Review Session 1

1 Probability

Sample Space. Sample space of a probabilistic experiment is the set of all possible outcomes of the experiment.

- We will only consider finite sample spaces
- In most cases, the sample space will be the set $\{0,1\}^k$ of size 2^k.

Event. Event is a subset of the sample space.

Union Bound. If S is a sample space and $A, A' \subseteq S$, then the probability that either A or A' occurs is $\Pr[A \cup A'] \leq \Pr[A] + \Pr[A']$.

Random Variables. A random variable is a function that maps elements of the sample space to another set. It assigns values to each of the experiment's outcomes.

Example. In the uniform distribution $\{0,1\}^3$, let Random Variable N denote the number of 1s in the chosen string, i.e., for every $x \in \{0,1\}^3$, $N(x)$ is the number of 1s in x.

$$\Pr_{x \in \{0,1\}^3}[N(x) = 2] = 3/8$$

Expectation: The expectation of a random variable is its weighted average, where the average is weighted according to the probability measure on the sample space. The expectation of random variable N is defined as:

$$\mathbb{E}[N] = \sum_{x \in S} N(x).p_x \tag{1}$$

where S is the sample space and p_x is the probability of obtaining x when sampling from S.

Example. If N is defined as in the above example,

$$\mathbb{E}[N] = 0.(1/8) + 1.(3/8) + 2.(3/8) + 3.(1/8) = 1.5$$

Expectation is a linear function, i.e.,

$$\mathbb{E}[N + M] = \mathbb{E}[N] + \mathbb{E}[M] \tag{2}$$

Variance: Variance of a random variable N is defined as the expectation of the square of the distance of N from its expectation.

$$Var[N] = \mathbb{E}[(N - \mathbb{E}[N])^2] \tag{3}$$

Example. If N is defined as in the first example,

$$Var[N] = (0 - 1.5)^2.(1/8) + (1 - 1.5)^2.(3/8) + (2 - 1.5)^2.(3/8) + (3 - 1.5)^2.(1/8) = 0.75$$

Variance is a measure of how spread out the values in a distribution are. A low variance means the outcomes will usually be very close to one another.

Standard Deviation: Standard deviation of N is the square root of $Var[N]$

Independent Events: We say that B is independent from A if $\Pr[B|A] = \Pr[B]$, i.e.,

$$\Pr[A \cap B] = \Pr[A].\Pr[B] \tag{4}$$

Example. Tossing a coin. The probability that heads shows up on two consecutive coin tosses,

$$\Pr[HH] = \Pr[H].\Pr[H] = \frac{1}{2}.\frac{1}{2} = 0.25$$

Each toss of a coin is an Independent Event.

Pairwise Independent Random Variables: Let $X_1, X_2.....X_n$ be random variables. We say that the X_i's are pairwise-independent if for every $i \neq j$ and all a and b, it holds that:

$$\Pr[X_i = a \text{ and} X_j = b] = \Pr[X_i = a].\Pr[X_j = b] \tag{5}$$

Example. We throw two dice.
Let A be the event "the sum of the points is 7"
B the event "die #1 came up 3" and
C the event "die #2 came up 4".

$$\Pr[A] = \Pr[B] = \Pr[C] = \frac{1}{6}$$

$$\Pr[A \cap B] = \Pr[B \cap C] = \Pr[A \cap C] = \frac{1}{36}$$

But,

$$\Pr[A \cap B \cap C] = \frac{1}{36} \neq \Pr[A].\Pr[B].\Pr[C]$$

A, B and C are pairwise independent but not independent as a triplet.

