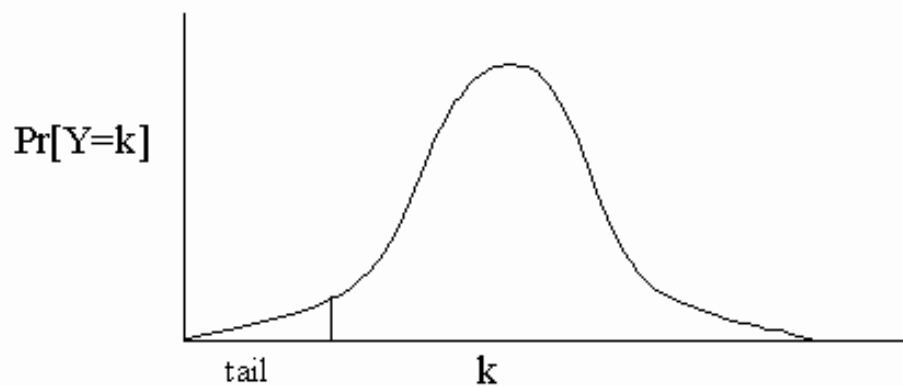


Tail Bounds

Last time we looked at occupancy problems and derived some results on the distribution of some random variables. We derived bounds on the probability of a bin containing more than k balls, and the expected number of bins containing exactly k balls for various k . We had to do specific analysis for balls and bins. This lecture gives two bounds that work for any probability distribution. One (Markov) requires almost no knowledge about the distribution but gives very weak bounds. The other (Chebyshev) needs knowledge of standard deviation but gives good bounds.



The bounds we study are called tail bounds because they correspond to the area or total probability of tails under a probability distribution like the one above.

Markov Bounds

Let Y be a non-negative random variable, and t a positive number. The Markov bound is

$$\Pr[Y \geq t] \leq \frac{E[Y]}{t}$$

The bound doesn't depend on any knowledge of the distribution of Y , except that it's non-negative. The proof is easy once the bound is rewritten as:

$$t\Pr[Y \geq t] \leq E[Y]$$

That inequality comes from the definition of $E[Y]$. $E[Y]$ is the sum of all possible values of Y times the probability of each value, which we can write as

$$E[Y] = \sum_{v < t} v\Pr[Y = v] + \sum_{v \geq t} v\Pr[Y = v]$$

The first term is non-negative because Y is, so we can remove it and make the inequality:

$$E[Y] \geq \sum_{v \geq t} v \Pr[Y = v] \geq \sum_{v \geq t} t \Pr[Y = v]$$

And then the last term rewrites as

$$E[Y] \geq t \sum_{v \geq t} \Pr[Y = v] = t \Pr[Y \geq t]$$

Which is the inequality we wanted to prove.

Examples

The Markov bound is usually not very exciting. Consider placing n balls into n bins as per last lecture. Then the number of balls in any bin is a non-negative random variable, whose expected value is 1. Let Y be the number of balls in bin 1. $E[Y] = 1$, so by Markov

$$\Pr[Y \geq k] \leq \frac{1}{k}$$

Or if $k = 10$, the probability of more than k balls is less than 0.1. But we know from last lecture that that probability is much smaller (about 10^{-6}). So the Markov is not a tight bound. That's not surprising, because it assumes nothing about the distribution of Y , and this distribution like many others, has small tails.

Chebyshev Bounds

Chebyshev bounds give us a lot more because they use more information about the distribution. Specifically, they use information about the standard deviation of the random variable. Remember that for a random variable X , the variance $\text{Var}[X]$ is defined as

$$\text{Var}[X] = E[(X - \bar{X})^2]$$

Where the expected value of X is

$$\bar{X} = E[X]$$

And the standard deviation σ_X is defined as the square root of the variance. Then the Chebyshev bound for a random variable X with standard deviation σ_X is:

$$\Pr[|X - \bar{X}| \geq t\sigma_X] \leq \frac{1}{t^2} \quad \text{or equivalently} \quad \Pr[|X - \bar{X}| \geq s] \leq \frac{\text{Var}[X]}{s^2}$$

