Tail Inequalities

William Hunt
LDCSEE,
West Virginia University,
Morgantown, WV
{ William.Hunt@mail.wvu.edu }

1 Introduction

In this chapter, we are interested in obtaining significantly stronger and sharper ways for making the statement: "a particular event occurs with high probability". Using Markov's Inequality just doesn't give us a strong enough statement.

Chernoff bounds are another kind of tail bound. They bound the probability that some random variable (say Y) assumes a value that is away from its expected value ($\mathbf{E}$[Y]), during an experiment. The Chernoff bound technique applies to a class of random variables that can be expressed as a sum of independent random variables, for instance Occupancy problems involving balls and bins

2 The Chernoff Technique for 0/1 Poisson Trials

The random variables we will look at are sums of independent Bernoulli trials. Let $X_1, \cdots, X_n$ be independent Bernoulli trials such that, for $1 \leq i \leq n$, $\mathbf{Pr}[X_i = 1] = p$ and $\mathbf{Pr}[X_i = 0] = 1 - p$. Let $X = \sum_{i=1}^{n} X_i$; then X is said to have the binomial distribution. If we let $X_1, \cdots, X_n$ be independent coin tosses such that, for $1 \leq i \leq n$, $\mathbf{Pr}[X_i = 1] = p_i$ and $\mathbf{Pr}[X_i = 0] = 1 - p_i$, then such tosses are referred to as Poisson trials. We will focus on the random variable $X = \sum_{i=1}^{n} X_i$ where the X_i are Poisson trials. All our bounds apply to the special case where X is binomially distributed.
For any experiment outcome X, given $\delta > 0$ and μ, we wish to answer the following questions about the deviation of the observed value from the expected value.

1. What is the probability that X exceeds $(1 + \delta)\mu$?
   $\mathbf{Pr}[X > (1 + \delta)\mu]$
2. How large should δ be so the probability that X exceeds $(1 + \delta)\mu$ is less than ε?
   $\mathbf{Pr}[X > (1 + \delta)\mu] < \varepsilon$

The Chernoff Bound technique provides a means of answering the above questions.

Definition: 2.1 *For a random variable X, $\mathbf{E}[e^{tX}]$ is called the moment generating function.*

We are going to focus on the sum of independent random variables. In particular, we are going to look at bounds on the sum of independent Poisson trials. The chief idea of the Chernoff Bound technique is to take the moment generating function of X, i.e $\mathbf{E}[e^{tX}]$ and apply the Markov inequality to it.

Theorem: 2.1 *Let $X_1, X_2, \cdots, X_n$ be independent Poisson trials such that, for $1 \le i \le n$, $\mathbf{Pr}[X_i = 1] = p_i$ with $0 < p_i < 1$. Then, for $X = \sum_{i=1}^{n} X_i$, $\mu = \mathbf{E}[X] = \sum_{i=1}^{n} p_i$, and for any $\delta > 0$,*

$$\mathbf{Pr}[X > (1+\delta)\mu] < \left[\frac{e^{\delta}}{(1+\delta)^{(1+\delta)}}\right]^{\mu}$$

Proof:

For any positive real number t,

$$\mathbf{Pr}[X > (1+\delta)\mu] = \mathbf{Pr}[e^{tX} > e^{t(1+\delta)\mu}]$$

Applying the Markov inequality to the right hand side, we get

$$\mathbf{Pr}[e^{tX} > e^{t(1+\delta)\mu}] < \frac{\mathbf{E}[e^{tX}]}{e^{t(1+\delta)\mu}}$$

This is easy to calculate because the products in the exponents become sums in the expected value expression and since each X_i is independent, the expectation of the product becomes the product of the expectations. So, the numerator can be calculated the expectation as:

$$\mathbf{E}[e^{tX}] = \mathbf{E}[e^{t\sum_{i=1}^{n} X_i}] = \mathbf{E}[\prod_{i=1}^{n} e^{tX_i}] = \prod_{i=1}^{n} \mathbf{E}[e^{tX_i}]$$