Conditional Probability: The conditional probability of event B in relation to event A is the probability of event B occurring when we know that A has already occurred.

$$\Pr[B|A] = \Pr[A \cap B] / \Pr[A] \tag{6}$$

Example. Drawing Kings from a deck of cards.
Event A is drawing a King first, and Event B is drawing a King second.

$$\Pr[A] = \frac{4}{52}$$

$$\Pr[B|A] = \frac{3}{51}$$

2 Tail Bounds

Markov's Inequality: Let X be a *non-negative random variable* and let $k \geq 1$. Then,

$$\Pr[X \geq k] \leq \mathbb{E}[X]/k \tag{7}$$

or

$$\Pr[X \geq k.\,\mathbb{E}[X]] \leq 1/k \tag{8}$$

Example. Suppose we roll a single fair die and let X be the outcome.
Then,

$$\mathbb{E}[X] = 3.5$$

$$Var[X] = 35/12$$

Suppose we want to compute $p = \Pr[X \geq 6]$. We can easily see that $p = \Pr[X \geq 6] = \Pr[X = 6] \approx 0.167$.
Using Markov's Inequality we can get an upper bound,

$$\Pr[X \geq 6] \leq 3.5/6 \approx 0.583$$

Markov's inequality gives an answer to the question "what is the probability that the value of the r.v., X, is 'far' from its expectation?".

Chebyshev's Inequality: Let X be a random variable, and $k \geq 1$. Then,

$$\Pr[|X - \mathbb{E}[X]| \geq k] \leq Var[X]/k^2 \tag{9}$$

Another answer to the question of "what is the probability that the value of X is far from its expectation" is given by Chebyshev's Inequality, which works for *any random variable* (not necessarily a non-negative one).

Example. Let X be as defined in the above example.
We can get a better bound on $\Pr[X \geq 6]$ using Chebyshev's inequality,

$$\Pr[X \geq 6] \leq \Pr[X \geq 6 ORX \leq 1] = \Pr[|X - 3.5| \geq 2.5] \leq \frac{35/12}{(2.5)^2} = \frac{7}{15} \approx 0.46$$

This inequality is a particularly useful for analysis of the error probability of approximation via repeated sampling.

Chernoff Bound: Let $X_1, X_2.....X_n$ be independent random variables with $0 \leq X_i \leq 1$ (they need not have the same distribution). Let $X = X_1 + X_2 + \ldots + X_n$ and $\mu = \mathbb{E}[X] = \mathbb{E}[X_1] + \ldots + \mathbb{E}[X_n]$, then for any $\epsilon \geq 0$

$$\Pr[X \geq (1+\epsilon)\mu] \leq e^{-\frac{\epsilon^2}{2+\epsilon}\mu} \tag{10}$$

Example. A biased coin, which lands heads with probability 1/10 each time it is flipped, is flipped 200 times consecutively. We want an upper bound on the probability that it lands heads at least 120 times.

The number of heads is a binomially distributed r.v., X, with parameters $p = 1/10$ and $n = 200$. Thus, the expected number of heads is

$$\mathbb{E}[X] = np = 200.(1/10) = 20$$

Using Markov's inequality

$$\Pr[X \geq 120] \leq \frac{20}{120} = \frac{1}{6}$$

Using Chernoff bound we get,

$$\Pr[X \geq 120] = \Pr[X \geq (1+5)20] \leq e^{-\frac{5^2}{2+5}20} \approx e^{-71.4}$$

which is a much better bound.

For specific random variables, particularly those that arise as sums of many independent random variables, we can get much better bounds on the probability of deviation from expectation.

3 Running Time

Polynomial Running Time: An algorithm A is said to run in polynomial time if there exists a constant c such that A runs in time $T(n) = n^c$.

We say an algorithm is efficient if it runs in polynomial time. If an algorithm runs exponential or super-polynomial time, i.e., $T(n) = 2^n$ or $T(n) = n^{(\log n)}$, then we will say it is inefficient. Note that here, c could be an arbitrary constant. In particular, it may not be small.

4 Turing Machine

A Turing machine is a hypothetical machine that can simulate any computer algorithm. It can be thought of as a set of infinite tapes consisting of equal sized cells. All the computations are performed by making changes to the contents on these tapes. The running time of an algorithm is measured by the number of steps taken by the Turing machine to halt.

