Supplementary Lecture I

Tail Bounds

In probabilistic analysis, we often need to bound the probability that a random variable deviates far from its mean. There are various formulas for this purpose. These are called *tail bounds*. The weakest of these is the *Markov bound*, which states that for any nonnegative random variable X with mean $\mu = \mathcal{E}X$,

$$\Pr(X \geq k) \quad \leq \quad \mu/k \tag{I.1}$$

(Miscellaneous Exercise 83). A better bound is the *Chebyshev bound*, which states that for a random variable X with mean $\mu = \mathcal{E}X$ and standard deviation $\sigma = \sqrt{\mathcal{E}((X-\mu)^2)}$, for any $\delta \geq 1$,

$$\Pr(|X - \mu| \geq \delta\sigma) \quad \leq \quad \delta^{-2} \tag{I.2}$$

(Miscellaneous Exercise 84).

The Markov and Chebyshev bounds converge linearly and quadratically, respectively, and are often too weak to achieve desired estimates. In particular, for the special case of Bernoulli trials (sum of independent, identically distributed 0,1-valued random variables) or more generally Poisson trials (sum of independent 0,1-valued random variables, not necessarily identically distributed), the convergence is exponential.

Consider Poisson trials X_i, $1 \le i \le n$, with sum $X = \sum_i X_i$ and $\Pr(X_i = 1) = p_i$. An exact expression for the upper tail is

$$\Pr(X \ge k) = \sum_{\substack{A \subseteq \{1,\ldots,n\} \\ |A| \ge k}} \prod_{i \in A} p_i \prod_{i \notin A} (1 - p_i).$$

In the special case of Bernoulli trials with success probability p, this simplifies to the *binomial distribution*

$$\Pr(X \ge k) = \sum_{i \ge k} \binom{n}{i} p^i (1-p)^{n-i}.$$

However, these expressions are algebraically unwieldy. A more convenient formula is provided by the *Chernoff bound*.

The Chernoff bound comes in several forms. One form states that for Poisson trials X_i with sum $X = \sum_i X_i$ and $\mu = \mathcal{E}X$, for any $\delta > 0$,

$$\Pr(X \ge (1+\delta)\mu) < \left(\frac{e^\delta}{(1+\delta)^{1+\delta}} \right)^\mu. \tag{I.3}$$

Equivalently,

$$\Pr(X \ge (1+\delta)\mu) < \left(e \left(1 - \frac{\delta}{1+\delta} \right)^{(1+\delta)/\delta} \right)^{\delta\mu}. \tag{I.4}$$

In (I.4), the subexpression

$$\left(1 - \frac{\delta}{1+\delta} \right)^{(1+\delta)/\delta} \tag{I.5}$$

is a special case of the function $(1 - 1/x)^x$, which arises frequently in asymptotic analysis. It is worth remembering that this function is bounded above by e^{-1} for all positive x and tends to that value in the limit as x approaches infinity. Similarly, the function $(1 + 1/x)^x$ is bounded above by e for all positive x and tends to that limit as x approaches infinity (Miscellaneous Exercise 57(a)).

A third form equivalent to (I.3) and (I.4) is: for all $k > \mu$,

$$\Pr(X \ge k) < e^{k-\mu}(\mu/k)^k. \tag{I.6}$$

One can see clearly from (I.4) and (I.6) that the convergence is exponential with distance from the mean.

These formulas bound the upper tail of the distribution. There are also symmetric versions for the lower tail: for any δ such that $0 \le \delta < 1$ and k

such that $0 < k \le \mu$,

$$\Pr(X \le (1-\delta)\mu) \quad < \quad \left(\frac{e^{-\delta}}{(1-\delta)^{1-\delta}}\right)^{\mu} \tag{I.7}$$

$$= \quad \left(e^{-1}\left(1+\frac{\delta}{1-\delta}\right)^{(1-\delta)/\delta}\right)^{\delta\mu} \tag{I.8}$$

$$\Pr(X \le k) \quad < \quad e^{k-\mu}(\mu/k)^k. \tag{I.9}$$

In the case of the lower tail, we also have a fourth version given by

$$\Pr(X \le (1-\delta)\mu) \quad < \quad e^{-\delta^2\mu/2}. \tag{I.10}$$

This bound is slightly weaker than (I.7)–(I.9), but is nevertheless very useful because of its simple form.

Proof of the Chernoff Bound

We now prove the Chernoff bound (I.3) for Poisson trials X_i. It is easy to show that the other forms (I.4) and (I.6) are equivalent, and these are left as exercises (Miscellaneous Exercise 87). The proofs of the corresponding bounds (I.7)–(I.9) for the lower tail are similar and are also left as exercises (Miscellaneous Exercise 88). The weaker bound (I.10) requires a separate argument involving the Taylor expansion of $\ln(1-\delta)$, but is not difficult (Miscellaneous Exercise 89).

Although the success probabilities of the X_i may differ, it is important that the trials be independent. At a crucial step of the proof, we use the fact that the expected value of the product of independent trials is the product of their expectations (Miscellaneous Exercise 82).

Let X_i be Poisson trials with success probabilities p_i, sum $X = \sum_i X_i$, and mean $\mu = \mathcal{E}X = \sum_i p_i$. Fix $a > 0$. By the monotonicity of the exponential function and the Markov bound (I.1), we have

$$\begin{aligned}\Pr(X \ge (1+\delta)\mu) \quad &= \quad \Pr(e^{aX} \ge e^{a(1+\delta)\mu}) \\ &\le \quad \mathcal{E}(e^{aX}) \cdot e^{-a(1+\delta)\mu}.\end{aligned} \tag{I.11}$$

Because the expected value of the product of independent trials is the product of their expectations (Miscellaneous Exercise 82), and because the e^{aX_i} are independent if the X_i are, we can write

$$\begin{aligned}\mathcal{E}(e^{aX}) \quad &= \quad \mathcal{E}(e^{\sum_i aX_i}) \quad = \quad \mathcal{E}(\prod_i e^{aX_i}) \quad = \quad \prod_i \mathcal{E}(e^{aX_i}) \\ &= \quad \prod_i (p_i e^a + (1-p_i)) \quad = \quad \prod_i (1 + p_i(e^a - 1)).\end{aligned} \tag{I.12}$$

It follows from $(1+1/x)^x < e$ that $1+y < e^y$ for all positive y. Applying this with $y = p_i(e^a - 1)$, we have $1 + p_i(e^a - 1) < e^{p_i(e^a-1)}$, thus (I.12) is strictly bounded by

$$\prod_i e^{p_i(e^a-1)} \quad = \quad e^{\sum_i p_i(e^a-1)} \quad = \quad e^{(e^a-1)\mu}.$$

Combining this with the expression $e^{-a(1+\delta)\mu}$ gives a strict bound

$$e^{(e^a-1)\mu} \cdot e^{-a(1+\delta)\mu} \quad = \quad e^{(e^a-1-a-a\delta)\mu} \tag{I.13}$$

on (I.11). Now we wish to choose a minimizing $e^a - 1 - a - a\delta$. The derivative vanishes at $a = \ln(1+\delta)$, and substituting this value for a in (I.13) and simplifying yields (I.3).