

## 1

a. If  $\Pr[x \leftarrow D_{0,k} : x = \hat{x}] > \Pr[x \leftarrow D_{1,k} : x = \hat{x}]$ , then Dave should pick 0 since it is more likely that  $x$  was chosen from  $D_{0,k}$ . Otherwise, Dave should pick 1.

b. To see this, note that  $\Pr[x \leftarrow D_{0,k}; b \leftarrow T(x) : b = 1] = 1 - \Pr[x \leftarrow D_{0,k}; b \leftarrow T(x) : b = 0]$  and likewise that  $\Pr[x \leftarrow D_{1,k}; b \leftarrow T(x) : b = 1] = 1 - \Pr[x \leftarrow D_{1,k}; b \leftarrow T(x) : b = 0]$ . Therefore, we can write this term on the right hand side as  $|1 - \Pr[x \leftarrow D_{0,k}; b \leftarrow T(x) : b = 0] - (1 - \Pr[x \leftarrow D_{1,k}; b \leftarrow T(x) : b = 0])| = |\Pr[x \leftarrow D_{1,k}; b \leftarrow T(x) : b = 0] - \Pr[x \leftarrow D_{0,k}; b \leftarrow T(x) : b = 0]|$ , which by the symmetry of absolute value is also equal to  $\Pr[x \leftarrow D_{0,k}; b \leftarrow T(x) : b = 0] - \Pr[x \leftarrow D_{1,k}; b \leftarrow T(x) : b = 0]$ . Therefore, the whole right-hand side becomes

$$\begin{aligned} & \frac{1}{2}(|\Pr[x \leftarrow D_{0,k}; b \leftarrow T(x) : b = 0] - \Pr[x \leftarrow D_{1,k}; b \leftarrow T(x) : b = 0]| \\ & \quad + |\Pr[x \leftarrow D_{0,k}; b \leftarrow T(x) : b = 0] - \Pr[x \leftarrow D_{1,k}; b \leftarrow T(x) : b = 0]|) \\ &= \frac{1}{2}(2 \cdot |\Pr[x \leftarrow D_{0,k}; b \leftarrow T(x) : b = 0] - \Pr[x \leftarrow D_{1,k}; b \leftarrow T(x) : b = 0]|) \\ &= |\Pr[x \leftarrow D_{0,k}; b \leftarrow T(x) : b = 0] - \Pr[x \leftarrow D_{1,k}; b \leftarrow T(x) : b = 0]| \end{aligned}$$

which we recognize as  $\text{Adv}(\text{Test})$ . Therefore,  $\text{Adv}'(\text{Test}) = \text{Adv}(\text{Test})$ .

c. Ultimately, what we want to show is that  $\text{Adv}' = \Delta(D_{0,k}, D_{1,k})$ . To start, we need to break up our probability terms in  $\text{Adv}'$ . What each term essentially asks for is, given that an  $x$  comes from either  $D_{0,k}$  or  $D_{1,k}$ , the probability that our Test from part (a) will output either a 0 or a 1 when given that  $x$ . Let's consider each  $x$  value individually. In other words, our first term  $\Pr[x \leftarrow D_{0,k}; b \leftarrow \text{Test}(x) : b = 0]$  is equal to  $\sum_{\hat{x} \in S_{0,k}} \Pr[\hat{x} \leftarrow D_{0,k}; b \leftarrow \text{Test}(\hat{x}) : b = 0] = \sum_{\hat{x}} \Pr[x \leftarrow D_{0,k} : x = \hat{x}] \Pr[b \leftarrow \text{Test}(\hat{x}) : b = 0]$

