the probability that Jiacheng doesn't get a job.
(c) Now compute the exact probability that Jiacheng doesn't get a job.

18.9 **Bounding wealth**
Keshav's Robinhood stock trading account loss limit is \$1000 dollars. Thus on any given day Keshav's account value, V, can range from $-\$1000$ to ∞. Suppose that all we know about Keshav is that his average Robinhood account value is \$3000 dollars. Can we say anything about the fraction of time that Keshav's account value is at least \$9000? Find the tightest upper bound, t, such that

$$\mathbf{P}\{V \geq 9000\} \leq t.$$

(a) Find a bound t such that $\mathbf{P}\{V \geq 9000\} \leq t$.
(b) Prove that the t that you found in part (a) is tight. Specifically, show that there exists a distribution, V, such that $\mathbf{E}[V] = 3000$ and $V \geq -1000$ and $\mathbf{P}\{V \geq 9000\} = t$.

18.10 **The tightness of Markov's inequality**
Markov's inequality says that, for any non-negative r.v. X,

$$\mathbf{P}\{X \geq k\mathbf{E}[X]\} \leq \frac{1}{k}.$$

After reading this chapter, you likely got the impression that Markov's inequality is quite weak. Prove that Markov's Inequality is "tight" in the following sense: For any given $k \geq 1$, there exists a non-negative r.v. X such that $\mathbf{P}\{X \geq k\mathbf{E}[X]\} = \frac{1}{k}$.

18.11 **Tightness of Chebyshev's inequality**
Chebyshev's inequality tells us that for all random variables X,

$$\mathbf{P}\{|X - \mathbf{E}[X]| \geq a\} \leq \frac{\mathbf{Var}(X)}{a^2}.$$

Prove that Chebyshev's Inequality is "tight" in the following sense: Give a r.v. X (not equal to a constant) and a value $a > 0$ for which the above inequality is met at equality.

18.12 **Concentration bounds for pair-wise independent random variables**
Let $X_1, X_2, \ldots, X_n$ be pairwise-independent random variables, satisfying $\mathbf{Var}(X_i) \leq 10$ for all $i = 1, \ldots, n$. Let $X = \sum_{i=1}^{n} X_i$. Prove that for all $a > 0$,

$$\mathbf{P}\{|X - \mathbf{E}[X]| \geq a\} \leq \frac{10n}{a^2}.$$

18.13 **Weak Law of Large Numbers**
Let $X_1, X_2, X_3, \ldots,$ be i.i.d. with finite mean $\mathbf{E}[X]$ and finite variance σ^2. Let $S_n = \sum_{i=1}^n X_i$. Your goal is to prove the Weak Law of Large Numbers:

$$\forall \epsilon > 0, \qquad \lim_{n\to\infty} \mathbf{P}\left\{ \left| \frac{S_n}{n} - \mathbf{E}[X] \right| > \epsilon \right\} = 0,$$

where $S_n = \sum_{i=1}^n X_i$. [Hint: Use Chebyshev's Inequality.]

18.14 **Comparing bounds on tail of Exponential**
Let $X \sim \text{Exp}(\lambda)$, where $\lambda > 0$. We will evaluate $\mathbf{P}\left\{X \geq \frac{a}{\lambda}\right\}$, the probability that X is at least a times its mean, where $a > 1$.
(a) What is $\mathbf{P}\left\{X \geq \frac{a}{\lambda}\right\}$ exactly?
(b) What does the Markov bound tell us about $\mathbf{P}\left\{X \geq \frac{a}{\lambda}\right\}$?
(c) What does the Chebyshev bound tell us about $\mathbf{P}\left\{X \geq \frac{a}{\lambda}\right\}$?
(d) What does the Chernoff bound tell us about $\mathbf{P}\left\{X \geq \frac{a}{\lambda}\right\}$?
[Hint: Pick t s.t. $0 < t < \lambda$.]
(e) How far off is the Chernoff bound from the correct answer?

