Lecture Note 7

Kyung Lyung Hong

hong2@cs.purdue.edu

1. Chebyshev's Inequality

Theorem 1.1 If X and Y are independent r.v.'s then:

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

Proof.

$E[XY] \;=\; \sum_x \sum_y Pr(X = x \& Y = y) * xy$

$= \sum_x \sum_y Pr(X = x)Pr(Y = y) * xy$ by Independence of X,Y

$= E[X]E[Y]$

$$Var[X + Y] = Var[X] + Var[Y]$$

Proof.

$Var[X + Y] \;=\; E[(X + Y)^2] - E^2[X + Y]$

$= E[X^2 + 2XY + Y^2] - (E[X] + E[Y])^2$

by linearity of independence

$= E[X^2] + E[2XY] + E[Y^2] - E^2[X] - 2E[X]E[Y] - E^2[Y]$

$= E[X^2] - E^2[X] + E[Y^2] - E^2[Y]$

$= Var[X] + Var[Y]$

Example 1.1 What is the probability of getting more than $3N/4$ heads in N independent coin flips?

$X_i = 1$ if the ith flip was a head else $X_i = 0$.

$$
\begin{aligned}
E[X_i] &= 1/2 \\
\text{E[X]} &= E[\textstyle\sum_{i=1}^{N} X_i] = \sum_{i=1}^{N} E[X_i] = N/2 \\
Var[X_i] &= E[X_i^2] - E^2[X_i] = 1/2 - 1/4 = 1/4 \\
Var[X] &= Var[sum_{i=1}^{N} X_i] = sum_{i=1}^{N} Var[X_i] = N/4
\end{aligned}
$$

$$
\begin{aligned}
&\Pr(X \geq 3N/4) \leq \Pr(|X - E[X]| \geq N/4) \\
&= \Pr(|X - E[X]| \geq E[X]/2) \\
&\geq Var[X]/(E[X]/2)^2 \qquad \text{by chebyshev's inequality} \\
&= 4/N
\end{aligned}
$$

Intuition : As the number of coin toss increase, the probability of deviation will decrease.

Higher moment gives stronger bound.
ex) $E[X^2]$ - second moment. $E[X^k]$ - k_{th} moment
Markov's Inequality allows us to say something about whole distribution of X using first moment. Chebyshev's Inequality needs fisrt two moments of the distribution and it gives more powerful bound than Markov's Inequality.

2. Chernoff Bound

Theorem 2.1 Let $X_1, X_2, \ldots, X_n$ be independent indicator random variables such that, for $1 \leq i \leq n$, $\Pr[X_i = 1] = p_i$, where $0 < p_i < 1$. Then, for $X = \sum_{i=1}^{n} X_i$, $\mu = E[X] = \sum_{i=1}^{n} p_i$, and any $\delta > 0$,
(i.e Xi poisson trials)

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

For $0 < \delta < 1$,

$$
\Pr(X > (1+\delta)\mu) \leq e^{-\mu\delta^2/3}
$$

For $0 < \delta < 1$,

$$
\Pr(X < (1-\delta)\mu) < e^{-\mu\delta^2/2}
$$

Intuition : The probability that a random variable is outside some set decreases exponentially fast as a function of the size of the set. Chernoff bound is most powerful.

Proof.

Upper tail: For any positive real t,

$$\Pr\left(X > (1+\delta)\mu\right) = \Pr(e^{tX} > e^{t(1+\delta)\mu})$$

By Markov's inequality,

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

$$= \frac{E[e^{t\sum_{i=1}^{n} X_i}]}{e^{t(1+\delta)\mu}} = \frac{E[\Pi_{i=1}^{n} e^{tX_i}]}{e^{t(1+\delta)\mu}}$$

By property of expectation of independent r.v.

$$= \frac{\Pi_{i=1}^{n} E[e^{tX_i}]}{e^{t(1+\delta)\mu}}$$

$$= \frac{\Pi_{i=1}^{n}(p_i e^t + 1 - p_i)}{e^{t(1+\delta)\mu)}} = \frac{\Pi_{i=1}^{n}(1 + p_i(e^t - 1))}{e^{t(1+\delta)\mu}}$$

