Applications of Tail Inequalities

- Here is one more practical application of the Chernoff bounds.
- Consider dividing a data set with features of interest into two parts: a test set and a training set.
- To make sure that both are roughly similar copies, you want to divide so that both data sets have the same number of data items of any given feature.
- Similar requirements arise in also experiments related to drug trails.

Example

Candidate	Above 50 Yrs	Diabetic	Frontline worker	Hypertension	Female
C1	1	0	1	1	0
C2	1	0	0	0	1
C3	0	0	0	0	1
C4	0	1	1	1	0
C5	1	1	0	0	0

Applications of Tail Inequalities

- To simplify matters, we will think of n data items with n features.
- Prepare a matrix A of nxn with entries from {0, 1}.
- Rows are features, columns are data items
- The goal is to find a vector x of size n with entries from {-1, +1} such that Ax has the smallest possible maximum absolute entry.
  - Rows with +1 in x belong to one class and those with -1 in x belong to another class.
  - The maximum absolute entry in Ax indicates how many data items differ at feature i according to the division by x.

Applications of Tail Inequalities

- Example

$$A = \begin{bmatrix} 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 \\ 0 & 1 & 0 & 1 \end{bmatrix} \qquad x = \begin{bmatrix} 1 \\ -1 \\ 1 \\ -1 \end{bmatrix} \qquad Ax = \begin{bmatrix} -2 \\ 1 \\ -1 \\ -2 \end{bmatrix}$$

- The maximum absolute entry in Ax is 2.

Applications of Tail Inequalities

- No good deterministic algorithms are known.
- The brute-force algorithm has to try all possible 2^n vectors.
- However, a very simple randomized algorithm exists.
- Consider choosing each element of x uniformly at random from {1, -1}.
- We will show that the maximum absolute entry of Ax in such an x is bounded by $O((n \ln n)^{1/2})$ with high probability.

Applications of Tail Inequalities

- Consider choosing each element of x uniformly at random from {1, -1}.
- We will show that the maximum absolute entry of AX in such an X is bounded by $O((n \ln n)^{1/2})$ with high probability.
- Let the product AX = Y.
- Consider any Y_i, say Y_1.
- By definition of matrix multiplication, $Y_1 = A_{11}X_1 + A_{12}X_2 + \cdot\cdot\cdot + A_{1n}X_n$ where the A_{ij} denotes the element of A at the ith row and jth column.

Applications of Tail Inequalities

- Consider choosing each element of x uniformly at random from {1, -1}.
- We will show that the maximum absolute entry of AX in such an X is bounded by $O((n \ln n)^{1/2})$ with high probability.
- Let the product AX = Y.
- Consider any Y_i, say Y_1.
- By definition of matrix multiplication, $Y_1 = A_{11}X_1 + A_{12}X_2 + \cdot\cdot\cdot + A_{1n}X_n$
- Note that $E[X_i] = 0$ and by linearity of expectations, $E[Y_1] = 0$.

Applications of Tail Inequalities

- We will show that the maximum absolute entry of AX in such an X is bounded by $O((n \ln n)^{1/2})$ with high probability.
- Let the product $AX = Y$. Consider any Y_i, say Y_1.
- By definition of matrix multiplication, $Y_1 = A_{11}X_1 + A_{12}X_2 + \cdots + A_{1n}X_n$
- Note that $E[X_i] = 0$ and by linearity of expectations, $E[Y_1] = 0$.
- Let us imagine the case where the choices of X are made independently.
- We can then apply Chernoff bound on Y.

Applications of Tail Inequalities

- One small detour before we do that.
- In our Chernoff bound derived earlier, we studied the case of the rv T being a sum of iid Bernoulli rv's T_i.
- Now, each X_i is a +1/-1 valued random variable.
- Need a new version of Chernoff bound!

Applications of Tail Inequalities

- Consider X as the random variable that is the sum of n independent and identically distributed random variables X_i with X_i taking value among {-1, +1}.
- Let $Pr[X_i = +1] = Pr[X_i = -1] = 1/2$.
- Note that $E[X] = n.E[X_i] = n.0 = 0$.
- We want to know the Prob(X >= k) for some integral k.
  - Of course k has values between –n and +n.