18.15 **Chernoff bound for Binomial with 1/−1 variables**
Let $X_1, X_2, \ldots, X_n$ be i.i.d random variables, where

$$X_i = \begin{cases} 1 & \text{w/prob } 0.5 \\ -1 & \text{w/prob } 0.5 \end{cases}.$$

Let $X = \sum_{i=1}^n X_i$, where $\mu = \mathbf{E}[X] = 0$. Assume $a > 0$. Follow the steps below to prove from first principles that

$$\mathbf{P}\{X \geq a\} \leq e^{-\frac{a^2}{2n}}.$$

(a) Start by setting up the usual Chernoff-based inequality for $\mathbf{P}\{X \geq a\}$, based on exponentiating and then applying the Markov bound.
(b) Prove that $\mathbf{E}\left[e^{tX_i}\right] < e^{t^2/2}$, where $t > 0$. [Hint: Taylor series]
(c) Form a simple closed-form bound for $\mathbf{E}\left[e^{tX}\right]$ and use this to get a simple expression for $\mathbf{P}\{X \geq a\}$ in terms of t.
(d) Find the t that minimizes $\mathbf{P}\{X \geq a\}$ and use this to get the final result.
(e) What can you say about $\mathbf{P}\{|X| \geq a\}$?

18.16 **Chernoff change of variable**
Let $Y_1, Y_2, \ldots, Y_n$ be i.i.d. random variables, where

$$Y_i = \begin{cases} 1 & \text{w/prob } 0.5 \\ 5 & \text{w/prob } 0.5 \end{cases}.$$

Let $Y = \sum_{i=1}^n Y_i$, where $\mu = \mathbf{E}[Y] = 3n$. For $a > 0$, derive a bound on $\mathbf{P}\{Y - \mu \geq a\}$. To do this, you will exploit the result in Exercise 18.15 by defining a simple linear transformation between the Y_i's in this exercise and the $1/-1$ random variables in Exercise 18.15.

18.17 **Chernoff bound for sum of Exponentials**
Let $X = \sum_{i=1}^{n} X_i$, where the X_i's are i.i.d. and are Exponentially distributed with rate $\lambda > 0$. Use Chernoff bounds to derive an upper bound on the probability that X is at least twice its mean.

18.18 **Tail on the sum of Uniforms**
Let $X = X_1 + \cdots + X_n$ where the X_i's are i.i.d. with $X_i \sim \text{Uniform}(0, 1)$. What is an upper bound on $\mathbf{P}\left\{X \geq \frac{3n}{4}\right\}$? Please answer this question in two different ways:
(a) Derive a Chernoff bound from scratch, following the usual process involving $\mathbf{E}\left[e^{tX}\right]$. [Hint: You will come across a term of the form $e^t - 1$. Please upper bound this by e^t to make your analysis nicer.]
(b) Compute the answer given by the Hoeffding bound (Theorem 18.7).
(c) Which bound do you expect to be better, (a) or (b)? Is that what happened?

18.19 **Chernoff bound on Binomial**
Complete the proof of Theorem 18.4 by proving (18.12).

18.20 **Chernoff bound for Binomial with different probabilities**
Prove (18.16) from Theorem 18.6, with extensions. Let $X = \sum_{i=1}^{n} X_i$, with independent $X_i \sim \text{Bernoulli}(p_i)$ and $\mu = \mathbf{E}[X] = \sum_{i=1}^{n} p_i$.
(a) Prove that $\forall \epsilon > 0$,

$$\mathbf{P}\{X \geq (1+\epsilon)\mu\} < \left(\frac{e^{\epsilon}}{(1+\epsilon)^{(1+\epsilon)}}\right)^{\mu}.$$

