Moments and Deviations

Junhui Jia
{jia@csee.wvu.edu}

1 Introduction

We have bounded the expected running times of several randomized algorithms in first two chapters. While the expectation of a random variable (such as a running time) may be small, it may frequently assume values that are far higher. In analyzing the performance of a randomized algorithm, we often like to show that the behavior of the algorithm is good almost all the time. For example, it is more desirable to show that the running time is small with high probability, not just that it has a small expectation. The similar statement: for randomized algorithms, usually knowing the bound of expected running time is not enough. It is more desirable to show that the expected running time is small with high probability. To prove this statement, we will begin by examining a family of stochastic processes that is fundamental to the analysis of many types of randomized algorithms. They are Occupancy Problems. This motivates the study of general bounds on the probability that a random variable deviates far from its expectation, enabling us to avoid such custom-made analysis. The probability that a random variable deviates by a given amount from its expectation is referred to as a tail probability for that deviation.

2 Occupancy Problems

We begin with an example of an occupancy problem. In such problems we envision each of m indistinguishable objects ("balls") being randomly assigned to one of n distinct classes ("bins"). In other words, each ball is placed in a bin chosen independently and uniformly at random. We are interested in questions such as: what is the maximum number of balls in any bin? what is the expected number of bins with k balls in them?

A Simple Example of Occupancy Problem

1. We have m **indistinguishable** balls.
2. We have n **distinct** bins.
3. We throw those m balls independently, uniformly into those n bins.

   Questions:

   1. What is the **expected number** of balls in a bin?
   2. What is the **expected number** of bins with k balls in each of it?

Such problem are at the core of the analysis of many randomized algorithms ranging from data structures to routing in parallel computers. Later, we will encounter a variant of the occupancy problem, known as the coupon collector's problem; we also will apply sophisticated techniques to various random variables arising in occupancy problem. Our discussion of the occupancy problem will illustrate a recurrent tool in the analysis of randomized algorithms: that the probability of the union of events is no more than the sum of their probabilities, the following give this theorem and the definition of tail probability.

Definition: 2.1 *The probability that a random variable deviates from its expectation is referred to as the* **tail probability** *of that deviation.*

Theorem: 2.1 *The probability of the Union of events is no more than the sum of their probabilities.*

$$Pr[\bigcup_{i=1}^{n} E_i] \leq \sum_{i=1}^{n} Pr[E_i]$$

We will need Theorem (2.1) often in this portion of the occupancy problems. The proof of Theorem (2.1) is given in Appendix A.

Now, let's consider the case $m = n$. For $1 \leq i \leq n$, let X be the number of balls in the ith bin. Let us try to make a statement with very high probability, no bins receives more than k balls, for a chosen k. Let $E_j(k)$ denote the event that bin j has k or more balls in it. The probability that bin j receives exactly i balls is

$$\binom{n}{i}\left(\frac{1}{n}\right)^i\left(1-\frac{1}{n}\right)^{n-i} \leq \binom{n}{i}\left(\frac{1}{n}\right)^i \leq \left(\frac{ne}{i}\right)^i\left(\frac{1}{n}\right)^i = \left(\frac{e}{i}\right)^i$$

Thus,

$$Pr[E_j(k)] \leq \sum_{i=k}^{n}\left(\frac{e}{i}\right)^i \leq \left(\frac{e}{k}\right)^k\left(1+\frac{e}{k}+\left(\frac{e}{k}\right)^2+\cdots\right)$$

Let $k^* = \lceil (e \lg n)/\lg\lg n \rceil$. Substitute k^* into the above equation

$$Pr[E_j(k^*)] \leq \left(\frac{e}{k^*}\right)^{k^*}\left(1+\frac{e}{k^*}+\left(\frac{e}{k^*}\right)^2+\cdots\right)$$

That is can be simplified as

$$Pr[E_j(k^*)] \leq \left(\frac{e}{k^*}\right)^{k^*}\frac{1}{1-e/k^*}$$

Using $k^* = \lceil (e \lg n)/\lg\lg n \rceil$, we get

$$Pr[E_j(k^*)] \leq n^{-2}$$

We invoke the theorem (2.1). Then we have

$$Pr[\bigcup_{j=1}^{n} E_j(k^*)] \leq \sum_{j=1}^{n} Pr[E_j(k^*)] \leq \frac{1}{n}$$

