Tail Estimates and High-probability Performance Bounds in Computational Geometry

Ying Xiao
Departmant of Mathematics and Statistics
Carleton University

Abstract

We will discuss some tail estimates from probability theory. They are used quite often in computational geometry, especially in proving high-probability performance bounds.

1 Some Definitions in Probability Theory

We begin with some basic definitions in probability theory.

Random Variable(intuitive definition): a function which takes its value by chance.

If the possible value set of a random variable X is finite or countably infinite(for example: all non-negative interger N={0,1,2,...}), then X is called a **discrete** random variable. All the random variables we will discuss in our course are discrete random variables.

The Expected Value: If X is a discrete random variable, the possible value set of X is N={0,1,2,...}, then its expected value $E[X]$ is defines as:

$$E[X] = \sum_{k=0}^{\infty} kP[X = k]$$

Intuitively, $E[X]$ is some sort of average value that X takes, because of this reason, we also call $E[X]$ the **mean** of X. For continuous random variable, we need use integration to define its expected value.

The Variance: Let X be a random variable, its expected value is $\mu = E[X]$. The Variance σ^2 of X is defined as the expected value of the random variable $(X - \mu)^2$. i.e.

$$\sigma^2 = E[(X - \mu)^2]$$

Intuitively, the variance σ^2 gives some idea of how far the random variable X derivates from its expected value μ.

The Standard Deviation: The Standard Deviation σ is defined as:

$$\sigma = \sqrt{\sigma^2} = \sqrt{E[(X-\mu)^2]}$$

What we are really interested in?

Bounding the probability that a discrete random variable X deviates from its expected value($E[X]$) by a given amount. More precisely, if $\mu = E[X]$, we might ask "what is the probability that X exceeds $(1+\epsilon)\mu$?", where $\epsilon > 0$ is a constant.

The following Markov's inequality is the simplest bound:

Theorem 1(Markov's Inequality): For any nonnegative random variable X and $\forall \alpha \geq 0$,

$$P[X \geq \alpha] \leq \frac{E[X]}{\alpha} \tag{1}$$

Proof We can easily prove this inequality by using the definition of $E[X]$ in the case X is a discrete random variable.

Markov's inequality only give a bound for the probability that X bigger than some value. It gives us no idea about the probability that X deviates from its expected value $E[X]$ by a given amount. The following Chebychev's inequality can do this job.

Theorem 2(Chebychev's Inequality): Let $\mu = E[X]$ be the expected value of X and σ be the standard deviation of X, then for $\forall t > 0$, we have

$$P[|X-\mu| \geq t\sigma] \leq \frac{1}{t^2} \tag{2}$$

Proof Using Markov's inequality to nonnegative random variable $(X-\mu)^2$.

2 Chernoff's Technique

We are looking for a much better bound than Markov's and Chebychev's inequalities. If the random variable X can be expressed as the summation of some independent random variables $X_i (i = 1, 2, \ldots, n)$, i.e. $X = \Sigma_{i=1}^{n} X_i$, then we can get a much better bound by using so-called Chernoff's Technique, but in order to do that, we need to define the independence of the random variables first:

Independence of the Random Variables: The discrete random variables X_1 and X_2 are said to be independent if for any x_1, x_2,

$$P[X_1 = x_1, X_2 = x_2] = P[X_1 = x_1]P[X_2 = x_2]$$

In general , we say $X_1, X_2, ...X_n$ to be independent if for any x_i(i=1,2,...n)

$$P[X_1 = x_1, X_2 = x_2, ..., X_n = x_n] = P[X_1 = x_1]P[X_2 = x_2] \cdots P[X_n = x_n]$$

Some properties of the independent random variables:

(i) If X_i's are independent, then $e^{\lambda X_i}$'s are also independent.