The random variable e^{tX_i} has the value of e^t with probability p_i and the value 1 with probability $1-p_i$. Therefore, the expected value is $p_i(e^t) + (1-p_i)(1)$ making the numerator:

$$\prod_{i=1}^{n} (p_i e^t + 1 - p_i) = \prod_{i=1}^{n} \left(1 + p_i(e^t - 1)\right)$$

Using the inequality $1 + x < e^x$ with $x = p_i(e^t - 1)$, we get the right hand side to be:

$$\mathbf{Pr}[e^{tX} > e^{t(1+\delta)\mu}] < \frac{\prod_{i=1}^{n} e^{p_i(e^t-1)}}{e^{t(1+\delta)\mu}}$$

$$\mathbf{Pr}[e^{tX} > e^{t(1+\delta)\mu}] < \frac{e^{\sum_{i=1}^{n} p_i(e^t-1)}}{e^{t(1+\delta)\mu}}$$

$$\mathbf{Pr}[e^{tX} > e^{t(1+\delta)\mu}] < \frac{e^{(e^t-1)\mu}}{e^{t(1+\delta)\mu}} \qquad (1)$$

In order to obtain a sharp bound, we choose a t such that the right hand side is minimized. Clearly, we can write the exponent of the RHS in System (1) as: $f(t) = [e^t - 1 - t - t\delta]\mu$.

Putting $f'(t) = 0$, *we get*

$$[e^t - 1 - \delta]\mu = 0$$

$$e^t - 1 - \delta = 0$$

$$e^t = 1 + \delta$$

$$ln(e^t) = ln(1 + \delta)$$

$$t = ln(1 + \delta)$$

Also, note that

$$f''(t)|_{t=\ln(1+\delta)} = \mu e^t|_{t=ln(1+\delta)}$$

$$= \mu e^{ln(1+\delta)} \text{ which is } > 0 \text{ since } \delta > 0 \text{ and } \mu > 0$$

Therefore, $t = ln(1 + \delta)$ *is a minimum for* $f(t)$.

Substituting $t = \ln(1 + \delta)$ *in System (1) gives us our theorem.* □

Definition: 2.2 *The upper tail bound function for the sum of Poisson trials is*

$$\mathbf{Pr}[X > (1 + \delta)\mu] < F^+(\mu, \delta) \triangleq \left[\frac{e^\delta}{(1 + \delta)^{(1+\delta)}}\right]^\mu$$

We can use a similar argument to find the lower tail bound function for the sum of Poisson trials.

Theorem: 2.2 *Let* $X_1, X_2, \cdots, X_n$ *be independent Poisson trials such that, for* $1 \leq i \leq n$, $\mathbf{Pr}[X_i = 1] = p_i$ *with* $0 < p_i < 1$. *Then, for* $X = \sum_{i=1}^{n} X_i$, $\mu = \mathbf{E}[X] = \sum_{i=1}^{n} p_i$, *and for any* $\delta > 0$,

$$\mathbf{Pr}[X < (1 - \delta)\mu] < e^{\frac{-\mu\delta^2}{2}}$$

Proof: *For any positive real number t,*

$$\mathbf{Pr}[X < (1 - \delta)\mu] = \mathbf{Pr}[-X > -(1 - \delta)\mu]$$

$$\mathbf{Pr}[X < (1 - \delta)\mu] = \mathbf{Pr}[e^{-tX} > e^{-t(1-\delta)\mu}]$$

Applying the Markov inequality to the right hand side, we get

$$\mathbf{Pr}[e^{-tX} > e^{-t(1-\delta)\mu}] < \frac{\mathbf{E}[e^{-tX}]}{e^{-t(1-\delta)\mu}}$$

Following the same steps as done in the upper bound proof, the numerator becomes:

$$\mathbf{E}[e^{-tX}] = \mathbf{E}[e^{-t\sum_{i=1}^{n} X_i}] = \mathbf{E}[\prod_{i=1}^{n} e^{-tX_i}] = \prod_{i=1}^{n} \mathbf{E}[e^{-tX_i}]$$