Proof

The proof is by defining

$$Y = (X - \bar{X})^2$$

And then applying the Markov bound to Y . First notice that

$$\Pr[Y \geq s^2] = \Pr[(X - \bar{X})^2 \geq s^2] = \Pr[|X - \bar{X}| \geq s]$$

Now apply Markov to this probability:

$$\Pr[|X - \bar{X}| \geq s] = \Pr[Y \geq s^2] \leq \frac{E[Y]}{s^2} = \frac{\text{Var}[X]}{s^2}$$

since $E(Y)$ is the variance $\text{Var}[X]$.

Variance of a Bernoulli trial

Let Y be an indicator random variable for a Bernoulli trial, where

$$Y = \begin{cases} 1 & \text{if the trial succeeds} \\ 0 & \text{if it fails} \end{cases}$$

Let $p = \Pr[Y = 1]$. Then the expected value $E[Y] = p$. The variance is

$$E[(Y - \bar{Y})^2] = p^2\Pr[Y = 0] + (1 - p)^2\Pr[Y = 1] = p^2(1 - p) + (1 - p)^2p$$

Or

$$\text{Var}[Y] = p(1 - p)$$

Variance of a Sum of Independent Random variables

Now suppose Y_i for $i = 1, \dots, n$ are independent random variables, and $Y = \sum Y_i$. Then

$$\text{Var}[Y] = \sum_{i=1}^n \text{Var}[Y_i]$$

You can check this yourself. Prove it for two random variables, and then extend to n using induction.

Example

In order to apply the Chebyshev bound to the occupancy problem, we need to compute the variance of Y , the number of balls in bin 1. Let

$$Y_j = \begin{cases} 1 & \text{if ball } j \text{ goes into bin 1} \\ 0 & \text{otherwise} \end{cases}$$

Then because the Y_j 's are independent their variances add:

$$\text{Var}[Y] = \sum_{i=1}^n \text{Var}[Y_i] = n\text{Var}[Y_i]$$

because the bins are indistinguishable and all the Y_j have the same variance.

Now since Y_j is an indicator, $E[Y_j] = \Pr[Y_j = 1] = 1/n$. It represents a Bernoulli trial with success probability $p = 1/n$, and from the earlier result on Bernoulli r.v.'s, its variance is:

$$\text{Var}[Y_j] = p(1 - p) = (n - 1)/n^2$$

And therefore

$$\text{Var}[Y] = n\text{Var}[Y_j] = (n - 1)/n$$

Which is very close to 1. If $Y \geq k$, then $|Y - E[Y]| \geq k - 1$, and by Chebyshev:

$$\Pr[Y \geq k] \leq \Pr[|Y - E[Y]| \geq k - 1] \leq \frac{\text{Var}[Y]}{(k - 1)^2} \approx \frac{1}{(k - 1)^2}$$

So Chebyshev gives us a bound that falls off as the inverse of the square of k . For $k = 10$, the bound is $1/81$, or about 0.012

Application: Randomized Selection

Selection is the problem of finding an element of rank k from some set. The rank $r_S(x)$ of an element x in a set S is k if there are $k - 1$ elements in S that are smaller than x . The problem is only interesting if S is represented as an unsorted array. If the array is sorted, it is of course trivial to find the element of rank k . You might have seen a (complex) linear-time selection algorithm in CS170. We present a simpler randomized algorithm here. The idea is to randomly pick a “smallish” subset of elements from the set, and to sort them. Then you find a pair of elements a and b from the sorted subset such that $[a, b]$ should contain the element you’re looking for. You compare all the elements with $[a, b]$ and create a new subset P of elements in that interval. By knowing how many elements are less than a , you figure out x 's rank m in P . Then sort P and pick the element in the m^{th} position.

Algorithm Lazyselect (from Motwani and Raghavan)

Input: An (unsorted) array S of n elements, and an integer k in the range $1, \dots, n$.

Output: The k^{th} smallest element of S .