(ii) If X_i's are independent, then

$$E[X_1 X_2 \cdots X_n] = E[X_1]E[X_2] \cdots E[X_n]$$

Let the random variable X can be expressed as the summation of some independent random variables $X_i (i = 1, 2, ..., n)$, i.e. $X = \Sigma_{i=1}^n X_i$, then **the basic idea of Chernoff's Technique** is as following:

(1) Instead consider X_i, we consider random variables $e^{\lambda X_i}$, since X_i's are independent , then those $e^{\lambda X_i}$'s are also independent, hence we have:

$$E[e^{\lambda X}] = E[e^{\lambda(X_1 + X_2 + ... + X_n)}] = E[\Pi_{i=1}^n e^{\lambda X_i}] = \Pi_{i=1}^n E[e^{\lambda X_i}] \qquad (3)$$

Notice we already use properties(ii) here.

(2) Using Markov's Inequality to the random variable $e^{\lambda X}$. For any x and any $\lambda > 0$, $P[X > x] = P[e^{\lambda X} > e^{\lambda x}] \leq e^{-\lambda x} E[e^{\lambda X}]$, hence

$$P[X > x] \leq e^{-\lambda x} E[e^{\lambda X}] = e^{-\lambda x} \Pi_{i=1}^n E[e^{\lambda X_i}] \qquad (4)$$

Actually, we do not really need the full independence among all X_i's, all we need is that equality(3) holds. It is enough if X_i is independent of all X_j(for any $j > i$), this condition is hold for all applications in our course.

3 Some Examples

Example 1: Let $X_1, X_2, ...X_n$ be independent identically distributed random variables taking 1 and -1 with equal probability(i.e. $P[X_i = 1] = P[X_i = -1] = 1/2$ for $i = 1, 2, ..., n$). Here we say identically distributed means $P[X_1 = a] = P[X_2 = a] = ... = P[X_n = a]$ for all possible value a. Let $X = X_1 + X_2 + ... + X_n$, then

$$E[X] = E[X_1 + X_2 + ... + X_n] = E[X_1] + E[X_2] + ... + E[X_n]$$

Since $E[X_i] = 1 \times 1/2 + (-1) \times 1/2 = 0$ for i=1,2,...n, we have $E[X] = 0$. The chernoff technique provides a very good bound on the probability that X takes values at a distance ϵn away from its expected value 0.

Theorem 3 Let X be defined as before, then for any $\epsilon > 0$, we have $P[X > \epsilon n] \leq e^{-\frac{\epsilon^2 n}{2}}$.

Proof For any $\lambda \geq 0$, using (4), we have

$$P[X > \epsilon n] \leq e^{-\epsilon n \lambda} \Pi_{i=1}^{n} E[e^{\lambda X_i}]$$

while $E[e^{\lambda X_i}] = e^{\lambda \times 1} P[X_i = 1] + e^{\lambda \times (-1)} P[X_i = -1] = \frac{e^{\lambda} + e^{-\lambda}}{2} \leq e^{\frac{\lambda^2}{2}}$, the last inequality can be shown by using taylor series of the corresponding functions. Thus

$$P[X > \epsilon n] \leq e^{-\epsilon n \lambda} \times (e^{\frac{\lambda^2}{2}})^n = e^{\frac{\lambda^2 n}{2} - \epsilon n \lambda}$$

By choosing $\lambda = \epsilon$, we prove the theorem.

If $X_1, X_2, ...X_n$ be independent identically distributed random variables taking 1 and 0 with equal probability(i.e. $P[X_i = 1] = P[X_i = 0] = 1/2$), then the corresponding result is the following:

Corollary 4 Let $X_1, X_2, ...X_n$ be independent identically distributed random variables taking 1 and 0 with equal probability, $X = X_1 + X_2 + ... + X_n$, then for any $\epsilon > 0$, $P[X - E[X] > \epsilon E[X]] \leq e^{-\frac{\epsilon^2 n}{2}}$.

Proof Let $Z_i = 2X_i - 1 (i = 1, 2, ..., n)$, then $Z = \sum_{i=1}^{n} Z_i = 2X - n$, $P[Z_i = 1] = P[Z_i = -1] = 1/2$, using theorem 1, we can prove the corollary easily.