Also, because we are doing this for all possible values of  $x$ , we can represent this as a sum over all  $x$ . Doing this for each term in  $\text{Adv}'$ , we see that  $\text{Adv}' = \frac{1}{2}(|\sum_{\hat{x}} \Pr[x \leftarrow D_{0,k} : x = \hat{x}] \cdot \Pr[b \leftarrow \text{Test}(\hat{x}) : b = 0] - \sum_{\hat{x}} \Pr[x \leftarrow D_{1,k} : x = \hat{x}] \cdot \Pr[b \leftarrow \text{Test}(\hat{x}) : b = 0]| + |\sum_{\hat{x}} \Pr[x \leftarrow D_{0,k} : x = \hat{x}] \cdot \Pr[b \leftarrow \text{Test}(\hat{x}) : b = 1] - \sum_{\hat{x}} \Pr[x \leftarrow D_{1,k} : x = \hat{x}] \cdot \Pr[b \leftarrow \text{Test}(\hat{x}) : b = 1]|)$ . (Note that to save space, I haven't written out the whole second probability terms). Now, we examine the algorithm used in part (a). We know that the second term in all of our sum sequences is going to either be 0 or 1. If  $\Pr[x \leftarrow D_{0,k} : x = \hat{x}] > \Pr[x \leftarrow D_{1,k} : x = \hat{x}]$ , then we know that the probability that  $b = 0$  is going to be 1; otherwise it will be 0. To simplify things, let's denote  $\Pr_0(x) = \Pr[x \leftarrow D_{0,k} : x = \hat{x}]$  and  $\Pr_1(x) = \Pr[x \leftarrow D_{1,k} : x = \hat{x}]$ . Now, we simply throw out the  $x$  values where our second term

will be 0 and rewrite  $\text{Adv}'$  as

$$\begin{aligned} \text{Adv}' &= \frac{1}{2} \left( \left| \sum_{x \mid \Pr_0(x) > \Pr_1(x)} \Pr[x \leftarrow D_{0,k}] - \sum_{x \mid \Pr_0(x) > \Pr_1(x)} \Pr[x \leftarrow D_{1,k}] \right| \right. \\ &\quad \left. + \left| \sum_{x \mid \Pr_0(x) \leq \Pr_1(x)} \Pr[x \leftarrow D_{0,k}] - \sum_{x \mid \Pr_0(x) \leq \Pr_1(x)} \Pr[x \leftarrow D_{1,k}] \right| \right) \\ &= \frac{1}{2} \left( \left| \sum_{x \mid \Pr_0(x) > \Pr_1(x)} \Pr[x \leftarrow D_{0,k}] - \Pr[x \leftarrow D_{1,k}] \right| \right. \\ &\quad \left. + \left| \sum_{x \mid \Pr_0(x) \leq \Pr_1(x)} \Pr[x \leftarrow D_{0,k}] - \Pr[x \leftarrow D_{1,k}] \right| \right). \end{aligned}$$

Now, we can apply the triangle inequality to see that

$$\text{Adv}' \leq \frac{1}{2} \left( \sum_{x \mid \Pr_0(x) > \Pr_1(x)} |\Pr[x \leftarrow D_{0,k}] - \Pr[x \leftarrow D_{1,k}]| + \sum_{x \mid \Pr_0(x) \leq \Pr_1(x)} |\Pr[x \leftarrow D_{0,k}] - \Pr[x \leftarrow D_{1,k}]| \right).$$

In fact, we can do better than this, and say that the two sides are actually equal. In our first term, we can see that we are summing over all  $x$  where the first term in the sum sequence is greater than the second term; in other words, summing over all  $x$  such that the part inside the absolute value stays positive. In our second term, we are dealing with the exact opposite. But, because we know that  $|-x| = |x|$ , we can negate this whole second term to achieve something that is always non-negative, and therefore we achieve equality (clearly  $|\sum x| = \sum |x|$  if all the  $x$  terms are non-negative). Writing this all out, we end up with

$$\text{Adv}' = \frac{1}{2} \left( \sum_{x \mid \Pr_0(x) > \Pr_1(x)} |\Pr[x \leftarrow D_{0,k}] - \Pr[x \leftarrow D_{1,k}]| + \sum_{x \mid \Pr_0(x) \leq \Pr_1(x)} |\Pr[x \leftarrow D_{1,k}] - \Pr[x \leftarrow D_{0,k}]| \right).$$

Now, by the symmetry of absolute value (i.e.  $|x - y| = |y - x|$ ), we find that this becomes

$$\begin{aligned} \text{Adv}' &= \frac{1}{2} \left( \sum_{x \mid \Pr_0(x) > \Pr_1(x)} |\Pr[x \leftarrow D_{0,k}] - \Pr[x \leftarrow D_{1,k}]| + \sum_{x \mid \Pr_0(x) \leq \Pr_1(x)} |\Pr[x \leftarrow D_{0,k}] - \Pr[x \leftarrow D_{1,k}]| \right) \\ &= \sum_{x \in S_{1,k} \cup S_{0,k}} |\Pr[x \leftarrow D_{0,k}] - \Pr[x \leftarrow D_{1,k}]|. \end{aligned}$$

