CS2800 HW 13 Adam Skrocki

1) If we have n vertices, there are $\binom{n}{3}$ possible triangles that can be formed because a triangle is a combination of 3 vertices that are all connected. The probability that a combination of 3 vertices is a triangle is $p \cdot p \cdot p = p^3$, the probability that the 3 vertices are all connected by edges. Let our sample space be the set of all possibilities for n vertices to be connected. We can use an indicator $I_{v_1 v_2 v_3} = 1$ if vertices v_1, v_2, and v_3 form a triangle, and $I_{v_1 v_2 v_3} = 0$ otherwise. Because the probability of 3 vertices forming a triangle is p^3, this is also the probability that $I_{v_1 v_2 v_3} = 1$. The total number of triangles can be given by $\sum_{n=1}^{\binom{n}{3}} I_{v_1 v_2 v_3}$, so the expected number of triangles is $E\left(\sum_{n=1}^{\binom{n}{3}} I_{v_1 v_2 v_3}\right) = \sum_{n=1}^{\binom{n}{3}} E(I_{v_1 v_2 v_3})$

$E(I_{v_1 v_2 v_3}) = 1(p^3) + 0(1-p^3) = p^3$, so we get

$$\sum_{n=1}^{\binom{n}{3}} E(I_{v_1 v_2 v_3}) = \sum_{n=1}^{\binom{n}{3}} p^3 = \boxed{\binom{n}{3} p^3}$$

(the sum is over $\binom{n}{3}$ because there are $\binom{n}{3}$ possible triangles)

2) (a) A point in the unit square can be represented as a pair of numbers on the interval $[0,1]$, so $[0,1]\times[0,1]$ is a point. The sample space is all the outcomes possible when choosing n points in sequence, so $S = \underbrace{([0,1]\times[0,1]) \times ([0,1]\times[0,1])}_{\text{2 times shown}} \ldots$ n times

(b) We want to find the expected # of points in the unit circle. Let I_m be the indicator variable for whether the mth point is inside the circle. $I_m = 1$ if the mth point is in the unit circle, $I_m = 0$ otherwise. The number of points in the unit circle

$N = \sum_{m=1}^{n} I_m$, so $E(N) = \sum E(I_m)$. The probability that $I_m = 1$ is the portion of the area of the unit square that is the unit circle, so $\frac{\pi(1)^2}{4} / 1^2 = \frac{\pi}{4}$,

so $E(I_m) = \frac{\pi}{4}(1) + (1 - \frac{\pi}{4})0 = \frac{\pi}{4}$. So

$E(N) = \sum_{n=1}^{n} E(I_n) = \frac{n\pi}{4}$

(c) If $E(N) = \frac{n\pi}{4}$, then $\frac{4}{n}E(N) = \pi$. Let $P = \frac{4N}{n}$, and

$E(P) = E(\frac{4N}{n}) = \frac{4}{n}E(N)$, which we saw equals

π. $P = \frac{4N}{n}$

(d) $Var(P) = Var(\frac{4N}{n}) = \frac{16}{n^2} Var(N)$

$N = \sum I_m$, so $Var(N) = \sum Var(I_m) = n\,Var(I_n)$

$Var(I_m) = E(I_m^2) - (E(I_m))^2 = E(I_m) - (\frac{\pi}{4})^2$ because it is always the case that $I_m^2 = I_m$ as I_m can only be 0 or 1. $E(I_m) - (\frac{\pi}{4})^2 = \frac{\pi}{4} - (\frac{\pi}{4})^2$, so

$Var(N) = n(\frac{\pi}{4} - (\frac{\pi}{4})^2)$ and $Var(P) = \frac{16}{n^2} n (\frac{\pi}{4} - (\frac{\pi}{4})^2)$

$= \frac{4}{n}(\pi - \frac{\pi^2}{4})$

(e) Chebychev's inequality: $Pr(|X - E(X)| \geq a) \leq \frac{Var(X)}{a^2}$

The probability that our estimate is within $\frac{1}{1000}$ of π being at least 50% is the same as the probability of our estimate not being within $\frac{1}{1000}$ of π with at most 50% probability. So to apply chebychev's to this situation, $Pr(|P - \pi| \geq \frac{1}{1000}) \leq \frac{Var(P)}{(\frac{1}{1000})^2}$

The RHS must be $\leq 50\% = \frac{1}{2}$, so $\frac{4}{n}(\pi - \frac{\pi^2}{4})(1000)^2 \leq \frac{1}{2}$

$n \geq 8000000(\pi - \frac{\pi^2}{4}) \approx 5393532.43$

$n \geq 5393532.43$

3) (a) Let $|X| = k$, $|Y| = m$. The set of inputs besides x_1 is $X \setminus \{x_1\}$, so the number of inputs besides x_1 is $|X \setminus \{x_1\}| = |X| - 1 = k-1$

Let I_n be an indicator for whether the nth element of $X \setminus \{x_1\}$ hashes to the same value as x_1 does. ($I_n = 1$ if so, 0, if not)

$N = I_1 + I_2 \cdots I_{k-1} = \sum_{n=1}^{k-1} I_n$ (where N is the # of inputs besides x_1 that hash to the same value)

$E(I_n) = \Pr(\text{nth elem of } X \setminus \{x_1\} \text{ hashes to same as } x_1) \cdot 1 + 0 \cdot (1 - \ldots)$

$E(I_n) = \frac{1}{m}$ for all elements of $X \setminus \{x_1\}$ because our hash function is nice and random, i.e.

(1) $\Pr(Y_n = y) = \frac{1}{|Y|} = \frac{1}{m}$ and

(2) Y_i and Y_j are independent, i.e. previous mappings don't affect later ones at all, which is why we can sum I_n's to get N.

So $E(N) = \sum_{n=1}^{k-1} E(I_n) = \frac{k-1}{m}$ ✓

(b) Consider an H such that hashing any input is the equivalent of hashing the input $x_1 \in X$ (the set of inputs), and that for hashing x_1, all outputs are equally likely, i.e. $\Pr(Y_1 = y) = \frac{1}{|Y|}$. This hash function satisfies the first property of a nice hash function but not the second because

$$1 = \Pr(Y_2 = y \cap Y_1 = y) \neq \Pr(Y_2 = y)\Pr(Y_1 = y) = \left(\frac{1}{m}\right)\left(\frac{1}{m}\right)$$

Y_1 and Y_2 (or any other Y_n) are not independent because the probability of them mapping to the same thing would be 1, not $\frac{1}{m}\left(\frac{1}{m}\right)$, as it would be if Y_1 and the other arbitrary Y_n were independent. In this case, the number of inputs N that mapped to the same thing as x_1 would be all the elements besides x_1 in X, or $k-1$ if we let $k = |X|$. So in this case, $N = k-1$, or $E(N) = k-1$, not $\frac{k-1}{m}$ like in a nice hash function.

(because $k-1$ is a constant)