Example 2: Recall what we have done in last class, using randomized binary search tree $\tilde{\mathrm{H}}(\mathrm{N})$ to locate any query point $p \in R$ in H(N) quickly. Fixed $p \in R$, using backward analysis. Starting with $H(N) = H(N^n)$, imagine deleting the points in N, one at a time, in the reserve order $S_n, S_{n-1}, ..., S_1$, this gives rise to a sequence of partitions $H(N^n), H(N^{n-1}), ..., H(N^0)$.

For $1 \leq j \leq n$, define **0-1 random variable** V_j such that $V_j = 1$ iff the interval in $H(N^j)$ containing p is adjacent to the deleted point S_j, otherwise it is 0. The interval is destroyed during the imaginary deletion of S_j from $H(N^j)$.

The length of the search path is $V = \sum_{j=1}^{n} V_j$. Due to the randomized construction of our search structure, we have $P[V_j = 1] \leq 2/j$(at most two points in N^j are adjacent to the interval in $H(N^j)$ containing p). Hence

$E[V_j] \leq 2/j$. Thus

$$E[V] = E[V_1 + V_2 + ... + V_n] \leq \sum_{j=1}^{n} \frac{2}{j} = O(\log n)$$

Hence for fixed query point p, the expected length of the search path is $O(\log n)$. Using Chernoff Technique, We can do more than that. Actually, we can show the length of the search path is $\tilde{\mathrm{O}}(\log n)$, i.e. the logrithmic bound holds with high probability.

Notice $V = \sum_{i=1}^{n} V_i$ and each V_i is independent of V_j for $j > i$, hence we can use (4). In order to show the length of the search is $\tilde{\mathrm{O}}(\log n)$, we will prove the following theorem:

Theorem 5 Let $H_n = \sum_{i=1}^{n} \frac{1}{i} \sim \log n$ and c is a constant, then

$$P[V \geq cH_n] \leq e^{-H_n[2+c\ln(\frac{c}{e^2})]}$$

Proof For any $\lambda > 0$,

$$E[e^{\lambda V_i}] \leq (1 - \frac{2}{i})e^{\lambda \cdot 0} + \frac{2}{i}e^{\lambda \cdot 1} = 1 + \frac{2}{i}(e^{\lambda} - 1)$$

Let $x = cH_n$, using (4), we have

$$P[V \geq cH_n] \leq e^{-\lambda cH_n} \Pi_{i=1}^{n} E[e^{\lambda V_i}] \leq e^{-\lambda cH_n} \cdot e^{\sum_{i=1}^{n}(e^{\lambda}-1)2/i)}$$

Hence we have

$$P[V \geq cH_n] \leq e^{-H_n(2+c\lambda-2e^{\lambda})}$$

Let $\lambda = \ln c$, we prove the theorem.

Since $H_n = \sum_{i=1}^{n} \frac{1}{i} \sim \log n$, if $c \geq e^2$, then $2 + c\ln\frac{c}{e^2} > 0$, let $\alpha = 2 + c\ln\frac{c}{e^2} > 0$, we got

$$P[V \geq cH_n] \leq e^{-H_n \cdot \alpha}$$

Thus we have V=$\tilde{\mathrm{O}}(H_n)$=$\tilde{\mathrm{O}}(\log n)$, i.e. the length of the search path is $\tilde{\mathrm{O}}(\log n)$.

The proceeding high-probability bound is for a fixed query point p, we can also prove that the same bound holds regardless of the query point.

References

[1] H. Chernoff, A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations. *Annals of Math. Statistics*, 23:493-507, 1952.

[2] K. L. Clarkson, K. Mehlhorm and R. Seidel. Four results on randomized incremental constructions. *Proceedings of the 9th symposium on Theoretical Aspects of Computer Science, lecture Notes in Computer Science*, volume 577, 463-474. Springer-Verlag, 1992.

[3] J. Matousek, Approximations and optimal geometric divide-and-conquer. *23rd STOC*, 505-511, 1991.

[4] J. H. Reif and S. Sen, Optimal randomized parallel algorithms for computational geometry. *Proceedings of the 16th International conference on parallel processing*, 1987.

[5] J. Spencer, Ten lectures on probabilistic method. *CBMS-NSF, SIAM*, 1987.