1. Let $(S, \mathcal{S}, P)$ be any probability space. Show that if $\{E_j\}$ is a partition of S with $P(E_j) > 0$, then for any events A, B:

$$P(A\,|B) = \sum_j P(A\,|B \cap E_j)P(E_j\,|B)$$

since $\{E_i\}$ is a partition, $A \cap B = [\underbrace{\cup_i E_i}_{S}] \cap [A \cap B] = \cup_i \underbrace{(A \cap B \cap E_i)}_{\text{disjoint}}$

then,

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

$$= \frac{P[\cup_i (A \cap B \cap E_i)]}{P(B)}$$

$$= \frac{\sum_i P[A \cap (B \cap E_i)]}{P(B)} \quad \text{(countable additivity)}$$

$$= \sum_i \frac{P(A \mid B \cap E_i)\, P(B \cap E_i)}{P(B)}$$

$$= \sum_i P(A \mid B \cap E_i) \underbrace{\frac{P(B \cap E_i)}{P(B)}}_{P(E_i|B)}$$

$$= \sum_i P(A|B \cap E_i)\,P(E_i|B)$$

in addition to the scores for correctness of individual problems:

- check whether problem solutions are numbered and in order and award…
0 pt: not at all
1/2 pt: mostly
1 pt: completely

- make an assessment of the clarity of presentation of the solutions checked and award…
0 pt: difficult to follow
1 pt: about average
2 pt: exceptionally clear

- record, not including problem 5, whether all problems were attempted and award…
0 pt: not all problems attempted
1 pt: all problems attempted

sum up points and award a score out of 4 for organization, clarity, and completeness

2. Sickle-cell disease is an inherited condition that causes pain and damage to organs and muscles. It is recessive: people with two copies of the relevant allele have the disease, but people with only one copy are healthy. That is, if A is the sickle-cell allele and a is the neutral allele, people with AA have the disease, and people with Aa or aa do not. Suppose that for some study population the probability that an individual has each combination is as shown below.

$$P(AA) = 0.02$$
$$P(Aa) = 0.18$$
$$P(aa) = 0.8$$

Assume that parents reproduce independently of their having any of these genotypes and inherited alleles are selected independently and with equal probability from each parent.

i. What are the unique combinations of parent genotypes?
ii. For each possible combination from (i), find the probability that a child has sickle-cell disease given parent genotypes.
iii. Find the probability that a child from this population has sickle-cell disease.

[redacted]

ii. note that $P(AA \mid (p_1, p_2)) = P(A \text{ from } p_1 \cap A \text{ from } p_2) = P(A \text{ from } p_1)\, P(A \text{ from } p_2)$

parent 1,2 — independence of allele selection

and $P(A \text{ from } p_i) = \begin{cases} 0, & p_i = aa \\ 1/2, & p_i = Aa \\ 1, & p_i = AA \end{cases}$

so $P(AA \mid (p_1, p_2))$ is:

p_2 \ p_1	AA	Aa	aa
AA	1	1/2	0
Aa	1/2	1/4	0
aa	0	0	0

iii. now note $P((p_1, p_2)) = P(p_1) P(p_2)$, so

independence of parent selection

$$P(AA) = \sum_{p_1, p_2} P(AA \mid (p_1, p_2))\, P(p_1) P(p_2)$$
$$= 1 \cdot (0.02)^2 + 2 \cdot \tfrac{1}{2} \cdot (0.02) \cdot (0.18) + \tfrac{1}{4} \cdot (0.18)^2$$
$$= 0.0121$$

score out of 8

for (i), answers should
- (2pt) enumerate all 9 combinations (see table in (ii))
  (award 1pt partial credit if order is ignored)

for (ii), answers should:
- (1pt) consider each case (parent combination) identified in (i)
- (1pt) use independence of allele selection to obtain P(AA) for each case
- (1pt) express probabilities for *at least* the non-zero cases
- (1pt) be numerically correct

for (iii), answers should:
- (1pt) apply the law of total probability consistent with answers in (ii)
- (1pt) use independence of parent couplings to obtain probabilities for each case

score out of 9

for (i), answers should use total probability OR multiplication rule (1pt)

for (ii), answers should be consistent calculations with result in (i) (1pt)

for (iii), answers should
- (1pt) apply bayes' rule
- (1pt) use correct probabilities for each piece

for (iv), answers should be consistent with result from (iii) (1pt)

for (v), answers should be consistent with result from (iii) (1pt)

for (vi), answers should:
- (1pt) set up an inequality using result in (iii)
- (1pt) rearrange to solve for the correct term (1 - b)
- (1pt) obtain a numerically correct final answer

