Quantitative Genomics and Genetics - Spring 2016
BTRY 4830/6830; PBSB 5201.01

Homework 1 (version 1) - Key

Assigned February 3; Due 11:59PM February 8

Problem 1 (Easy)

Consider a coin (system) that you would like to learn about and two types of experiments: 1. One flip of the coin (Experiment 1), and Two flips of the coin (Experiment 2). Consider the case where you are going to perform TWO experimental trials for Experiment 1 and ONE experimental trial for Experiment 2.

a. Write out BOTH the sample spaces AND the Sigma Algebras for BOTH Experiments 1 and 2.

Experiment 1: $\Omega = \{H, T\}$, σ-algebra= $\emptyset, \{H\}, \{T\}, \{H, T\}$
Experiment 2: $\Omega = \{HH, HT, TH, TT\}$
σ-algebra= $\emptyset, \{HH\}, \{HT\}, \{TH\}, \{TT\}$
$\{HH, HT\}, \{HH, TH\}, \{HH, TT\}, \{HT, TH\}, \{HT, TT\}, \{TH, TT\}$
$\{HH, HT, TH\}, \{HH, HT, TT\}, \{HH, TH, TT\}, \{HT, TH, TT\}, \{HH, HT, TH, TT\}$

b. Using one sentence at most, explain why the sets $\{H_1, T_2\}$ describing the sample of a 'heads' on the first trial and a 'tails' on the second trial, resulting from the two trials of Experiment 1, is distinct from the set $\{HT\}$ describing the results of one trial of Experiment 2.

Among the acceptable answers: The two flips in Experiment 2 are coupled together such that it is not possible to consider the probability of each flip separately, i.e., the result 'HT' is a single possible outcome among four total. For Experiment 1, $\{H_1, T_2\}$ is a result of two experimental trials where the result of each could be 'H' or 'T', where the probability of obtaining either of these outcomes for each of the trials, could be considered.

c. Answer the following questions by considering possible sample outcomes for Experiment 1 (two total trials). What is $\{H_1, H_2\} \cap \{H_1, T_2\}$?

$\{H_1, H_2\} \cap \{H_1, T_2\} = \{H_1\}$

c. Answer the following questions by considering possible sample outcomes for Experiment 2 (one total trial). What is $\{HH\} \cap \{HT\}$?

$\{HH\} \cap \{HT\} = \emptyset$

Problem 2 (Medium)

Consider a coin that you plan to learn about with the following experiment: 3 flips of the coin.

a. Write out the sample space of this experiment.

$\Omega = \{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT\}$

b. For the (appropriate) Sigma Algebra of this experiment, how many subsets will be the null / empty set? How many subsets will have one element? How many subsets will have two elements? How many subsets will have eight elements? Note that you DO NOT need to write out these sets or the entire Sigma Algebra!

σ-algebra subsets with the empty set: 1 (the empty set is unique)
σ-algebra subsets with one element: 8 (the number of element in the sample space)
σ-algebra subsets with two elements: 28 (calculate by starting with HHH and considering all pairs, then HHT and all pairs excluding HHH, etc. producing 7+6+5+4+3+2+1 or you could use the formula $\binom{8}{2} = \frac{8!}{(2!)(6!)}$)
σ-algebra subsets with eight elements: 1 (the set of the entire sample space)

c. Define a probability model such that each of the subsets of the Sigma Algebra that have ONE element each are assigned THE SAME probability. Explain why from this starting point, you can calculate the probabilities of all of the other subsets of the Sigma Algebra. Note that you DO NOT need to write out these probabilities for the other subsets of the Sigma Algebra - just explain your reasoning (which may include an example or two).

$Pr(HHH) = Pr(HHT) = Pr(HTH) = Pr(HTT) = Pr(THH) = Pr(THT) = Pr(TTH) = Pr(TTT) = 0.125$

From the third Axiom of probability, we know the probability of the union of disjoint sets is the sum of the probability of each of the sets. Since all of the other sets in the sigma algebra are a union of (some) combination of these sets, we can calculate the probability of each (and note that the probability of the empty set is always zero), e.g., $Pr(\{HHH, HHT, HTH\}) = Pr(HHH \cup HHT \cup HTH) = Pr(HHH) + Pr(HHT) + Pr(HTH) = 0.375$

d. Define the random variables X_1 that is the 'number of heads on the first flip' and X_2 that is the 'number of heads on the second flip'. For X_1 and X_2, write out the values that these random variables take for each of the possible sample outcomes that could result from a single experimental trial.

$X_1(HHH) = 1, X_1(HHT) = 1, X_1(HTH) = 1, X_1(HTT) = 1, X_1(THH) = 0, X_1(THT) = 0, X_1(TTH) = 0, X_1(TTT) = 0$

$X_2(HHH) = 1, X_1(HHT) = 1, X_1(HTH) = 0, X_1(HTT) = 0, X_1(THH) = 1, X_1(THT) = 1, X_1(TTH) = 0, X_1(TTT) = 0$

e. What is $Pr(X_1 = 1)$? What is the $Pr(X_2 = 1)$? Explain how you arrived at these answers.