Applications of Tail Inequalities

- Consider X as the random variable that is the sum of n independent and identically distributed random variables X_i with X_i taking value among $\{-1, +1\}$.
- Let $Pr[X_i = +1] = Pr[X_i = -1] = 1/2$.
- Note that $E[X] = n.E[X_i] = n.0 = 0$.
- We want to know the Prob($X >= k$) for some integral k.
- Instead of doing the entire calculation again, let us do the following.
- Define $T_i = (1+X_i)/2$. Now, T_i is $\{0,1\}$ valued.
- Define T as the sum of T_i's.

Applications of Tail Inequalities

- Consider X as the random variable that is the sum of n independent and identically distributed random variables X_i with X_i taking value among {-1, +1}.
- We want to know the Prob(X >= k) for some integral k.
- Define $T_i = (1+X_i)/2$. Now, T_i is {0,1} valued.
- Define T as the sum of T_i's.
- $T = \Sigma\, T_i = \Sigma(1+X_i)/2 = n/2 + X/2$
- Note that ET = n/2.
- Also, X >= k if and only if T >= n/2 + k/2.
- Now, Pr(X >= k) = Pr(T >= n/2 + k/2) = Pr(T >= n/2(1+k/n)).
  - So, we will use δ = k/n when applying Chernoff bounds on T.

Applications of Tail Inequalities

- Also, $X >= k$ if and only if $T >= n/2 + k/2$.
- Now, $Pr(X >= k) = Pr(T >= n/2 + k/2)$.
- Rewrite as $Pr(T >= E[T](1+(k/n))$ with $\delta = k/n < 1$.
- Applying Chernoff bounds with the above δ, we get that $Pr(T >= E[T](1+\delta)) <= \exp(-E[T]\delta^2/4) = \exp(-\delta^2/8n)$.

Applications of Tail Inequalities

- Back to our problem of set balancing....
- We now get $\Pr(Y_1 >= 8 \sqrt{n \ln n}) \le \exp(-64 n \ln n / 8n) = \exp(-8 \ln n) = 1/n^8$.
- But we are interested in a two-sided bound.
- That is, since we want to minimize the absolute value of Y_1, we need to compute $\Pr[Y_1 \le -8 \sqrt{n \ln n}]$ also.
- But by symmetry, we have that $\Pr(Y_1 <= -8 \sqrt{n \ln n}) <= 1/n^8$.
- So, $\Pr[|Y_1| >= 8(n \ln n)^{1/2}] <= 2/n^8$.

Applications of Tail Inequalities

- Back to our problem of set balancing....
- So, $\Pr[|Y_1| >= 8(n \ln n)^{1/2}] <= 2/n^8$.
- But, what about Y_2, Y_3, etc.
- This is where another classical probability result aids us.
- Boole's inequality. For any events, $E_1, E_2, \ldots, E_n$,
- $\Pr(E_1 \cup E_2 \cup .. \cup E_n) <= \Pr(E_1) + \Pr(E_2) + \ldots + \Pr(E_n)$.
- Apply the above to get that with probability at least $1-2/n^7$, every Y_i has an absolute value that is within $8(n \ln)^{1/2}$.

Another Application of Tail Inequalities

- We consider the randomized rounding problem defined as follows.
- Consider a Boolean matrix A of size nxn. Let X be a vector of size n with each element from [0, 1]. We want to find a Boolean vector Y of size n such that Y is as close to X as possible in the following sense.
  - The maximum absolute entry in A(Y – X) is minimized.

Another Application of Tail Inequalities

- Here is an example.
- Let $A = \begin{bmatrix} 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix}$ and $X = \begin{bmatrix} 0.2 \\ 0.4 \\ 0.3 \\ 0.7 \end{bmatrix}$. With $Y = \begin{bmatrix} 0 \\ 1 \\ 0 \\ 1 \end{bmatrix}$ we have A(Y – X) = $\begin{bmatrix} 0.3 \\ -0.2 \\ 0.7 \\ 0.3 \end{bmatrix}$.
- The largest absolute entry is 0.7.

Another Application of Tail Inequalities

- Finding such a vector Y using deterministic methods is very inefficient.
- So, we appeal to randomization as follows.
- Irrespective of the matrix A, we just let each Y_i be 0 with probability $1 – X_i$ and 1 with probability X_i.
- We now show how good this simple choice of Y is.