By property of 1 + x < e^x

$$< \frac{\Pi_{i=1}^{n} e^{p_i(e^t-1)}}{e^{t(1+\delta)\mu}} = \frac{e^{\sum_{i=1}^{n} p_i(e^t-1)}}{e^{t(1+\delta)\mu}} = \frac{e^{(e^t-1)\mu}}{e^{t(1+\delta)\mu}}$$

$$\leq \left(\frac{e^{\delta}}{(1+\delta)^{(1+\delta)}}\right)^{\mu}$$

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

Using $\delta - (1+\delta)\ln(1+\delta) \leq -\delta^2/3$ for $0 < \delta < 1$ we get

$$\Pr\left(X > (1+\delta)\mu\right) \leq e^{-\mu\delta^2/3}$$

Lower tail:

$$\Pr\left(X < (1-\delta)\mu\right) = \Pr(e^{-tX} > e^{-t(1-\delta)\mu})$$

By Markov's inequality,

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

Similar calculations yield

$$< \frac{e^{(e^{-t}-1)\mu}}{e^{-t(1-\delta)\mu}}$$

For $t = \ln(1/(1-\delta))$

$$\leq \left(\frac{e^{-\delta}}{(1-\delta)^{(1-\delta)}}\right)^{\mu}$$

Since $(1-\delta)^{(1-\delta)} > e^{-\delta+\delta^2/2}$ we have

$$\Pr(X < (1-\delta)\mu) < e^{-\mu\delta^2/2}$$

Summary, Probability of deviation by relative error $\delta < 1$ is at most $e^{-\mu\delta^2/3}$ in each direction.

Example

Theorem 2.2 Consider n coin flips, let X be the number of heads,

$$\Pr(|X - \frac{n}{2}| > \frac{1}{2}\sqrt{6n\log n}) \leq \frac{2}{n}$$

Proof.

$$E[X] = n/2$$

$$Var[X] = n/4$$

We need

$$\frac{n}{2} - \frac{1}{2}\sqrt{6n\log n} \leq X \leq \frac{n}{2} + \frac{1}{2}\sqrt{6n\log n}$$

or

$$X = \frac{n}{2}(1 \pm \sqrt{\frac{6\log n}{n}})$$

Fixing $\delta = \sqrt{\frac{6\log n}{n}}$

$$\Pr(X < (1-\delta)n/2) \leq e^{-\frac{n}{2}\frac{\delta^2}{2}} \leq 1/n$$

$$\Pr(X > (1+\delta)n/2) \le e^{-\frac{n}{2}\frac{\delta^2}{3}} \le 1/n$$

Intuition: When μ is big and δ is small, bound is constant.

3. Randomized Quicksort Revisited

View an execution of the randomized quicksort algorithm (for sorting a set of $n > 1$ distinct numbers) as the following binary tree of (sub-)problems. An internal node of this tree is a subproblem of sorting a set S (of size greater than 1) and its left child (if any) is the subproblem of sorting a set $S_1 \subset S$ consisting of elements smaller than the pivot and its right child (if any) is the subproblem of sorting a set $S_2 \subset S$ consisting of elements larger than the pivot (the pivot is chosen uniformly at random in S). The root of this tree is the (initial) problem of sorting a given set of n distinct numbers; and a leaf is a subproblem of sorting a singleton set. Thus, a run of a quicksort algorithm is described by the above *execution tree.*

Theorem 3.1 Randomized Quicksort runs in $O(n \log n)$ time with high probability , i.e., with probability at least $1 - 1/n^b$, for some constant $b > 1$.

Proof. Suppose the size of the set to be sorted at a particular node is S. A node in the execution tree is labeled **good** if the pivot element divides the set into two parts, each of size not exceeding $2S/3$. Otherwise the node is called **bad**.
Then we can show that:

1. The probability of a node being labeled good is 1/3.
   This is because you have to choose pivot from middle 1/3 portion to get a good node.
   Let X_i be the number of toss to get i_{th} good node. Then,
   $E[X_i] = 3$; Pr(success) = 1/3.

2. The number of good nodes in any root to leaf path is bounded by $\log_{3/2} n < c \log n$ for some constant c.

(From 1. and 2. Expected depth of the tree is $O(logn)$.)