Deterministic Turing Machine: In a deterministic Turing machine, the set of rules prescribe at most one action to be performed for any given situation. Any polynomial time computation can be captured by a polynomial time turing machine.

Probabilistic Polynomial Time Turing Machine Any probabilistic polynomial time (PPT) algorithm can be captured by this turing machine. A probabilistic polynomial time turing machine is a Turing machine

that runs in polynomial time and is equipped with an extra randomness tape. Each bit of randomness tape is uniformly and independently chosen.The output of a randomized algorithm is a distribution.

Non-Deterministic Turing Machine: A non-deterministic Turing machine (NTM) may have a set of rules that prescribes more than one action for a given situation. It non-deterministically decides which action to perform at run-time.

5 Complexity Classes

Complexity Class P: A language L is recognizable in (deterministic) polynomial time if there exists a Turing machine M and a polynomial p(.) such that:

- on input strings x, machine M halts after atmost p($|x|$) steps and
- M(x)=1 if and only if x∈L

P is the class of languages that can be recognized in (deterministic) polynomial time.

Complexity Class NP: A language L is in NP if there exists a Boolean relation $R_L \subseteq \{0,1\}* \times \{0,1\}*$ and a polynomial p(.) such that R_L can be recognized in (deterministic) polynomial time, and $x \in$ L if and only if there exists a y such that $|y| \leq$ p($|x|$) and $(x, y) \in R_L$. Such a y is called a witness for membership of $x \in$ L.

NP-Completeness: A language is NP-complete if it is in NP and every language in NP is polynomially reducible to it.
A language L is polynomially reducible to a language L' if there exists a polynomial-time-computable function f such that $x \in$ L if and only if $f(x) \in$ L'.

Bounded error Probabilistic Polynomial Time, BPP: We say that L is recognized by the probabilistic polynomial Time Turing machine M if:

- for $x \in$ L, then Pr$[M(x) = 1] \geq 2/3$
- for $x \notin$ L, then Pr$[M(x) = 0] \geq 2/3$

6 Asymptotic Notations

Asymptotic bounds on the growth rate of runtime of an algorithm:

Big-O: $f(n)$ is O($g(n)$) if for some constants $c(c > 0), n_0(n_0 > 0)$, $f(n) \leq c.g(n)$ for $n > n_0$

Big-Omega: $f(n)$ is $\Omega(g(n))$ if for some constants $c(c > 0), n_0(n_0 > 0)$, $f(n) \geq c.g(n)$ for $n > n_0$

Small-o: $f(n)$ is o($g(n)$) if for some constants $c(c > 0), n_0(n_0 > 0)$, $f(n) < c.g(n)$ for $n > n_0$

Small-omega: $f(n)$ is $\omega(g(n))$ if for some constants $c(c > 0), n_0(n_0 > 0)$, $f(n) > c.g(n)$ for $n > n_0$

Theta: $f(n)$ is $\Theta(g(n))$ if for some constants $c_1(c_1 > 0), c_2(c_2 > 0)n_0(n_0 > 0)$, $c_1.g(n) < f(n) < c_2.g(n)$ for $n > n_0$

7 References:

1. http://www.boazbarak.org/cs127/chap001-mathematical-background.pdf

2. https://learnxinyminutes.com/docs/asymptotic-notation/
3. http://faculty.washington.edu/fm1/394/Materials/2-3indep.pdf
4. http://www.ams.sunysb.edu/ jsbm/courses/311/cheby.pdf
5. https://www.inf.ed.ac.uk/teaching/courses/dmmr/slides/13-14/chebi-Ch7.pdf
6. https://www.cs.princeton.edu/courses/archive/fall09/cos521/Handouts/probabilityandcomputing.pdf
7. Foundations of Cryptography: Volume 1, Basic Tools, Oded Goldreich