Thus we established

Theorem: 2.2 *With probability at least* $1 - 1/n$*, no bin has more than* $k^* = (e\ln n)/\ln\ln n$ *balls in it.*

Suppose that m balls are randomly assigned to n bins. We study the probability of the event that they all land in distinct bins. Consider the assignment of the balls to bins as a sequential process: we throw the first ball into a random bin, then next ball, and so on. For $2 \leq i \leq m$, let E_i denote the event that the *ith* ball lands in a bin not containing any of the first $i-1$ balls. Note that $Pr[E_1] = 1$. From the probability of the intersection for a collection of events

$$Pr[\bigcap_{i=1}^{k} E_i] = Pr[E_1] \times Pr[E_2|E_1] \times Pr[E_3|E_1 \cap E_2] \cdots Pr[E_k| \cap_{i=1}^{k-1} E_i]$$

We have

$$Pr[\bigcap_{i=2}^{m} E_i] = Pr[E_2]Pr[E_3|E_2]Pr[E_4|E_2 \cap E_3] \cdots Pr[E_m| \cap_{i=2}^{m-1} E_i]$$

The probability that *ith* ball lands in an empty bin given that the first $i-1$ balls fell into distinct bins is

$$Pr[E_i | \cap_{j=2}^{i-1} E_j] = 1 - \frac{i-1}{n}$$

Making use of the fact that

$$1 - x \leq e^{-x}$$

We have

$$Pr[\bigcap_{i=2}^{m} E_i] \leq \prod_{i=2}^{m} \left(1 - \frac{i-1}{n}\right) \leq \prod_{i=2}^{m} e^{-(i-1)/n} = e^{-m(m-1)/2n}$$

Corollary: 2.1 *For $m = \lceil\sqrt{2n} + 1\rceil$, the probability that m balls land in distinct bins $\leq \frac{1}{e}$*

When m increases beyond this value, the probability drops rapidly. A special case that is popular in mathematics is the birthday problem. The 365 days of the year (ignoring leap years) correspond to 365 bins, and the birthday of each of m people is chosen independently and uniformly from 365 days. How large must m be before two people in the same room are likely to share their birthdays? From the Corollary 1.1,

$$m = \lceil\sqrt{2n} + 1\rceil = \lceil\sqrt{2 \times 365} + 1\rceil = 28$$

Thus, when more than 28 people are in the same room, the probability of two people likely to share their birthdays is $(1 - \frac{1}{e})$.

3 The Markov and Chebyshev Inequalities

Theorem: 3.1 **(Markov Inequality):** Let Y be a random variable assuming only non-negative values. Then for all $t \in R^+$,

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

Proof: *Let*

$$f(y) = 1, \; if \; y \geq t;$$

$$f(y) = 0, \; otherwise.$$

For all y, it satisfies

$$f(y) \leq \frac{y}{t}$$

Then we have

$$Pr[Y \geq t] = E[f(Y)] \leq E\left[\frac{Y}{t}\right] = \frac{E[Y]}{t}$$

□

Corollary: 3.1 *Let Y be a random variable assuming only non-negative values. Then for all $t \in R^+$,*

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

Proof: *Apply Theorem 3.1*

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

with

$$t = kE[Y]$$

This gives

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

□

Theorem: 3.2 **(Chebyshev's Inequality):** Let X be a random variable with expectation μ_X and standard deviation σ_X. Then for all $t \in R^+$,

$$Pr[|X - \mu_X| \geq t\sigma_X] \leq \frac{1}{t^2}$$

Proof:
First note that

$$|X - \mu_X| \geq t\sigma_X$$

$$(X - \mu_X)^2 \geq t^2\sigma_X^2$$

The random variable

$$Y = (X - \mu_X)^2$$

has expectation

$$E(Y) = \sigma_X^2$$

Note that $Y \geq 0$, then from Markov Inequality, we have

$$Pr[Y \geq t^2\sigma_X^2] \leq \frac{E[Y]}{t^2\sigma_X^2} \leq \frac{\sigma_X^2}{t^2\sigma_X^2} = \frac{1}{t^2}$$

□

4 Randomized Selection

Consider the problem of selecting the k^{th} smallest element in a set S of n element. We assume that the elements of S are distinct. Let $r_S(t)$ denote the rank of an element t (the k^{th} smallest element has rank k) and let $S_{(i)}$ denote the i^{th} smallest element of S. Thus the problem becomes that we seek to identify $S_{(k)}$. LazySelect algorithm is introduced. some important properties of independent random variables in order to perform the analysis of LazySelect algorithm are given in Appendix A .