Follow these steps, where $t > 0$:
(i) Prove $\mathbf{E}\left[e^{tX}\right] < e^{(e^t-1)\mu}$. [Hint: Use $1 + x < e^x$ from (1.12).]
(ii) Apply the usual Chernoff bound technique to upper bound $\mathbf{P}\{X \geq (1+\epsilon)\mu\}$. Write your answer as compactly as possible.
(iii) Find a $t > 0$ that minimizes the answer in the previous step.
(iv) Substitute in that t to yield the desired bound on $\mathbf{P}\{X \geq (1+\epsilon)\mu\}$.
(b) Follow the steps below to prove that, if $0 < \epsilon \leq 1$,

$$\mathbf{P}\{X \geq (1+\epsilon)\mu\} < e^{-\frac{\epsilon^2\mu}{3}}.$$

(i) Using the result of part (a), write what you need to show as an inequality where the right-hand side is $(1+\epsilon)\ln(1+\epsilon)$.
(ii) Derive the Taylor series expansion of $\ln(1+\epsilon)$ where $0 < \epsilon \leq 1$. Then substitute this into your prior expression to prove the needed result.
(c) From the result in part (b), deduce this immediate corollary:

$$\text{For } 0 < \gamma < \mu,\ \mathbf{P}\{X - \mu \geq \gamma\} \leq e^{-\frac{\gamma^2}{3\mu}}.$$

18.21 **Chernoff bound for Binomial with different probabilities, continued**
Prove (18.17) from Theorem 18.6, with extensions. Let $X = \sum_{i=1}^{n} X_i$, with independent $X_i \sim \text{Bernoulli}(p_i)$ and $\mu = \mathbf{E}[X] = \sum_{i=1}^{n} p_i$.
(a) Show that for $0 < \epsilon < 1$,
$$\mathbf{P}\{X \le (1-\epsilon)\mu\} \le \left(\frac{e^{-\epsilon}}{(1-\epsilon)^{(1-\epsilon)}}\right)^{\mu}.$$
(b) Show that for $0 < \epsilon < 1$,
$$\mathbf{P}\{X \le (1-\epsilon)\mu\} < e^{-\frac{\epsilon^2\mu}{2}}.$$
[Hint: start by proving that $\ln\left((1-\epsilon)^{(1-\epsilon)}\right) > -\epsilon + \epsilon^2/2$ by using a Taylor series around 0.]

18.22 **Approximating the tail of the Normal distribution**
[Proposed by Arisha Kulshrestha] Recall that we have no closed-form expression for the tail of the Normal distribution, which must be computed by numerically evaluating the integral. Let $X \sim \text{Normal}(0, 1)$. Your goal is to produce upper bounds on $\mathbf{P}\{X \ge a\}$, where $a > 0$.
(a) Use Markov's inequality to bound $\mathbf{P}\{X \ge a\}$. Note: This is not as trivial as it might seem because X is *not* non-negative. It will help to observe that:
$$\mathbf{P}\{X \ge a\} = \mathbf{P}\{X \ge a \mid X > 0\} \cdot \mathbf{P}\{X > 0\}.$$
Now define the non-negative r.v. $Y \equiv [X \mid X > 0]$ and note that $\mathbf{P}\{X \ge a \mid X > 0\} = \mathbf{P}\{Y \ge a\}$.
(b) Use Chebyshev's inequality to bound $\mathbf{P}\{X \ge a\}$.
(c) Use Chernoff bounds following these steps:
(i) Derive $\mathbf{E}\left[e^{tX}\right]$.
(ii) Derive the Chernoff bound for $\mathbf{P}\{X \ge a\}$.