The random variable e^{-tX_i} *has the value of* e^{-t} *with probability* p_i *and the value 1 with probability* $1-p_i$*. Therefore, the expected value is* $p_i(e^{-t}) + (1-p_i)(1)$ *making the numerator:*

$$\prod_{i=1}^{n} (p_i e^{-t} + 1 - p_i) = \prod_{i=1}^{n} \left(1 + p_i(e^{-t} - 1)\right)$$

Using the inequality $1 + x < e^{-x}$ *with* $x = p_i(e^{-t} - 1)$*, we get the right hand side to be:*

$$\mathbf{Pr}[e^{-tX} > e^{-t(1-\delta)\mu}] < \frac{\prod_{i=1}^{n} e^{p_i(e^{-t}-1)}}{e^{-t(1-\delta)\mu}}$$

$$\mathbf{Pr}[e^{-tX} > e^{-t(1-\delta)\mu}] < \frac{e^{\sum_{i=1}^{n} p_i(e^{-t}-1)}}{e^{-t(1-\delta)\mu}}$$

$$\mathbf{Pr}[e^{-tX} > e^{-t(1-\delta)\mu}] < \frac{e^{(e^{-t}-1)\mu}}{e^{-t(1-\delta)\mu}} \tag{2}$$

Once again, we choose t *to make the bound as tight as possible. This means minimizing the righthand side with respect to* t*. The exponent on the RHS of (2) can be rewritten as:* $f(t) = [e^{-t} - 1 + t - t\delta]\mu$*.*

Setting $f'(t) = 0$ *gives:*

$$[-e^{-t} + 1 - \delta]\mu = 0$$

$$\frac{1}{e^t} = 1 - \delta$$

$$\ln\left(\frac{1}{e^t}\right) = \ln(1-\delta)$$

$$t = \ln\left(\frac{1}{(1-\delta)}\right)$$

Also note that

$$f''(t)|_{t=ln(\frac{1}{1-\delta})} = \mu e^{-t}$$

$$= \mu e^{-ln(\frac{1}{1-\delta})}$$

The second derivative is positive since for $0 < \delta < 1$, $\ln(\frac{1}{1-\delta}) > 0$ *and* $\mu > 0$*.*
Therefore, $t = \ln\frac{1}{(1-\delta)}$ *is a minimum for* $f(t)$*.*

Substituting this value t *in System (2) gives*

$$\mathbf{Pr}[X < (1-\delta)\mu] < \left[\frac{e^{-\delta}}{(1-\delta)^{(1-\delta)}}\right]^{\mu}$$

This can be simplified since for $\delta \in (0, 1]$*, we note that* $(1-\delta)^{(1-\delta)} > e^{(-\delta + \frac{\delta^2}{2})}$*. And by using the MacLaurin expansion for* $\ln(1-\delta)$*, we get the result of the theorem (see Appendix §A).*

□

Definition: 2.3
The lower tail bound function for the sum of Poisson trials is

$$\mathbf{Pr}[X < (1-\delta)\mu] < F^{-}(\mu,\delta) \stackrel{\Delta}{=} e^{\frac{-\mu\delta^2}{2}}$$

Example (1): Suppose that sports commentators say that WVU wins each game that they play with a probability of $\frac{1}{3}$. Assuming that the winning the games are independent of each other, what is the upper bound on the probability that they will have a winning season in a season that lasts n games?

We can characterize this event by the following: Let X_i be 1 if WVU wins the i^{th} game and 0 otherwise. Let $Y_n = \sum_{i=1}^{n} X_i$. We want to find $\mathbf{Pr}\left[Y_n > \frac{n}{2}\right]$, that is, what is the probability that more than half the games are won.

