SCRIBE NOTES from CSE 203 - Classes on 29th Jan and 3rd Feb

Vignesh Nallur Cheluve Gowda, vngowda@eng.ucsd.edu

February 5, 2015

Concentration Bounds

1. Objective : Obtain stronger bounds than the ones obtained from Chebyshev's Inequailty for a special case. We obtain the upper bound b on a random variable x by proving that probability that x takes values greater than the bound b is very low.

2. The special case for which we want to obtain a stronger bound is defined below:

   (a) Consider n independent random variables $X_1, X_2, X_3.....X_i....X_n$ such that

   $$for\ all\ i,\ Pr(X_i = 1) = Pr(X_i = -1) = 1/2.$$

   (b) Consider another random variable $X = \sum_{i=1}^{n} X_i$.

   (c) Expected value of the n independent random variables X_i's :

   $$for\ all\ i,\ E(X_i) = 1 * 1/2 - 1 * 1/2 = 0$$

   (d) Variance of the n independent random variables X_i's :

   $$for\ all\ i,\ Var(X_i) = (1 - 0)^2 * 1/2 + (-1 - 0)^2 * 1/2 = 1$$

   (e) Expected value of the random variable X :

   $$E(X) = \sum_{i=1}^{n} E(X_i) = 0$$

   (f) Variance of the random variable X :

   $$Var(X) = \sum_{i=1}^{n} Var(X_i) = n$$

   Our objective is to find a bound on the deviation of X from the expected value.

   Lets try obtaining a bound on the deviation of X from the expected value using Chebyshev's inequality initially.

3. Bound using Chebyshev's Inequality :

   $$Pr(X - E(X) > \rho\lambda) \leq Var(X)/\rho^2\lambda^2, \text{ where } \lambda = \text{standard deviation of X.}$$

   $$Pr(X - 0 > \rho\lambda) \leq \lambda^2/\rho^2\lambda^2$$

   $$Pr(X > \rho\lambda) \leq 1/\rho^2$$

   Thus the bound describing the deviation from the expected value, obtained using Chebyshev's inequality reduces polynomially in ρ, where ρ represents an integer multiple of standard deviation.

4. Stronger bound: For the random variable X defined in (2) we try to find a stronger bound.

(a) We know, $for\ t > 0,\ X > a \Rightarrow tX > ta \Rightarrow e^{tx} > e^{ta}$.

(b) Therefore, $Pr(X > a) = Pr(e^{tX} > e^{ta})$

(c) Expected value of e^{tX_i} :

$$E(e^{tX_i}) = (1/2 * e^t + 1/2 * e^{-t})$$

(d) Expected value of e^{tX} :

$$\begin{aligned} E(e^{tX}) &= E(e^{t\sum_{i=1}^{n} X_i}) \\ &= \prod_{i=1}^{n} (E(e^{tX_i}) \\ &= \prod_{i=1}^{n} (e^t + e^{-t})/2 \end{aligned}$$

(e) We could simplify the above expression using Taylor series. From Taylor series we have the following,

i. $e^t = 1 + t + t^2/2! + t^3/3! + t^4/4! + ...$

ii. $e^{-t} = 1 - t + t^2/2! - t^3/3! + t^4/4! + ...$

iii. $e^t + e^{-t} = 2(1 + t^2/2! + t^4/4! + ...)$

$\leq 2e^{t^2/2}$ (Since the first two terms of the taylor series of $e^{t^2/2}$ are equal to the first two terms above and the subsequent terms in the taylor series of $e^{t^2/2}$ are greater than the subsequent terms above)

(f) Using the expression for $e^t + e^{-t}$ from (e) in (d) we get the following,

$$\begin{aligned} E(e^{tX_i}) &\leq \prod_{i=1}^{n} (e^{t^2/2}) \\ &= e^{nt^2/2} \end{aligned}$$

(g) Using Markov's inequality we have,

$$\begin{aligned} Pr(X > a) &= Pr(e^{tX} > e^{ta}) \\ &\leq E(e^{tX})/e^{ta} \\ &\leq e^{nt^2/2}/e^{at} \text{ (using the expression for } E(e^{tX}) \\ &\qquad \text{from (f))} \end{aligned}$$

(h) We want to obtain a bound B on the value of X. We do this by showing that the probability that X takes values greater than this bound B is very low. Hence, we need to choose a suitable value for t that minimizes $e^{-at+nt^2/2}$. We get this by equating the differential of the above expression with respect to t, to zero and finding the value of t which satisfies the equality. Thus, it turns out that for $t = a/n$, $e^{-at+nt^2/2}$ has the minimum value. This minimum value is $e^{-a^2/2n}$.