Thus the idea of the algorithm is to identify two elements a and b in S such that both of the following statements hold with high probability:

1. The element $S_{(k)}$ that we seek is in P.
2. The set P of elements between a and b is not very large, so that we can sort P inexpensively in step 5.

Algorithm LazySelect:

Input: A set S of n elements, and an integer k in $[1, n]$.
Output: The k^{th} smallest element of S, $S_{(k)}$.

1: Pick $n^{3/4}$ elements from S, chosen independently and uniformly at random with replacement; call this multiset of elements R.
2: Sort R in $O(n^{3/4} \log n)$ steps using any optimal sorting algorithm.
3: Let $x = kn^{-1/4}$. For $l = max\{\lfloor x - \sqrt{n} \rfloor, 1\}$ and $h = min\{\lceil x + \sqrt{n} \rceil, n^{3/4}\}$, let $a = R_{(l)}$ and $b = R_{(h)}$. By comparing a and b to every element of S, determine $r_S(a)$ and $r_S(b)$.
4: **if** $k < n^{1/4}$, **then** $P = \{y \in S | y \leq b\}$ noted as P_b;
else if $k > n - n^{1/4}$, let $P = \{y \in S | y \geq a\}$ noted as P_a;
else if $k \in [n^{1/4}, n - n^{1/4}]$, let $P = \{y \in S | a \leq y \leq b\}$ noted as P_{ab};
Check whether $S_{(k)} \in P$ and $|P| \leq 4n^{3/4} + 2$. If not, repeat Steps $1 - 3$ until such a set P is found.
5: By sorting P in $O(|P| \log |P|)$ steps, identify $P_{(k - r_S(a) + 1)}$, which is $S_{(k)}$.

Algorithm 4.1: LazySelect Algorithm

Theorem: 4.1 *With probability $1 - O(n^{-1/4})$, LazySelect finds $S_{(k)}$ on the first pass through Steps 1-5. The running time of LazySelect algorithm is $2n + o(n)$.*

Proof:

The time bound is easily established by examining the algorithm; Step 3 requires $2n$ comparisons, and all other steps perform $o(n)$ comparisons, provided the algorithm finds $S_{(k)}$ on the first pass through Steps $1-5$. We measure the running time of LazySelect algorithm in terms of the number of comparisons performed on it, therefore, the running time of LazySelect algorithm is $2n + o(n)$.

There are three possible ways in which P is chosen

1. *P_a, that is $P = \{y \in S | y \geq a\}$ for $k > n - n^{1/4}$.*
2. *P_b, that is $P = \{y \in S | y \leq b\}$ for $k < n^{1/4}$.*
3. *P_{ab}, that is $P = \{y \in S | a \leq y \leq b\}$ for $k \in [n^{1/4}, n - n^{1/4}]$.*

For each case, there are two possibilities in which the algorithm fails:

1. *The element $S_{(k)}$ that we seek is not in P.*
2. *P is too big.*

Case P_{ab}:

The first possibility is that the element $S_{(k)}$ that we seek is not in P. We now consider the mode of failure: $a > S_{(k)}$ because fewer than l of the samples in R are less than or equal to $S_{(k)}$ (so that $S_{(k)} \notin P$). Set

$$X_i = 1, \ if \ R_{(i)} \leq S_{(k)},$$

$$X_i = 0, \ otherwise.$$

Thus

$$Pr[X_i = 1] = k/n$$

and

$$Pr[X_i = 0] = 1 - k/n$$

Let

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

be the number of samples of R, that are at most S_k. Note that we really do mean the number of samples, and not the number of distinct elements. The random variables X_i are Bernoulli random variables. Then the expectation and the variance of a Bernoulli random variable with success probability p obtained by using the theorems in Appendix A.

$$\mu_X = \sum_{i=1}^{n^3/4} \mu_{X_i} = \frac{k}{n}(n^{3/4}) = kn^{-1/4}$$

$$\sigma_X^2 = \sum_{i=1}^{n^3/4} \sigma_{X_i}^2 = n^{3/4}\left(\frac{k}{n}\right)\left(1 - \frac{k}{n}\right)$$