Using the fact that the probability of winning one game is $\frac{1}{3}$, we can calculate the expected value of winning n games to be

$$\mathbf{E}[Y] = \frac{n}{3}$$

We now need to rewrite the probability in terms of δ and μ. In other words, we need to rewrite $Y_n > \frac{n}{2}$ to be in the form of $Y_n > (1+\delta)\mu$. So, we just do simple algebra:

$$(1+\delta)\mu = \frac{n}{2}$$

$$(1+\delta)\frac{n}{3} = \frac{n}{2}$$

$$1+\delta = \frac{3}{2}$$

$$\delta = \frac{1}{2}$$

So, now we have the parameters needed and can now plug into the upper bound formula:

$$\mathbf{Pr}[Y_n > \frac{n}{2}] < F^{+}\left(\frac{n}{3},\frac{1}{2}\right) < \left[\frac{e^{\frac{1}{2}}}{\left(1+\frac{1}{2}\right)^{\left(1+\frac{1}{2}\right)}}\right]^{\frac{n}{3}} < (0.965)^n$$

Is this good or bad? As n increases, this number drops off very rapidly. That means that the longer WVU plays the more likely they will have a losing season.

Example (2): Suppose WVU now has a probability of winning each game to be $\frac{3}{4}$. What is the probability that they suffer a losing season assuming that the outcomes of their games are independent of one another?

We can characterize this event by let X_i be 1 if WVU wins the i^{th} game and 0 otherwise. Let $Y_n = \sum_{i=1}^{n} X_i$. We want to find $\mathbf{Pr}\left[Y_n < \frac{n}{2}\right]$. We can use the formula for finding the lower bound by again finding δ and μ.

$$(1-\delta)\mu = \frac{n}{2}$$

$$(1-\delta)\frac{3n}{4} = \frac{n}{2}$$
$$1-\delta = \frac{2}{3}$$
$$\delta = \frac{1}{3}$$

$$\mathbf{Pr}[Y_n < \frac{n}{2}] < F^-\left(\frac{3n}{4}, \frac{1}{3}\right) < e^{\frac{-(\frac{3n}{4})(\frac{1}{3})^2}{2}} < (0.9592)^n$$

This probability is also exponentially small in n.

3 The Chernoff Technique for $-1/1$ trials

Thus far we have studied the sum X of $0/1$ random variables X_i, i.e. each X_i takes on the value 1 with probability p_i and 0, with probability $1-p_i$. We now study the sum X of random variables X_i that assume values of -1 or 1. We will only consider the case, where $p_i = \frac{1}{2}$; the more general case is a simple extension of this case [Mul93].

Theorem: 3.1 *Let $X_1, \cdots, X_n$ be independent, identically distributed random variables taking values 1 and -1 with equal probability of $\frac{1}{2}$. Let $X = X_1 + \cdots + X_n$. Thus, $\mathbf{E}[X] = \mathbf{E}[\sum_{i=1}^{n} X_i] = \sum_{i=1}^{n} \mathbf{E}[X_i] = 0$. Then*

$$\mathbf{Pr}[X > \epsilon n] \leq e^{\frac{-\epsilon^2 n}{2}}, \text{ for any } \epsilon > 0 \tag{3}$$

Proof: *Consider the random variables $e^{\lambda X_i}$, where λ is a real parameter. Since the $X_i's$ are independent, the random variables $e^{\lambda X_i}$ are also independent. Therefore,*

$$\mathbf{E}[e^{\lambda X}] = \mathbf{E}[e^{\lambda(X_1+\cdots+X_n)}] = \mathbf{E}\left[\prod_{i=1}^{n} e^{\lambda X_i}\right] = \prod_{i=1}^{n} \mathbf{E}[e^{\lambda X_i}] \tag{4}$$

By applying Markov's inequality to the nonnegative random variable $e^{\lambda X}$ we get:

$$\mathbf{Pr}[X > x] = \mathbf{Pr}[e^{\lambda X} > e^{\lambda x}] \leq e^{-\lambda x}\mathbf{E}[e^{\lambda X}], \text{ for } \lambda > 0 \text{ and any real number } x$$

