CS/ECE 473 Fall 2020 | Ryan Prendergast (ryanp4)
Homework 5 Problem 2 | Noah Watson (nwatson3)
Lawson Probasco (lawsonp2)

(a) **Solution:**

$$
\begin{aligned}
Pr_{h\in\mathcal{M}}[h(x)=h(y)] &= Pr[\bigoplus_{i:x_i=1} M_i = \bigoplus_{i:y_i=1} M_i] &(1)\\
&= Pr[(\bigoplus_{i:x_i=y_i=1} M_i)\oplus(\bigoplus_{i:y_i=0,x_i=1} M_i) = (\bigoplus_{i:x_i=y_i=1} M_i])\oplus(\bigoplus_{i:x_i=0,y_i=1} M_i)] &(2)\\
&= Pr[(\bigoplus_{i:y_i=0,x_i=1} M_i) = (\bigoplus_{i:x_i=0,y_i=1} M_i)] &(3)\\
&= Pr[(\bigoplus_{i:y_i=0,x_i=1} M_i)\oplus(\bigoplus_{i:x_i=0,y_i=1} M_i) = 0] &(4)\\
&= Pr[h(z)=0] &(5)\\
&\leq \frac{1}{2^l} &(6)\\
&\leq \frac{1}{m} &(7)
\end{aligned}
$$

Annotating the equations down here to avoid clutter. Essentially, the rewriting in (1) is simply using the definition of h_M. (2) uses the fact that each pair of distinct input vectors x, y have at least one differing bit, meaning they produce some zero or more shared columns of M and some one or more different columns of M. The shared columns on each side are dropped in (3), and the right side is moved over in (4). The resulting equation in (4) is the xor of a set of at least one unique column(s) of M, which in turn can be rewritten as some $h(z)$ since this is exactly what the hash function produces anyways. This probability can be evaluated in the following manner - fix one non-zero column M_j representing by a corresponding 1-bit in z. This means that $M_j = \bigoplus_{i:x_1=1,i\neq j} M_i$. This probability is bounded by the l random bits in M_j, meaning that $Pr[h(z)=0] \leq \frac{1}{2^l} = \frac{1}{m}$. ■

(b) **Solution:** In order to be uniform, $Pr_{h\in\mathcal{M}}[h(x)=i] = \frac{1}{m}$ for all x and i. However, in the case where x is $\overrightarrow{0}$, and i is $\overrightarrow{1}$, $Pr_{h\in\mathcal{M}}[h(x)=i]=0$. ■

(c) **Solution:** In order to be 2-uniform (as above), $Pr_{h\in\mathcal{M}^+}[h(x)=i] = \frac{1}{m}$ for all x and i. Considering that there will be (without loss of generality, since independent $h(x), h(y)$ can be flipped) at least one bit in y with a value of 0, we come to the conclusion that there will always be some column in M - refer to it as M_i, in which M_i does not contribute in any way to the calculation of v_2 but does contribute to the calculation of v_1 as in $Pr[h(x)=v_1|h(y)=v_2]$. As a result, we can fix this column M_i s.t. $v_2 \oplus M_i = v_1$, e.g. $M_i = v1 \oplus v_2$, more precisely the xor of the unique columns of v_1 and v_2. Since there are l-bits in this vector, the likelihood is then $\frac{1}{2^l} = \frac{1}{m}$. Thus $Pr[h(x)=a \text{ and } h(y)=b] = Pr[h(x)=a] * Pr[h(x)=b] = \frac{1}{m^2}$. ■

(d) **Solution:** ■

In order to be 4-uniform, $Pr_{h\in\mathcal{M}^+}[\bigwedge_{j=1}^4 h(x_j)=i_j] = \frac{1}{m^4}$ for all distinct $x_1,...,x_4$ and $i_1,...,i_4$. However, consider the case where x_2 is the input vector with each even bit set, x_3 is the input vector with each odd bit set, and x_4 is the input vector $\overrightarrow{0}$. In this case then, $Mx_4 = \overrightarrow{0}$, meaning $i_4 = b$. If x_1 is the input vector $\overrightarrow{1}$, this would mean that $i_1 = i_2 \oplus i_3 \oplus i_4$, and as such, these values are not independent and $Pr_{h \in \mathscr{M}^+}[\bigwedge_{j=1}^{4} h(x_j) = i_j] \neq \frac{1}{m^4}$.