$Pr(X_1 = 1) = 0.5$ and $Pr(X_2 = 1) = 0.5$. In each case, half of the outcomes mapped to '1' and since the probability of each outcome is the same, the probability of each of these variables being '1' is 0.5.

f. What is $Pr(X_2 = 1|X_1 = 1)$? Show your work by making use of the formula for conditional probability.

$Pr(X_2 = 1|X_1 = 1) = \frac{Pr(X_2=1 \cap X_1=1)}{Pr(X_1=1)} = \frac{0.25}{0.5} = 0.5$

Note the numerator is the sum of the probabilities of the two occurrences 'HHH' and 'HHT'.

g. Explain why your answer to part 'f' is consistent with X_1 and X_2 being independent and why this makes sense given the probability model.

If the two random variables are independent, then $Pr(X_2 = 1|X_1 = 1) = 0.05 = Pr(X_2 = 1) = 0.5$.

That these random variables are independent makes sense, given that there are an equal number of outcomes that have 'HH', 'HT', 'TH', and 'TT' for the first two flips and each of the possibilities has an equal probability of occurring, such that knowing that the first flip is 'H' provides no information on what the second flip will be.

h. Define a new the random variable X_3 that is 'the number of heads observed for all three flips'. Write out the values that this random variable takes for each of the possible outcomes of a single experimental trial.

$X_3(HHH) = 3, X_3(HHT) = 2, X_3(HTH) = 2, X_3(HTT) = 1, X_3(THH) = 2, X_3(THT) = 1, X_3(TTH) = 1, X_3(TTT) = 0$

i. What is $Pr(X_3 = 3|X_1 = 1)$? Show your work by making use of the formula for conditional probability.

$Pr(X_3 = 3|X_1 = 1) = \frac{Pr(X_3=3 \cap X_1=1)}{Pr(X_1=1)} = \frac{0.125}{0.5} = 0.25$

Note the numerator is the probability of the single occurrences 'HHH'.

j. Explain why your answer to part 'i' is consistent with X_1 and X_3 NOT being independent and why this makes sense intuitively given how these random variables are defined.

The two random variables are not independent, since $Pr(X_3 = 3|X_1 = 1) = 0.25 \neq Pr(X_3 = 1) = 0.125$.

That these random variables are not independent makes sense, since the value of X_3 depends in part on whether there is a 'H' on the first flip, such that knowing X_1 is '1' (i.e., first flip is a 'H') provides information that some (none zero probability) values of X_3 are not possible, changing the probability of the values X_3 can take.

Problem 3 (Difficult)

a. Consider a system and associated experiment where the sample space is the set of Natural numbers. Make use of the Axioms of Probability to show why we cannot define a probability function on the (appropriate) Sigma Algebra of this sample space that assigns an equal probability to each (non-empty set) subset of the Sigma algebra that has a single element.

If $\Omega = \mathbb{N} = \{1, 2, 3, ...\}$ there are a countably infinite number of disjoint outcomes. There are therefore countably infinite subsets in the Sigma Algebra. From the third Axiom, we know that the probability of the union of disjoint sets is the sum of the probabilities. If we assigned an equal probability to each of the natural numbers, no matter how small we made this probability, the probability of a subset in the Sigma Algebra with countably infinite elements would be infinite, which is greater than one, violating the second Axiom. If we assigned a probability of zero to each of the natural numbers, we would also violate the second axiom.

Note: not necessary for full credit but another answer would make use of a test of series convergence such as a comparison test, e.g., for all n the sequence $c * a_n = c * (a_1 + a_2 + a_3 + ...) = c * (1 + \frac{1}{2} + \frac{1}{3} + ...)$ is less than or equal to $c * b_n = c * (b_1 + b_2 + b_3 + ...) = c * (1 + 1 + 1 + ...)$ and since a_n diverges (i.e., goes to infinite), so does b_n, where this holds for $c > 0$ since any non-zero number times infinite, is infinite.

b. Consider a system and associated experiment where the sample space is the subset of the Reals from zero to one (including zero and one) and the number 2 (i.e., $\Omega = [0, 1] \cup 2$). Consider a random variable X on this sample space that is the identity function. Write out the equation of the probability distribution function of X where the probability of each interval within $[0, 1]$ of the same size has an equal probability AND where the probability of the entire interval $[0, 1]$ and the number 2 is the same. Also, write out the equation for the cumulative distribution function for X for the probability distribution function just described.

Note that in both cases, the distribution equations will not be in 'closed' form (i.e., you cannot write them out as a single equation), such that you will need to write out the equation in two pieces, part for $X \in [0, 1]$ and part for $X \in 2$. On a side note, what you are writing out a 'mixture' distribution for X, which is a mixture of continuous and discrete.

$f_X(x) = 0.5$ for $x \in [0, 1]$; $f_X(x) = 0.5$ for $x \in [2]$

$F_X(x) = 0$ for $x \in (-\infty, 0)$;
$F_X(x) = 0.5x$ for $x \in [0, 1)$;
$F_X(x) = 0.5$ for $x \in [1, 2)$;
$F_X(x) = 1.0$ for $x \in [2, +\infty)$

Note that since the question states (incorrectly) that both equations will have two pieces, any attempt to write the cdf in two pieces (that gets the critical elements) will be assigned full credit.