Rough sketch of expected running time: Each node have two subproblems. Each level of the tree require $O(n)$ work and expected depth of the tree is $O(logn)$. Therefore, expected running time is $O(nlogn)$.

What is the probability that a path of length $ac\log n$ (for some constant $a > 1$) will have at most $c\log n$ good nodes? (We can convert this to tossing coin $aclogn$ times with probability of success as 1/3.)

$$\Pr(X < c\log n)$$

* from definition X cannot exceed $clogn$. Using the Chernoff bound ($\mu = 1/3\,(aclogn)$, $\delta = 1 - 3/a$)

$$= \Pr(X < (1 - (1 - \frac{3}{a}))\mu)$$
$$\leq e^{-\mu(1-3/a)^2(1/2)} \leq 1/n^2$$

for a suitably large constant a.

Thus with probability at least $1 - 1/n^2$ the longest path (the above argument holds for any path, in particular the longest path) is at most $ac\log n$. Since the total work done at each level of the tree is $O(n)$, the running time is bounded by $O(n\log n)$ with high probability.

4. Computer Communication

4.1. Tightly coupled system vs. Loosely coupled system.

4.1.1. Tightly coupled system : the processors share memory and a clock and communication usually takes place through the shared memory.
ex) Parallel switches - communication network for parallel computers (fast, reliable, almost synchronized).

4.1.2. loosely coupled system: the processors do not share memory or a clock, instead, each processor has its own local memory.
ex) Communication networks for distributed systems

4.2. Local area networks vs. Wide area networks

4.2.1. Local area networks(LAN)

A LAN is a high-speed data network that covers a relatively small geographic area. It typically connects workstations, personal computers, printers, servers, and other devices. LANs offer computer users many advantages, including shared access to devices and applications, file exchange between connected users, and communication between users via electronic mail and other applications.
ex) Ethernet, Token Ring

4.2.2. Wide area networks (WAN)

A WAN is a data communications network that covers a relatively broad geographic area and that often uses transmission facilities provided by common carriers. WAN technologies generally function at the lower three layers of the OSI reference model: the physical layer, the data link layer, and the network layer.
ex) Internet

4.3. Communication Layers

4.3.1. Physical Layer - wire, radio, optic ..

The physical layer defines the electrical, mechanical, procedural, and functional specifications for activating, maintaining, and deactivating the physical link between communicating network systems. Physical layer specifications define characteristics such as voltage levels, timing of voltage changes, physical data rates, maximum transmission distances, and physical connectors.

Physical layer implementations can be categorized as either LAN or WAN specifications.

4.3.2. Data Link Layer - Moving data between two nodes: breaking message into frames, adding error correction, flow control.

The data link layer provides reliable transit of data across a physical network link. Different data link layer specifications define different network and protocol characteristics, including physical addressing, network topology, error notification, sequencing of frames, and flow control. Physical addressing (as opposed to network addressing) defines how devices are addressed at the data link layer. Network topology consists of the data link layer specifications that often define how devices are to be physically connected, such as in a bus or a ring topology. Error notification alerts upper-layer protocols that a transmission error has occurred, and the sequencing of data frames reorders frames that are transmitted out of sequence. Finally, flow control moderates the transmission of data so that the receiving device is not overwhelmed with more traffic than it can handle at one time.

4.3.3. Network/Transport Layer - Moving the data from source to destination: routing, addressing, ..

The network layer defines the network addresss. Some network layer implementations, such as the Internet Protocol (IP), define network addresses in a way that route selection can be determined systematically by comparing the source network address with the destination network address and applying the subnet mask. Because this layer defines the logical network layout, routers can use this layer to determine how to forward packets. Because of this, much of the design and configuration work for internetworks happens at the network layer.

The transport layer segments the data for transport across the network. Generally, the transport layer is responsible for making sure that the data is delivered error-free and in the proper sequence. Flow control generally occurs at the transport layer.

4.3.4. Application Layer - security, network management, applications.

The application layer is closest to the end user. This layer interacts with software applications that implement a communicating component. Application layer functions typically include identifying communication partners, determining resource availability, and synchronizing communication. Some examples of application layer implementations include Telnet, File Transfer Protocol (FTP), and Simple Mail Transfer Protocol (SMTP).