## 2

a. This can be written  $\Pr[m \leftarrow \{0, 1\}^k; r \leftarrow \{0, 1\}^k; c = m \oplus r : c = \hat{c}]$ .

b. We denote  $S$  as our sample space of faces and answer as the event when the machine returns either  $M$  or  $F$  (i.e. doesn't produce an error). Then we denote our probability of overall success as:

$$\Pr[\text{face} \leftarrow S; a \leftarrow \text{test}(\text{face}); \text{if } a = \text{answer then } b = a; \text{ else } b = M : b = \text{gender}(\text{face})] \quad (1)$$

where  $gender()$  represents the gender of each face. Or equivalently:

$$\Pr[\text{gender} \leftarrow M, F; \text{face} \leftarrow S_{\text{gender}}; a \leftarrow \text{test}(\text{face}); \text{if } a = \text{answer} \text{ then } b = a; \text{ else } b = M : b = \text{gender}] \quad (2)$$

where  $S_F$  and  $S_M$  are sets of female and male faces. Now, analyzing the algorithm using the first notation gives us

$$(1) = \Pr[\text{face} \leftarrow S; \text{test}(\text{face}) = \text{answer} : \text{answer} = \text{gender}(\text{face})] \cdot \Pr[a \leftarrow \text{test}(\text{face}) : a = \text{answer}] \\ + \Pr[\text{face} \leftarrow S; \text{test}(\text{face}) \neq \text{answer} : M = \text{gender}(\text{face})] \cdot \Pr[a \leftarrow \text{test}(\text{face}) : a \neq \text{answer}]$$

Now we can stop and figure out what each of these probabilities actually represents. The first probability is asking us, for any face and assuming the test gives us an answer ( $M$  or  $F$ ), what the probability is that the answer is correct. We know this probability is 1, and likewise that the probability of getting an answer is .7. Therefore,

$$(1) = 1(.7) + \Pr[\text{face} \leftarrow S; \text{test}(\text{face}) \neq \text{answer} : M = \text{gender}(\text{face})] \cdot \Pr[a \neq \text{answer}]$$

To analyze this second term, we note that there are only two things the test can return: an answer or an error. Therefore, saying that  $a \neq \text{answer}$  is the same as saying  $a = \text{error}$ . The first term in this expression asks for, given a face the machine could not identify and guessing it is male, the probability that we are correct. Since we did this analysis in the last homework, we know that this probability is .8. Likewise, we know that the probability the machine returns an error is .3. Therefore,  $(1) = 1(.7) + .8(.3) = .94$ , which is our success probability.

### 3

a. Basically, we should always guess that  $x$  is 0, so pick whatever we are given as  $y$  as our output. By the nature of how  $x$  is chosen, we know that we will guess correctly with probability  $\frac{1}{2} + \frac{1}{k}$ . Since  $\frac{1}{k}$  is not a negligible function, this proves that the two sets are not 1/2-indistinguishable.

b. Assume that there exists no such negligible function, that is that  $p_T^{(M,y)} - (1 - p_T^{(F,y)}) = \epsilon$  for some non-negligible  $\epsilon$  and for either  $y = 0$  or  $y = 1$ . Now, we analyze the success of an algorithm Test. We can see that

$$\Pr[\text{success}] - \Pr[\text{fail}] = \Pr[y = 0](\Pr[\text{success}|y = 0] - \Pr[\text{fail}|y = 0]) \\ + \Pr[y = 1](\Pr[\text{success}|y = 1] - \Pr[\text{fail}|y = 1]) \\ = \Pr[y = 0](p_T^{(M,0)} - (1 - p_T^{(F,0)})) + \Pr[y = 1](p_T^{(M,1)} - (1 - p_T^{(F,1)})).$$

Now it is clear to see that if either  $p_T^{(M,0)} - (1 - p_T^{(F,0)})$  or  $p_T^{(M,1)} - (1 - p_T^{(F,1)})$  are non-negligible, the whole right-hand side will also be non-negligible. Since  $F_k$  and  $M_k$  were assumed to be 1/2-indistinguishable, this provides a contradiction and therefore our original assumption that  $|p_T^{(M,y)} - (1 - p_T^{(F,y)})| \neq \nu_y(k)$  (for either  $y$ ) must be false.