18.23 **Negative Binomial tail**
Suppose we are flipping a coin that lands on heads with probability $p > 0.5$. Let X be the number of heads that we see in n flips. Let Y be the number of flips until we see the kth head. We say that $X \sim \text{Binomial}(n, p)$ and $Y \sim \text{NegBinomial}(k, p)$.
(a) Derive $\mathbf{E}[Y]$ and $\mathbf{Var}(Y)$.
(b) Prove that $\mathbf{P}\{Y > n\} = \mathbf{P}\{X < k\}$. (Just use words to explain why each side implies the other.)
(c) Is $\mathbf{P}\{Y = n\} = \mathbf{P}\{X = k\}$? Explain.
(d) Use the above results and a Chernoff bound to derive an upper bound on $\mathbf{P}\{Y > a\mathbf{E}[Y]\}$, where $Y \sim \text{NegBinomial}(k, p)$ and $a > 1$. You should find that your upper bound decreases as k increases. (Please don't worry about the fact that some quantities might not be integers.)

(e) In part (d) we used the Chernoff bound to derive an upper bound on Y's tail. Now instead use CLT. Apply CLT to approximate the probability that Y is at least twice its mean. You can leave your answers in terms of $\Phi(\cdot)$.

18.24 **Hoeffding's inequality**

[Proposed by Misha Ivkov] In this problem, you will prove Hoeffding's Inequality, Theorem 18.7, which states the following: Let $X_1, X_2, \ldots, X_n$ be independent random variables satisfying $a_i \le X_i \le b_i$ for all i where $a_i \le b_i$ are real numbers. Let $X = \sum_{i=1}^{n} X_i$. Then,

$$\mathbf{P}\{X - \mathbb{E}[X] \ge \delta\} \le \exp\left(-\frac{2\delta^2}{\sum_{i=1}^{n}(b_i - a_i)^2}\right),$$

$$\mathbf{P}\{X - \mathbb{E}[X] \le -\delta\} \le \exp\left(-\frac{2\delta^2}{\sum_{i=1}^{n}(b_i - a_i)^2}\right).$$

(a) Start with the usual Chernoff-based inequality for $\mathbf{P}\{X - \mathbb{E}[X] \ge \delta\}$, based on exponentiating and the Markov bound.

(b) Recall from Definition 5.21 that a real-valued function, $g(\cdot)$, defined on interval $S \subseteq \mathbb{R}$ is *convex* if $\forall \lambda \in [0, 1]$, and $\forall \alpha, \beta \in S$,

$$\lambda g(\alpha) + (1 - \lambda) g(\beta) \ge g(\lambda\alpha + (1 - \lambda)\beta). \tag{18.25}$$

Draw a picture of (18.25) where $g(x) = e^x$ to illustrate that $g(x) = e^x$ is convex.

(c) Suppose that Y is a r.v. which satisfies $0 \le Y \le 1$ and has mean $\mathbf{E}[Y] = \mu$. Use the fact that e^x is convex to prove that

$$\mathbb{E}[e^{tY}] \le \mu e^t + (1 - \mu). \tag{18.26}$$

[Hint: You will start with (18.25), but replace λ with the r.v. Y, which is also in $[0, 1]$. You'll need to set $\alpha = t$, $\beta = 0$.]

(d) Use Lemma 18.5 to go from (18.26) to the expression below:

$$\mathbb{E}[e^{tY}] \le e^{t\mu + t^2/8}. \tag{18.27}$$

(e) Using part (d), derive a bound on $\mathbb{E}[e^{tX_i}]$ in terms of t, a_i, b_i, and μ_i, where μ_i is the mean of X_i. It will help to start by defining

$$Y = \frac{X_i - a_i}{b_i - a_i} \qquad \text{or, equivalently,} \qquad X_i = (b_i - a_i)Y + a_i.$$

(f) Form a simple closed-form bound for $\mathbb{E}[e^{tX}]$. Then use this bound to get a simple bound for $\mathbf{P}\{X - \mathbb{E}[X] \ge \delta\}$ in terms of t.

(g) Find the t that minimizes $\mathbf{P}\{X - \mathbb{E}[X] \ge \delta\}$ and use this to get the final result.

(h) Argue that the bound you showed for $\mathbf{P}\{X - \mathbb{E}[X] \ge \delta\}$ also works for $\mathbf{P}\{X - \mathbb{E}[X] \le -\delta\}$.