3. Suppose you are analyzing a diagnostic test for a condition that appears in 5% of the population. Let T_+, T_- denote the events that an individual obtaining a positive or negative test result, respectively, and C_+, C_- denote the events that an individual is a positive or negative case. Represent the detection rates for each type of case as follows:

$$P(T_+ | C_+) = a \qquad \text{(true positive rate)}$$
$$P(T_- | C_+) = 1 - a \qquad \text{(false negative rate)}$$
$$P(T_+ | C_-) = 1 - b \qquad \text{(false positive rate)}$$
$$P(T_- | C_-) = b \qquad \text{(true negative rate)}$$

i. Find the probability that the test correctly diagnoses a randomly chosen individual from the population in terms of a, b.
ii. If the test achieves an overall accuracy of 90% — so $P(\text{test result is correct}) = \frac{9}{10}$ — and the true positive rate is $a = \frac{8}{10}$, what is the true negative rate b?
iii. Find $P(C_+ | T_+)$ and $P(C_- | T_-)$ in terms of a, b.
iv. If the true positive rate is $a = 0.8$, what true negative rate b produces a test for which $P(C_+ | T_+) = 0.9$?
v. Under the scenario in ~~(ii)~~ (iv), what is $P(C_- | T_-)$?
vi. Suppose the test is being redesigned to ensure that $P(C_+ | T_+) \geq \frac{9}{10}$. If the best possible true positive rate for the test is $a = \frac{9}{10}$, what is the maximum false positive rate that achieves the redesign goal?

i. $P(\text{correct}) = P(\text{correct} | C_+) P(C_+) + P(\text{correct} | C_-) P(C_-)$

$= P(T_+ | C_+) P(C_+) + P(T_- | C_-) P(C_-)$

$= a \cdot \frac{5}{100} + b \, \frac{95}{100}$

ii. here $a = \frac{8}{10}$, $P(\text{correct}) = \frac{9}{10}$, so

$b = P(\text{correct}) - a \cdot \frac{5}{100} = \frac{9}{10} - \frac{8}{10} \cdot \frac{5}{100} = \frac{86}{100}$

iii. $P(C_+ | T_+) = \dfrac{P(T_+ | C_+) P(C_+)}{P(T_+ | C_+) P(C_+) + P(T_+ | C_-) P(C_-)}$

$= \dfrac{a \cdot 5/100}{a \cdot 5/100 + (1-b)\, 95/100}$

$= \dfrac{1}{1 + \frac{95}{5} \cdot \frac{1-b}{a}}$

$P(C_- | T_-) = \dfrac{P(T_- | C_-) P(C_-)}{P(T_- | C_-) P(C_-) + P(T_- | C_+) P(C_+)}$

$= \dfrac{b \cdot 95/100}{b \cdot 95/100 + (1-a)\, 5/100}$

$= \dfrac{1}{1 + \frac{5}{95} \cdot \frac{1-a}{b}}$

iv. $\dfrac{9}{10} = \dfrac{1}{1 + \frac{95}{5} \cdot \frac{1-b}{8/10}}$

$\dfrac{10}{9} = 1 + 190(1-b)/8$

$\dfrac{80}{9} = 1 + 190(1-b)$

$\dfrac{71}{9 \cdot 190} = 1 - b$

$b = 1 - \dfrac{71}{1710} \approx 0.958$

v. $P(C_- | T_-) = \dfrac{1}{1 + \frac{1}{19} \cdot \frac{0.2}{0.958}} \approx 0.99$

vi. $P(C_+ | T_+) \geq 9/10 \iff \dfrac{1}{1 + \frac{95}{5} \frac{1-b}{a}} \geq 9/10$

$1 + \frac{95}{5} \left(\frac{1-b}{a} \right) \leq \frac{10}{9}$

$\frac{1-b}{a} \leq \frac{1}{9} \cdot \frac{5}{95} = \frac{1}{171}$

$1 - b \leq \frac{a}{171}$

so if $a = 9/10$, $P(T_+ | C_-) = 1 - b \leq \frac{9}{1710} = 0.0053$