(i) Hence, $Pr(X > a) \leq e^{-a^2/2n}$

(j) If $a = \rho\sqrt{n}$, we have $Pr(X > \rho\sqrt{n}) \leq e^{-\rho^2/2}$.

(k) Thus the bound obtained is better than the one obtained using Chebyshev's inequality, since it decreases exponentially in ρ.

(l) If we describe an expression using moment generating function e^{tx}, we capture the distribution of x better than what is captured by Chebyshev's inequality. This is true, since in Chebyshev only the second moment is captured while the moment generating function captures all the moments.

5. Applications :

The above method to find stronger bounds finds applications in statistical experiments. Consider an experiment to test the efficacy of a medicine on a set of m individuals. Also, lets say that there are a total of n features which the m individuals may or may not exhibit.

Objective :

(a) Divide the set of indiviuals into two groups with equal cardinality. One of the groups serving as a control group for the experiment being run on the other group.

(b) Division into two groups should be done in such a way as to minimize the difference between the two groups as much as possible. That is, we want to minimize the difference between the number of individuals exhibiting a feature in the control group to the number of individuals exhibiting the same feature in the other group.

We can represent the presence and absence of these n features in the m individuals in the form of a matrix A of size (n x m) with entries in $\{-1, 1\}$. The index a_{ij} in the matrix A has value 1 if the j^{th} individual has the i^{th} feature, and 0 otherwise.

Now consider a vector $\overrightarrow{b}$ with entries $\{-1, 1\}$ of size n. If $b_i = -1$, it signifies the presence of the i^{th} individual in the control group and if it is 1 it signifies the presence of the i^{th} individual in the other group.

Such a matrix A and a vector $\overrightarrow{b}$ is represented below in the figure,

$$\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1m} \\ a_{21} & a_{22} & \cdots & a_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nm} \end{pmatrix} \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{pmatrix} = \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{pmatrix}.$$

In part (b) of the *Objective*, we said we need to reduce the difference between the two groups as much as possible. Thus, it is clear that we are looking to minimize

$$|A\overrightarrow{b}|_\infty = max_{i=1...n}|c_i|$$

We adopt a randomized process to divide the individuals into two groups. An individual is put into either of the two groups with equal probability (= 1/2). Thus, this involves setting the entries of vector $\overrightarrow{b}$ to either 1 or -1 with probability equal to 1/2. This ensures that the expected number of individuals in the both the groups is equal.

Finding a bound on $|A\overrightarrow{b}|_\infty$:

We try to prove the below bound for $|A\overrightarrow{b}|_\infty$.

$$Pr(||Ab|_\infty| \geq \sqrt{4mln(n)}) \leq 2/n$$

Consider the i^{th}row $a_i = a_{i1}, ...a_{im}$. Let k be the number of $1s$ in that row. If $k \leq \sqrt{4mln(n)}$, then clearly $|a_i.b| = c_i \leq \sqrt{4mln(n)}$. However, if $k > \sqrt{4mln(n)}$ then the k nonzero terms in the sum,

$$c_i = \sum_{i=1}^{m} a_{i.j} b_j$$

are independent random variables, each with probability 1/2 of being either +1 or −1.

Now using the stronger bound $e^{-a^2/2n}$ that we obtained earlier for such a random variable and substituting k for n and $\sqrt{4mln(n)}$ for a, we get

$$
\begin{aligned}
Pr(|c_i| > \sqrt{4mln(n)}) &\leq 2e^{-4mln(n)/2k}) \\
&= 2e^{-2mln(n)/k} \\
&\leq 2e^{-2ln(n)} \text{ (since } m \geq k\text{).} \\
&= 2/n^2
\end{aligned}
$$

Since we have have n rows, the probability that the bound does not fail for any row is obtained by the union bound of 2/n.