Using Equation (4), we get,

$$\mathbf{Pr}[X > x] \leq e^{-\lambda x} \prod_{i=1}^{n} \mathbf{E}[e^{\lambda X_i}], \text{ for } \lambda > 0 \tag{5}$$

For any λ,

$$\mathbf{E}[e^{\lambda X_i}] = \frac{e^{\lambda} + e^{-\lambda}}{2} = \cosh(\lambda) \leq e^{\frac{\lambda^2}{2}}$$

The last inequality follows from the power series expansion for $\cosh(\lambda)$ *(see Appendix §A). Substituting this bound in System (5) with $x = \epsilon n$, we get:*

$$\mathbf{Pr}[X > \epsilon n] \leq e^{-\lambda \epsilon n + \frac{\lambda^2 n}{2}}, \text{ for } \lambda > 0$$

Choosing $\lambda = \epsilon$ give the required bound. □

In identical fashion, we can show that

$$\mathbf{Pr}[X < \epsilon n] \leq e^{\frac{-\epsilon^2 n}{2}} \tag{6}$$

4 Deviation by δ

We now focus on the question "how large does δ need to be such that $\mathbf{Pr}[X > (1+\delta)\mu] \leq \varepsilon$?"

Definition: 4.1 *For any positive μ and ε, $\Delta^+(\mu,\varepsilon)$ is the value of δ that satisfies*

$$F^+(\mu, \Delta^+(\mu,\varepsilon)) = \varepsilon$$

Definition: 4.2 *For any positive μ and ε, $\Delta^-(\mu,\varepsilon)$ is the value of δ that satisfies*

$$F^-(\mu, \Delta^-(\mu,\varepsilon)) = \varepsilon$$

Since $F^-(\mu, \Delta^-(\mu,\varepsilon))$ has the more convenient form, we equate $e^{\frac{-\mu\delta^2}{2}}$ to ϵ, to get:

$$e^{\frac{-\mu\delta^2}{2}} = \epsilon$$
$$e^{\frac{\mu\delta^2}{2}} = \frac{1}{\epsilon}$$
$$\frac{-\mu\delta^2}{2} = \ln\frac{1}{\epsilon}$$
$$\delta = \sqrt{\frac{2ln\frac{1}{\epsilon}}{\mu}}$$

$$\Delta^-(\mu,\varepsilon) = \sqrt{\frac{2\ln\frac{1}{\varepsilon}}{\mu}}$$

Example (3): Consider the sum of n poisson trials, in which $p_i = 0.75$. Suppose that we want to find a δ such that $\mathbf{Pr}[X < (1-\delta)\mu] \leq n^{-5}$.
We can plug into the $\Delta^-(\mu,\varepsilon)$ formula for the result.

$$\mu = \sum p_i = \frac{3n}{4}$$
$$\varepsilon = \frac{1}{n^5}$$
$$\Delta^-\left(\frac{3n}{4}, \frac{1}{n^5}\right) = \sqrt{\frac{10\ln n}{\frac{3n}{4}}} = \sqrt{\frac{13.333\ln n}{n}}$$

This is very small, so it means that we do not need to go very far away from the expectation.

In the next two theorems, we focus on finding convenient formulae for $F^+(\mu,\delta)$.

Theorem: 4.1

$$F^+(\mu,\delta) < \left(\frac{e}{1+\delta}\right)^{(1+\delta)\mu}$$

If we let $\delta > 2e - 1$, then we can rewrite the above as

$$F^+(\mu,\delta) < 2^{-(1+\delta)\mu}$$

Proof: *By definition,*

$$F^+(\mu,\delta) = \left(\frac{e^\delta}{(1+\delta)^{(1+\delta)}}\right)^\mu \tag{7}$$
$$< \left(\frac{e}{1+\delta}\right)^{(1+\delta)\mu} \tag{8}$$

This is because $\delta \geq 0 \Rightarrow e^\delta \geq 1$.