Where let $f(k) = \left(\frac{k}{n}\right)\left(1 - \frac{k}{n}\right)$

$$f(k)' = \frac{1}{n} - \frac{2k}{n^2} = 0$$

$$k = \frac{n}{2}$$

$$f(k)'' = -\frac{2}{n^2} < 0$$

So when $k = \frac{n}{2}$, $f(k)$ has maximum value $\frac{1}{4}$, therefore

$$\sigma_X^2 \leq \frac{n^{3/4}}{4}$$

Since

$$l = max\{\lfloor x - \sqrt{n} \rfloor, 1\}$$
$$X < l$$

Then we have

$$X < x - \sqrt{n}$$
$$X - x < -\sqrt{n}$$
$$\mid X - x \mid \geq \sqrt{n}$$

The probability of the above is

$$Pr[\mid X - x \mid \geq \sqrt{n}] = Pr[\mid X - \mu_X \mid \geq \sqrt{n}]$$

Apply the Chebyshev bound to X and $\sigma_X \leq n^{3/8}/2$

$$Pr[\mid X - \mu_X \mid \geq \sqrt{n}] \leq Pr[\mid X - \mu_X \mid \geq 2n^{1/8}\sigma_X] = O(n^{-1/4})$$

An essentially identical argument shows that

$$Pr[b < S_k] = O(n^{-1/4})$$

Since the probability of the union of events is at most the sum of their probabilities, the probability that either of these events occurs (causing $S_{(k)}$ to lie outside P) is $O(n^{-1/4})$

The second possibility of failure occurs when P is too big. To study this, we define $k_l = max\{1, k - 2n^{3/4}\}$ and $k_h = min\{k + 2n^{3/4}, n\}$. To obtain an upper bound on the probability of this kind of failure, we will be pessimistic and say that failure occurs if either $a < S_{kl}$ or $b > S_{kh}$. The analysis is very similar to that above in studying the first mode of failure, with k_i and k_h playing the role of k. For $k \in [n^{1/4}, n - n^{1/4}]$ and $P = \{y \in S | a \leq y \leq b\}$ (that is P_{ab}); let us show

$$Pr[a < S_{kl}] = O(n^{-1/4})$$

Set

$$X_i = 1,\ if\ R_{(i)} \leq S_{(kl)},$$

$$X_i = 0,\ otherwise.$$

Thus

$$Pr[X_i = 1] = k/n$$

and

$$Pr[X_i = 0] = 1 - k/n$$

Let

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

be the number of samples of R, that are at most S_{kl}. The random variables X_i are Bernoulli random variables. Then the expectation and the variance of a Bernoulli random variable with success probability p

$$\mu_X = \frac{kn^{3/4}}{n} = kn^{-1/4}$$

$$\sigma_X^2 = n^{3/4}\left(\frac{k}{n}\right)\left(1 - \frac{k}{n}\right) \leq \frac{n^{3/4}}{4}$$

This implies that $\sigma_X \leq n^{3/8}/2$. Applying the Chebyshev bound to X,

$$Pr[\mid X - \mu_X \mid \geq \sqrt{n}] \leq Pr[\mid X - \mu_X \mid \geq 2n^{1/8}\sigma_X] = O(n^{-1/4})$$

Let us show

$$Pr[b > S_{kh}] = O(n^{-1/4})$$

Set

$$X_i = 1,\ if\ R_{(i)} \leq S_{(kh)},$$

$$X_i = 0,\ otherwise.$$

Thus

$$Pr[X_i = 1] = k/n$$

and

$$Pr[X_i = 0] = 1 - k/n$$

Let

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

be the number of samples of R, that are at most S_{kh}. The random variables X_i are Bernoulli random variables. Then the expectation and the variance of a Bernoulli random variable with success probability p

$$\mu_X = \frac{kn^{3/4}}{n} = kn^{-1/4}$$

$$\sigma_X^2 = n^{3/4}\left(\frac{k}{n}\right)\left(1 - \frac{k}{n}\right) \leq \frac{n^{3/4}}{4}$$

This implies that $\sigma_X \leq n^{3/8}/2$. Applying the Chebyshev bound to X, we have

$$Pr[| X - \mu_X | \geq \sqrt{n}] \leq Pr[| X - \mu_X | \geq 2n^{1/8}\sigma_X] = O(n^{-1/4})$$

Case P_b:

For $k < n^{1/4}$ and $P = \{y \in S | y \leq b\}$ (that is P_b), The first possibility is that the element $S_{(k)}$ that we seek is not in P, that is $b < S_{(k)}$. An essentially identical arguments to Case P_{ab} shows that

$$Pr[b < S_k] = O(n^{-1/4})$$

The second possibility of failure occurs when P is too big. let's show

$$Pr[a < S_{kl}] = O(n^{-1/4})$$

Set

$$X_i = 1,\ if\ R_{(i)} \leq S_{(kl)},$$

$$X_i = 0,\ otherwise.$$

Thus

$$Pr[X_i = 1] = k/n$$

and

$$Pr[X_i = 0] = 1 - k/n$$

Let

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

be the number of samples of R, that are at most S_{kl}. The random variables X_i are Bernoulli random variables. Then the expectation and the variance of a Bernoulli random variable with success probability p

$$\mu_X = \frac{kn^{3/4}}{n} = kn^{-1/4}$$

$$\sigma_X^2 = n^{3/4}\left(\frac{k}{n}\right)\left(1 - \frac{k}{n}\right) \leq \frac{n^{3/4}}{4}$$

This implies that $\sigma_X \leq n^{3/8}/2$. Applying the Chebyshev bound to X,

$$Pr[| X - \mu_X |\geq \sqrt{n}] \leq Pr[| X - \mu_X |\geq 2n^{1/8}\sigma_X] = O(n^{-1/4})$$

Let us show

$$Pr[b > S_{kh}] = O(n^{-1/4})$$

Set

$$X_i = 1,\ if\ R_{(i)} \leq S_{(kh)},$$

$$X_i = 0,\ otherwise.$$

Thus

$$Pr[X_i = 1] = k/n$$

and

$$Pr[X_i = 0] = 1 - k/n$$

Let

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

be the number of samples of R*, that are at most* S_{kh}*. The random variables* X_i *are Bernoulli random variables. Then the expectation and the variance of a Bernoulli random variable with success probability* p

$$\mu_X = \frac{kn^{3/4}}{n} = kn^{-1/4}$$

$$\sigma_X^2 = n^{3/4}\left(\frac{k}{n}\right)\left(1 - \frac{k}{n}\right) \leq \frac{n^{3/4}}{4}$$

This implies that $\sigma_X \leq n^{3/8}/2$*. Applying the Chebyshev bound to* X*, we have*

$$Pr[| X - \mu_X |\geq \sqrt{n}] \leq Pr[| X - \mu_X |\geq 2n^{1/8}\sigma_X] = O(n^{-1/4})$$

Case P_a*:*

For $k > n - n^{1/4}$ *and* $P = \{y \in S | y \geq a\}$ *(that is* P_a*); The first possibility is that the element* $S_{(k)}$ *that we seek is not in* P*, that is* $a > S_{(k)}$*. An essentially identical arguments to Case* P_{ab} *shows that*

$$Pr[a > S_k] = O(n^{-1/4})$$

The second possibility of failure occurs when P *is too big. let's show*

$$Pr[a < S_{kl}] = O(n^{-1/4})$$

Set

$$X_i = 1,\ if\ R_{(i)} \leq S_{(kl)},$$

$$X_i = 0,\ otherwise.$$

Thus

$$Pr[X_i = 1] = k/n$$

and

$$Pr[X_i = 0] = 1 - k/n$$

Let

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

be the number of samples of R*, that are at most* S_{kl}*. The random variables* X_i *are Bernoulli random variables. Then the expectation and the variance of a Bernoulli random variable with success probability* p

$$\mu_X = \frac{kn^{3/4}}{n} = kn^{-1/4}$$

$$\sigma_X^2 = n^{3/4}\left(\frac{k}{n}\right)\left(1 - \frac{k}{n}\right) \leq \frac{n^{3/4}}{4}$$

This implies that $\sigma_X \leq n^{3/8}/2$*. Applying the Chebyshev bound to* X*,*

$$Pr[\mid X - \mu_X \mid \geq \sqrt{n}] \leq Pr[\mid X - \mu_X \mid \geq 2n^{1/8}\sigma_X] = O(n^{-1/4})$$