c. We can write  $p_T = \Pr[\text{coin} = 0](\Pr[x = 0]a_0 + \Pr[x = 1]a_1) + \Pr[\text{coin} = 1](\Pr[x = 0](1 - a_1) + \Pr[x = 1](1 - a_0))$ . Since we are using  $B_{2/3}$ , we know that  $\Pr[\text{coin} = 0] = 2/3$  and likewise that  $\Pr[\text{coin} = 1] = 1/3$ . Likewise, we know that  $\Pr[x = 0] = \frac{1}{2} + \frac{1}{k}$  and  $\Pr[x = 1] = 1 - \Pr[x = 0] = \frac{1}{2} - \frac{1}{k}$ . We can plug these probabilities into our first equation to see that  $p_T = \frac{2}{3}((\frac{1}{2} + \frac{1}{k})a_0 + (\frac{1}{2} - \frac{1}{k})a_1) + \frac{1}{3}((\frac{1}{2} + \frac{1}{k})(1 - a_1) + (\frac{1}{2} - \frac{1}{k})(1 - a_0))$ . Expanding this out fully, we end up with  $\frac{1}{3}a_0 + \frac{2}{3k}a_0 + \frac{1}{3}a_1 - \frac{2}{3k}a_1 + \frac{1}{6} - \frac{1}{6}a_1 + \frac{1}{3k} - \frac{1}{3k}a_1 + \frac{1}{6} - \frac{1}{6}a_0 - \frac{1}{3k} + \frac{1}{3k}a_0$ . Now, sorting like terms together, we find that  $p_T = \frac{1}{3} + (\frac{1}{3}a_0 + \frac{1}{3}a_1 - \frac{1}{6}a_1 - \frac{1}{6}a_0) + (\frac{2}{3k}a_0 + \frac{1}{3k}a_0 - \frac{2}{3k}a_1 - \frac{1}{3k}a_1) = \frac{1}{3} + \frac{a_0 + a_1}{6} + \frac{a_0 - a_1}{k}$ , as desired.

d. First, we can rearrange our equation to express it in terms of  $k$ . We start by multiplying both sides by  $k$ , which gives us  $kp_T = \frac{k}{3} + \frac{k(a_0 + a_1)}{6} + a_0 - a_1$ . Now, we put all the terms with a factor of  $k$  on the same side and pull out  $k$ , giving us  $k(p_T - \frac{1}{3} - \frac{a_0 + a_1}{6}) = a_0 - a_1$ . Therefore,  $k = \frac{6(a_0 - a_1)}{6p_T - 2 - a_0 - a_1}$ . Now, we assume that  $p_T > \frac{2}{3}$ . Then

$$k < \frac{6(a_0 - a_1)}{6(\frac{2}{3}) - 2 - a_0 - a_1} = \frac{6(a_0 - a_1)}{2 - a_0 - a_1} \leq 6$$

since we have that  $a_0, a_1 \leq 1$  and therefore  $a_0 - a_1 \leq 2 - a_1 - a_0$ . But, we know that  $k \geq 6$  so this is a contradiction, which tells us that we must have  $p_T \leq \frac{2}{3}$ .

e. Repeating our logic from part (c) (and using  $c$  to represent the coin), we have the equation  $p_T = \Pr[c = 0](\Pr[x = 0]a_0 + \Pr[x = 1]a_1) + \Pr[c = 1](\Pr[x = 0](1 - a_1 + \nu_1) + \Pr[x = 1](1 - a_0 + \nu_0))$ . We can substitute in the same known probabilities of  $\Pr[c = 0] = \frac{2}{3}$ ,  $\Pr[c = 1] = \frac{1}{3}$ ,  $\Pr[x = 0] = \frac{1}{2} + \frac{1}{k}$ , and  $\Pr[x = 1] = \frac{1}{2} - \frac{1}{k}$  and reduce (I won't repeat all the steps here) to get  $p_T = \frac{1}{3} + \frac{a_0 - a_1}{k} + \frac{a_0 + a_1}{6} + \frac{1}{3}(\frac{1}{2} + \frac{1}{k})\nu_1 + \frac{1}{3}(\frac{1}{2} - \frac{1}{k})\nu_0$ . Now, we use the same algebraic manipulations from part (d) to see that  $p_T \leq \frac{2}{3} + \frac{1}{3}(\frac{1}{2} + \frac{1}{k})\nu_1 + \frac{1}{3}(\frac{1}{2} - \frac{1}{k})\nu_0$ . Since both the  $\nu_i$  are negligible, this whole second term is negligible as well and therefore our two sets are 2/3-indistinguishable.