Another Application of Tail Inequalities

- Let us first notice that for any i, $EY_i = 0x(1-X_i)+ 1xX_i = X_i$.
- Further, consider $Z := A(Y – X)$.
- We have $Z_1 = A_{11}(Y_1 – X_1) + A_{12} (Y_2 – X_2) + \ldots + A_{1n} (Y_n – X_n)$ by definition.
- We can without loss of generality assume that each $A_{1i} = 1$, so that the value of Z_1 is maximized.
- So, $EZ_1 = E(\Sigma_i (Y_i – X_i)) = \Sigma_i (EY_i – X_i) = \Sigma_i (X_i – X_i)= 0$.

Another Application of Tail Inequalities

- So, $EZ_1 = E(\Sigma_i (Y_i - X_i)) = \Sigma_i (EY_i - X_i) = \Sigma_i (X_i - X_i) = 0$.
- Let us consider a parameter $d = 2(n \ln n)^{1/2}$.
- We want to estimate $\Pr(Z_1 \geq d)$.
- Notice that Z_1 is the sum of independent random variables $(Y_i - X_i)$.
- Notice that the random variable $Y_i - X_i$ takes values $\{-X_i, 1-X_i\}$.
  - Instead of 0,1 values in our earlier discussions.
- So, these X_i rv's are not identically distributed.
- We can derive Chernoff bounds for this case too.

Another Application of Tail Inequalities

- Notice that Z_1 is the sum of independent random variables ($Y_i - X_i$).
- Notice that the random variable $Y_i - X_i$ takes values $\{-X_i, 1-X_i\}$.
  - Instead of 0,1 values in our earlier discussions.
- So, these X_i rv's are not identically distributed.
- We can derive Chernoff bounds for this case too.
- Theorem: Let each X_i , $1 \leq i \leq n$ be $\{a_i , a_i + 1\}$ valued for some real a_i and X_i's be independent. Let $X = \Sigma_i X_i$ and $E[X] = \mu$. Then for any $\delta > 0$, $\Pr[X \geq \mu + \delta n] \leq \exp\{-2 \delta^2 n\}$ and $\Pr[X \leq \mu - \delta n] \leq \exp\{-2 \delta^2 n\}$.
  - Proof is an offline exercise for you.

Another Application of Tail Inequalities

- Notice that Z_1 is the sum of independent random variables $(Y_i – X_i)$.
  - Notice that the random variable $Y_i – X_i$ takes values $\{-X_i, 1-X_i\}$.
- Theorem: Let each X_i , $1 \le i \le n$ be $\{a_i , a_i + 1\}$ valued for some real a_i and X_i's be independent. Let $X = \Sigma_i X_i$ and $E[X] = \mu$. Then for any $\delta > 0$, $\Pr[X \ge \mu + \delta n] \le \exp\{ -2 \delta^2 n\}$ and $\Pr[X \le \mu - \delta n] \le \exp\{ -2 \delta^2 n\}$.
  - Proof is an offline exercise for you.
- We will apply Chernoff bounds to Z_1 as follows with $E[X] = 0$, and $d = 2(n \ln n)^{1/2}$ and $d = \delta n$ resulting in $\delta = 2(\ln n/n)^{1/2}$.

Another Application of Tail Inequalities

- Notice that Z_1 is the sum of independent random variables ($Y_i - X_i$).
  - Notice that the random variable $Y_i - X_i$ takes values $\{-X_i, 1-X_i\}$.
- We will apply Chernoff bounds to Z_1 as follows with $E[X] = 0$, and $d = 2(n \ln n)^{1/2}$ and $d = \delta n$ resulting in $\delta = 2(\ln n/n)^{1/2}$.
- We get that $\Pr(Z1 \geq d) \leq \exp\{-2\delta^2 n\} = \exp\{-2x4\ln n/n \text{ x } n\} = \exp\{-8\ln n\} = 1/n^8$.

Another Application of Tail Inequalities

- We will apply Chernoff bounds to Z_1 as follows with $E[X] = 0$, and $d = 2(n \ln n)^{1/2}$ and $d = \delta n$ resulting in $\delta = 2(\ln n/n)^{1/2}$.
- We get that $\Pr(Z1 \geq d) \leq \exp\{-2\delta^2 n\} = \exp\{-2x4\ln n/n \text{ x } n\} = \exp\{-8\ln n\} = 1/n^8$.

- We have a few more steps to complete the calculations.
- First, obtain an estimate on $\Pr(Z_1 \leq d)$ also.
- Second, use the Boole's inequality to claim a bound on the event that some Z_i exceeds d in absolute value.