Note that, $\delta > 2e-1 \Rightarrow \frac{e}{1+\delta} < \frac{1}{2}$. *Conseqeuntly, we can rewrite System (7) as:*

$$F^+(\mu,\delta) < 2^{-(1+\delta)\mu}$$

□

We now present the following simplification of $F^+(\mu,\delta)$ for δ in a restricted range $(0,U]$.

Theorem: 4.2 *For* $0 < \delta \leq U$, $F^+(\mu,\delta) \leq e^{-c(U)\mu\delta^2}$, *where* $c(U) = \frac{[(1+U)\ln(1+U)-U]}{U^2}$.

Proof: *See [MR95].* □

For $U = 2e-1$, this simplifies to $F^+(\mu,\delta) < e^{\frac{-\mu\delta^2}{4}}$. Consequently, provided $\delta \leq 2e-1$, we can use the estimate:

$$\Delta^+(\mu,\epsilon) < \sqrt{\frac{4\ln\frac{1}{\epsilon}}{\mu}} \tag{9}$$

Example (4): Suppose we have n *balls and* n *bins. What is the number* m *such that the first bin has* m *balls in it? Each ball is thrown into a bin independently and at random.*

We can express this by letting Y *be a random variable denoting the number of balls in the first bin.*

We want to find m *such that* $\mathbf{Pr}[Y > m] \leq \frac{1}{n^2}$.

If we consider the Bernoulli trails of whether or not the i^{th} *ball falls into the first bin, we have each of these events occurring with* $p_i = \frac{1}{n}$. *Thus, the expected value of the number of balls in the first bin i.e.* $\mathbf{E}[Y] = 1$. *Thus, by applying the Chernoff bound, we have that*

$$\mathbf{Pr}[Y > m] = (1 + \left(\Delta^+(1, \frac{1}{n^2})\right))$$

5 Set Balancing

Before defining the Set Balancing problem, we need the concept of the norm of a vector in $\Re^n$.

Definition: 5.1 *The norm of a vector,*$||\vec{\mathbf{x}}||$, *is a non-negative real number, such that the following conditions are satisfied:*

$$||\vec{\mathbf{x}}|| \geq 0, \ \ ||\vec{\mathbf{x}}|| = 0 \textit{ iff } \vec{\mathbf{x}} = \vec{\mathbf{0}}$$

$$||c\vec{\mathbf{x}}|| = |c| \cdot ||\vec{\mathbf{x}}||$$

$$||\vec{\mathbf{x}} + \vec{\mathbf{y}}|| = ||\vec{\mathbf{x}}|| + ||\vec{\mathbf{y}}||$$

$$||\vec{\mathbf{x}}||_p = \sqrt[p]{\sum_{i=1}^{n} |\vec{\mathbf{x}}_i|^p}$$

If $p = 2$, the norm is the Euclidean norm (our well known formula for distance).

If $p = \infty$, the norm is the $\max_{1 \leq i \leq n} |\vec{\mathbf{x}}_i|$.

We now consider the Set Balancing problem

Given an $n \times n$ matrix A, all of whose entries are either 0 or 1, find a vector $\vec{\mathbf{b}}$, all of whose entries are –1 or 1, such that $\left\|A\vec{\mathbf{b}}\right\|_\infty$ is minimized.

Consider the vector $\vec{\mathbf{b}}$ constructed by Algorithm (5.1).

```
Function SET-BALANCING-SOLUTION
for (i = 1 to n) do
    choose b_i uniformly and at random with probability = 1/2 from {-1,+1}
end for
```

Algorithm 5.1: A Set Balancing Solution Algorithm

We are interested in bounding the value of $||\mathbf{A} \cdot \vec{\mathbf{b}}||_\infty$.

1. Fix any row $\vec{\mathbf{a_i}}$ of $\mathbf{A}$. What is the expected value of $X = \vec{\mathbf{a}}_i\vec{\mathbf{b}}$?

