CSC373 Week 10 Notes

Intro to Stats:

- A random variable (r.v.) is a math formalization of a quantity/object which depends on random events.

I.e. x is a r.v. if it takes values from an outcome set.

- The probability of an event, denoted as $P(X=k)$, is the num of times we observe k over N events.

I.e. $P(X=k) = \lim_{N \to \infty} \frac{\text{(\# of times k is observed)}}{N}$

- The Law of Total Probability states that

$$\sum_k P(X=k) = 1$$

- The expectation of f, denoted as $E(f)$, is:

$$E(f) = \sum_i P(i) \cdot f(i)$$

Note: Expectation is linear $\Rightarrow E(f+g) = E(f) + E(g)$

- The Union Bound states that $P(A \cup B) \leq P(A) + P(B)$

Proof:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$
$$\leq P(A) + P(B)$$

- X and Y are independent r.v. if $P(X \cap Y) = P(X) \cdot P(Y)$

Note: If X and Y are indep r.v., then $P(X|Y) = P(X)$

Proof:

$$P(X|Y) = \frac{P(X \cap Y)}{P(Y)} = \frac{P(X) \cdot P(Y)}{P(Y)} = P(X)$$

- The Uniform distribution is a distribution where each event has the same prob.
I.e. Say $X \in \{1, \ldots, n\}$. Then, $P(x) = \frac{1}{n}$, $E(x) = \frac{n+1}{2}$, $Var(x) = \frac{n^2-1}{2}$

- The Variance, denoted as $Var(x)$, is a measurement of how close the points are to the mean. A smaller variance means the points are close to the mean. A larger variance means the points are far from the mean.
$Var(x) = E(x^2) - (E(x))^2$

- The Geometric Distribution is the number of Bernoulli trials needed to get one success.
$P(X=k) = (1-p)^{k-1} p$
$E(k) = \frac{1}{p}$
$Var(k) = \frac{1-p}{p^2}$

- With the Normal Distribution/Gaussian Distribution,

$$P(x) = \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

$E(x) = \mu$
$Var(x) = \sigma^2$

Randomized Algo:

- The idea is to first devise a randomized algo and then use derandomization to yield a deterministic algo.

- E.g. Max k-SAT

Problem: Given an exact k-SAT formula $F = C_1 \wedge C_2 \wedge \dots \wedge C_m$ where each clause has exactly k literals and a weight $w_i \geq 0$, we want to output a truth assignment τ max the number/total weight of clauses satisfied under τ.

Soln:

Let $W(\tau)$ be the total weight of the clauses satisfied under τ.

We will let τ be a random assignment.
I.e. We will randomly choose each literal's value.

For each clause C_i, $P(C_i \text{ is NOT satisfied}) = \frac{1}{2^k}$
This is bc each clause has k values and the only way C_i is not satisfied is if all the literals are False.

$$P(C_i \text{ is satisfied}) = 1 - P(C_i \text{ isn't satisfied}) = 1 - \frac{1}{2^k} = \frac{2^k - 1}{2^k}$$

$$E(W(\tau)) = \sum_{i=1}^{m} w_i \cdot P(C_i \text{ is satisfied}) + w_i \cdot P(C_i \text{ isn't satisfied})$$

$$= \sum_{i=1}^{m} w_i \cdot P(C_i \text{ is satisfied}) + 0 \cdot P(C_i \text{ isn't satisfied})$$

$$= \sum_{i=1}^{m} P(C_i \text{ is satisfied}) \cdot w_i$$

$$\geq \frac{2^k - 1}{2^k} \text{ OPT}$$

$$\therefore \frac{E(w(\tau))}{OPT} \geq \frac{2^k-1}{2^k}$$

This is a random and "avg-case" ratio. Now, we'll derandomize it to get a deterministic ratio. We will use the Method of Conditional Prob to do so.

Recall: $P(x) = \sum_a P(x|a) \cdot P(a)$

We can rewrite $E(w(\tau))$ as:

* $E(w(\tau)) = P(X_1 = T) \cdot E(w(\tau) | X_1 = T) + P(X_1 = F) \cdot E(w(\tau) | X_1 = F)$

$= \frac{1}{2} E(w(\tau) | X_1 = T) + \frac{1}{2} E(w(\tau) | X_1 = F)$

We know that $\max(E(w(\tau) | X_1 = T), E(w(\tau) | X_1 = F)) \geq E(w(\tau))$
Hence, if we compute both $E(w(\tau) | X_1 = T)$ and $E(w(\tau) | X_1 = F)$ and we pick the better one, we can deterministically set X_1 without degrading the obj value.

We do the same thing for $X_2, X_3, \ldots$

$E(w(\tau)) = \sum_i w_i \cdot P(C_i \text{ is satisfied})$

$= \sum_i w_i \, P(C_i \text{ is satisfied} | X_i = T) \cdot P(X_i = T) + w_i \, P(C_i \text{ is satisfied} | X_i = F) \cdot P(X_i = F)$

$= \sum_i \frac{1}{2}(w_i \cdot P(C_i \text{ is satisfied} | X_i = T) + \frac{1}{2}(w_i \cdot P(C_i \text{ is satisfied} | X_i = F)$

$= \sum_i \frac{1}{2} E(w(\tau) | X_i = T) + \frac{1}{2} E(w(\tau) | X_i = F)$