Let us show

$$Pr[b > S_{kh}] = O(n^{-1/4})$$

Set

$$X_i = 1,\ if\ R_{(i)} \leq S_{(kh)},$$

$$X_i = 0,\ otherwise.$$

Thus

$$Pr[X_i = 1] = k/n$$

and

$$Pr[X_i = 0] = 1 - k/n$$

Let

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

be the number of samples of R*, that are at most* S_{kh}*. The random variables* X_i *are Bernoulli random variables. Then the expectation and the variance of a Bernoulli random variable with success probability* p

$$\mu_X = \frac{kn^{3/4}}{n} = kn^{-1/4}$$

$$\sigma_X^2 = n^{3/4}\left(\frac{k}{n}\right)\left(1 - \frac{k}{n}\right) \leq \frac{n^{3/4}}{4}$$

This implies that $\sigma_X \leq n^{3/8}/2$*. Applying the Chebyshev bound to* X*, we have*

$$Pr[\mid X - \mu_X \mid \geq \sqrt{n}] \leq Pr[\mid X - \mu_X \mid \geq 2n^{1/8}\sigma_X] = O(n^{-1/4})$$

Adding up the probability of all of these failure modes, we find that the probability that LazySelect algorithm fail to find a suitable set P *is* $O(n^{-1/4})$

□

A Some important theorems

Some important properties of independent random variables in order to perform the analysis of LazySelect algorithm are given in Appendix A.

Definition: A.1 *Set X and Y be two random variables defined on the sample space. The* **joint distribution** *of X and Y is given by*

$$Pr[x, y] = Pr[X = x, Y = y]$$

Theorem: A.1 *The random variable X and Y are independent if*

$$Pr[X = x, Y = y] = Pr[X = x]Pr[Y = y]$$

Theorem: A.2 *If X and Y are the independent random variable, then*

$$E[XY] = E[X]E[Y]$$

Theorem: A.3 *Let X_1, X_2, $\cdots X_m$ be the independent random variables, and $X = \sum_{i=1}^{m} X_i$. Then*

$$\sigma_X^2 = \sum_{i=1}^{m} \sigma_{X_i}^2$$

Proof:
Let μ_i denote $E[X_i]$, and $\mu = \sum_{i=1}^{m} \mu_i$. The variance of X is given by

$$E[(X - \mu)^2] = E[(\sum_{i=1}^{m}(X_i - \mu_i))^2]$$

Expanding the latter and using linearity of expectations, we obtain

$$E[(X - \mu)^2] = \sum_{i=1}^{m} E[(X_i - \mu_i)^2] + 2\sum_{i<j} E[(X_i - \mu_i)(X_j - \mu_j)]$$

Since all pairs of X_i, and X_j are independent, so are the pairs $(X_i - \mu_i)$, $(X_j - \mu_j)$. Each term in the latter summation can be replaced by $E[(X_i - \mu_i)]E[(X_j - \mu_j)]$. Since $E[(X_i - \mu_i)] = E[X_i] - \mu_i = 0$, the latter summation vanishes. It follows that

$$E[(X - \mu)^2] = \sum_{i=1}^{m} E[(X_i - \mu_i)^2] = \sum_{i=1}^{m} \sigma_{X_i}^2$$

□

The proof of Theorem (2.1)
Proof: *Base case $n = 2$*

$$P(E_1 \cup E_2) = P(E_1) + P(E_2) - P(E_1E_2) \leq P(E_1) + P(E_2)$$

That is

$$Pr[\bigcup_{i=1}^{n=2} E_i] \leq \sum_{i=1}^{n=2} Pr[E_i]$$

Suppose for $n \leq k$, it satisfies

$$Pr[\bigcup_{i=1}^{k} E_i] \leq \sum_{i=1}^{k} Pr[E_i]$$

For $n = k + 1$, *Then we have*

$$Pr[\bigcup_{i=1}^{k+1} E_i] = Pr[\bigcup_{i=1}^{k} E_i \cup E_{k+1}] \leq Pr[\bigcup_{i=1}^{k} E_i] + Pr[E_{k+1}]$$

Since

$$Pr[\bigcup_{i=1}^{k} E_i] \leq \sum_{i=1}^{k} Pr[E_i]$$

Therefore

$$Pr[\bigcup_{i=1}^{k+1} E_i] \leq \sum_{i=1}^{k} Pr[E_i] + Pr[E_{k+1}] \leq \sum_{i=1}^{k+1} Pr[E_i]$$

Thereom is proved. □

References

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