   Let $X_{ij} = a_{ij}.b_j$. Clearly,

$$\mathbf{E}[X] = \mathbf{E}[\vec{\mathbf{a}}_i\vec{\mathbf{b}}] = \mathbf{E}[\sum_{j=1}^{n} a_{ij}b_j] = \mathbf{E}[\sum_{j=1}^{n} X_{ij}] = \sum_{j=1}^{n} \mathbf{E}[a_{ij}b_j] = \sum_{j=1}^{n} \left(\frac{1}{2}a_{ij}(+1) + \frac{1}{2}a_{ij}(-1)\right) = 0$$

2. How much away from 0 does the expected value deviate?

   Observe that X is a sum of random variables $X_{ij}, j = 1, 2, \ldots n$, each of which assumes the value a_{ij} with probability $\frac{1}{2}$ and $-a_{ij}$ with probability $\frac{1}{2}$. We can thus use Theorem (3.1) (in fact, the theorem has to be rederived using $a_{ij}.\lambda$, instead of λ! Note that $\mathbf{E}[e^{a_{ij}.\lambda X_i}] = \frac{e^{a_{ij}.\lambda} + e^{-a_{ij}.\lambda}}{2} \leq \frac{e^{\lambda}+e^{-\lambda}}{2}!$) to show that the

$$\mathbf{Pr}[X > \epsilon n] \leq e^{\frac{-\epsilon^2 n}{2}}, \text{ for any } \epsilon > 0$$

   Substituting $\epsilon n = 4\sqrt{n \ln n}$, we conclude that

$$\mathbf{Pr}[X > 4\sqrt{n \ln n}] \leq n^{-2} \tag{10}$$

   In identical fashion, (along the lines of Theorem (3.1)), we can show that

$$\mathbf{Pr}[X < -4\sqrt{n \ln n}] \leq n^{-2} \tag{11}$$

   We can therefore conclude that the probability of the absolute value of $\vec{\mathbf{a_i}}.\vec{\mathbf{b}}$ exceeding $4\sqrt{n \ln n}$ is at most $\frac{2}{n^2}$, for fixed $\vec{\mathbf{a_i}}$. It follows that the probability that the absolute value of $\vec{\mathbf{a_i}}.\vec{\mathbf{b}}$ exceeding $4\sqrt{n \ln n}$ over *any row* $\vec{\mathbf{a_i}}$ is at most $\frac{2}{n}$. Therefore, with probability at least $1 - \frac{2}{n}$, we have $||\mathbf{A}.\vec{\mathbf{b}}||_\infty \leq 4\sqrt{n \ln n}$, for the randomly chosen vector $\vec{\mathbf{b}}$.

A Series Expansions

The Taylor series expansion for a function $f(x)$ about $x = a$ is given by :

$$f(x) = f(a) + \frac{f'(a)}{1!} \cdot (x-a) + \frac{f''(a)}{2!} \cdot (x-a)^2 + \cdots + \frac{f^{(n)}(a)}{n!} \cdot (x-a)^n + \cdots$$

The MacLaurin series expansion of a function $f(x)$ is the Taylor series expansion with $a = 0$:

$$f(x) = f(0) + \frac{f'(0)}{1!} \cdot (x) + \frac{f''(0)}{2!} \cdot (x)^2 + \cdots + \frac{f^{(n)}(0)}{n!} \cdot (x)^n + \cdots$$

The power series expansion for $\cosh(x) = \frac{e^x + e^{-x}}{2}$ is:

$$\cosh(x) = 1 + \frac{x^2}{2!} + \frac{x^4}{4!} + \cdots$$

(Check it!)

References

[MR95] Rajeev Motwani and Prabhakar Raghavan. *Randomized Algorithms.* Cambridge University Press, Cambridge, England, June 1995.

[Mul93] K. Mulmuley. *Computational Geometry, an Introduction through Randomized Algorithms.* Prentice Hall, 1993.