1. Pick $n^{3/4}$ elements of S , chosen independently and uniformly at random with replacement, and place them in an array R .
2. Sort R in $O(n^{3/4} \log n)$ steps using any optimal sorting algorithm.
3. Let $x = kn^{-1/4}$. (This is a guess at the rank of the nearest element in R to the element we want from S). Now define:

$$l = \max(\lfloor x - \sqrt{(n)} \rfloor, 1) \quad \text{and} \quad h = \min(\lceil x + \sqrt{(n)} \rceil, n^{3/4})$$

And let $a = R[l]$, $b = R[h]$. Compare a and b with all the elements of S (linear time) to determine the ranks $r_S(a)$ and $r_S(b)$. While doing this, put elements of S that are in the interval $[a, b]$ into an array P .

4. Check if $r_S(a) \leq k \leq r_S(b)$. This is equivalent to checking if the k^{th} smallest element of S is in P . Check also if $|P| \leq 4n^{3/4} + 2$. If either test fails, repeat steps 1-3 until success.
5. Sort P in $O(n^{3/4} \log n)$ steps, return $P[k - r_S(a) + 1]$ which is the k^{th} smallest element of S .

Theorem: With probability $1 - O(n^{-1/4})$ Algorithm LazySelect succeeds on the first iteration. It performs at most $2n + o(n)$ comparisons.

Proof: The test for success is at step 4. If $k < r_S(a)$ or $k > r_S(b)$ or if P is too big, we have to iterate. Consider $k > r_S(b)$ first. Let X_i be a random variable which is

$$X_i = \begin{cases} 1 & \text{if the } i^{\text{th}} \text{ random sample for } R \text{ has rank } \leq k \text{ in } S \\ 0 & \text{otherwise} \end{cases}$$

So $\Pr[X_i = 1] = k/n$ and $\Pr[X_i = 0] = 1 - k/n$. Define

$$X = \sum_{i=1}^{n^{3/4}} X_i$$

then X is the total number of elements of R that have rank $\leq k$ in S , and

$$\bar{X} = n^{3/4} \mathbb{E}[X_i] = \frac{k}{n} n^{3/4} = kn^{-1/4}$$

Which is also the value of x from the algorithm. Since X is a sum of indicator variables for independent Bernoulli trials X_i where each trial has success probability $p = k/n$, its variance is given by:

$$\text{Var}[X] = n^{3/4} p(1-p) = n^{3/4} \frac{k}{n} \left(1 - \frac{k}{n}\right) \leq n^{3/4}/4$$

The $1/4$ in the last bound follows because the maximum value of $p(1 - p)$ for $p \in [0, 1]$ is $1/4$ (Calculus).

Now if $k > r_S(b)$, then $\geq h$ elements of R (the number which are $\leq b$ in R) were chosen from S which had rank $\leq k$ in S . In other words, $X \geq h$. Now since $h = x + \sqrt{n}$, we have that $\Pr[X \geq h] = \Pr[X - x \geq \sqrt{n}]$. Since x is the expected value of X , this is in a form that we can apply Chebyshev to:

$$\Pr[|X - x| \geq \sqrt{n}] \leq \frac{\text{Var}[X]}{n} = \frac{n^{3/4}}{4n} = \frac{1}{4}n^{-1/4}$$

which is the bound we wanted. Applying a very similar argument to $k < r_S(a)$ shows that $\Pr[k \leq r_S(a)] = O(n^{-1/4})$. Adding up both these probabilities of failure gives us $O(n^{-1/4})$.

The last test is whether $|P| > 4n^{3/4} + 2$. Notice that $|P| = r_S(b) - r_S(a) + 1$, and for the test to fail we must have either $r_S(b) - k > 2n^{3/4} + 1$ or $k - r_S(a) > 2n^{3/4} + 1$. The proof follows as above, but start by defining X_i as a random variable which is 1 if the i^{th} random sample for R has rank $> k + 2n^{3/4}$, or rank $< k - 2n^{3/4}$ respectively, in S . The failure probability is once again $O(n^{-1/4})$, so the total failure probability is $O(n^{-1/4})$, which completes the proof.