Discrete Multivariate Modeling (SySc 551/651) Professor M. Zwick
MIDTERM EXAM Fall 2019 (Oct 30, 2019)

60 points; 2 hours; closed book, no notes. *Answer questions on exam sheets, and put **your name** on them.* You can use $\Gamma(a, b...) = -a \log a - b \log b -...$

I. BASICS [32 points]

1. [14 points]

(a) A deck of cards has 13 ranks (Ace, 2-10, Jack, Queen, King) and 4 suits (hearts, diamonds, clubs, spades), 52 different cards in all. I pick a card. Telling you (honestly) that I picked a diamond reduces your uncertainty by how much? (Circle one; 2 pts)

i. $\log_2(52/13)$ ii. $\log_2(52/4)$ iii. $\log_2(13/4)$
iv. $\log_2(13/52)$ v. $\log_2(4/52)$ vi. $\log_2(4/13)$

(b) Shannon entropy can be interpreted as diversity when probabilities are fractional proportions. Consider four counties of Oregon with fractions of employment in three economic sectors (A, B, C), as given below. D(1), D(2), D(3), D(4) are the employment diversities of the four counties. The ordinal ranking of these four diversity magnitudes is: D( ) > D( ) > D( ) > D( ): Put the right numbers in the parentheses on the left. (2 pts)

	Sector			
County	A	B	C	Diversity
1	.50	.50	.00	D(1)
2	.25	.25	.50	D(2)
3	.33	.33	.33	D(3)
4	.00	.00	1.00	D(4)

(c) For the following contingency table, with probabilities a+b+c+d=1, if X and Y are independent, then (circle all that are true; 2 pts)

	Y_1	Y_2
X_1	a	b
X_2	c	d

i. a/b = c/d ii. a/c = b/d iii. a/d = b/c iv. none of the above

(d) Let the level of knowledge of a student about information theory be a dichotomous variable, K, whose value is high (h) or low (l). Assume that an exam testing students on their knowledge makes only a binary discrimination, i.e., the grade, G, is also either h or l. The probabilities of possible situations give the following table, where a+b+c+d = 1:

		G: l	G: h
K	l	a	b
	h	c	d

(d1) Write an expression in terms of probabilities a, b, c, d for H(K|G), the uncertainty which remains about a student's knowledge when his/her grade is known. (2 pts)

(d2) It would clearly be desirable that H(K|G) be zero, i.e., that there were no uncertainty about a student's actual knowledge once his/her grade is known. If this were true, what must be true about the numerical values of the probabilities a, b, c, d? (2 pts)

(d3) Write an (equality and/or inequality) equation involving *only* the univariate uncertainties, H(K), H(G), and the transmission, T(K:G), which would hold if the exam not only resolved all uncertainty about K but was actually *more detailed and complex* than needed to resolve this uncertainty. (2 pts)

(e) Circle all choices which give a correct expression for transmission, T(X:Y) in terms of the following uncertainties and conditional uncertainties of x and y (2 pts):

i. H(x|y) + H(y|x) – H(x,y)
ii. H(x,y) – H(x|y) – H(y|x)
iii. H(x,y) + H(x|y) + H(y|x)
iv. H(x) – H(x|y)
v. H(x) – H(y|x)
vi. H(y) – H(y|x)
vii. H(y) – H(x|y)
viii. H(x|y) + H(y|x)
ix. H(x) + H(y) – H(x,y)
x. H(x,y) – H(x) – H(y)

2. [8 points] Let the data be the table below, with known probabilities, a…h.

	Z:	0		1	
	B:	0	1	0	1
A:	0	a	b	c	d
	1	e	f	g	h

(a) Give an expression for T(A:Z) in terms of parameters, a…h, by using the definition of T(A:Z) in terms of model and data entropy values (2 pts).

(b) Give an expression for T(A:Z) in terms of parameters, a…h, using the definition of T(A:Z) as a sum of p log (p/q) terms. Give only the *first two terms* of this sum (2 pts).

(c) Give an expression for T_B(A:Z) in terms of parameters, a…h (2 pts).

(d) Let the DV = Z and the IVs = A, B, C. If one interprets "the predictive effect of A on Z, *controlling for* B and C" to mean how much knowing A tells us about Z beyond what B and C tells us about Z, which transmission value gives the magnitude of the predictive effect of A, controlling for B and C? (circle one; 2 pts)

i. T(BCZ:AZ) ii. T(ABC:AZ) iii. T(ABC:BCZ) iv. None of these

3. [4 points] Consider relations defined over time.
(a) For an AB table where A = x(t) and B = x(t+1), which of the following, if true, implies that AB system *must* be stochastic (not deterministic)? (circle one; 2 pts)

i. H(A) = H(B)	ii. H(A) > H(B)	iii. H(B) > H(A)
iv. H(A|B) = 0	v. H(B|A) = 0	vi. None of these

(b) Let time be a variable, as follows. Consider ABC_1 and ABC_2 distributions below, where C_1 = t and C_2 = t+1, with probabilities, a+b+c+d = 1 and e+f+g+h = 1. Treating the AB distribution at time t as the reference, I wish to test the hypothesis that the AB distribution at t+1 is the same as it was earlier at t. Using the p log (p/q) expression for transmission, write the T in terms of a…h that will let me test this hypothesis. (2 pts)

C_1

	B_1	B_2
A_1	a	b
A_2	c	d

C_2

	B_1	B_2
A_1	e	f
A_2	g	h

T =

4. [6 points]
(a) For neutral system ABCD, I consider a disjoint model AB:CD. It is always the case that (circle one; 2 pts):

i. T(A:B) + T(C:D) = T(AB:CD)
ii. T(A:B) + T(C:D) + T(AB:CD) = 0
iii. T(A:B) + T(C:D) + T(AB:CD) = T(ABCD)
iv. T(A:B) + T(C:D) + T(AB:CD) = T(A:B:C:D)
v. None of these

(b) True or false? (circle; 2 pts): For a directed ABZ system, the entropy of Z equals the transmission between Z and one predictor, plus the transmission given this predictor between Z and the other predictor, plus an unexplained (not reduced) entropy, i.e., $H(Z) = T(A{:}Z) + T_A(B{:}Z) + H_{AB}(Z)$. (A proof of this assertion or its negation is *not* required for this question.)

(c) The following table is for a 3-variable Borromean rings, where there is non-zero constraint in ABC, but no constraint in any of the bivariate projections.

	C_1		C_2	
	B_1	B_2	B_1	B_2
A_1	.25	0	0	.25
A_2	0	.25	.25	0

It is possible to construct a 4-variable analog to the above table. Give this analog by filling in the following table with numerical values. (2 pts)

	D_1				D_2			
	C_1		C_2		C_1		C_2	
	B_1	B_2	B_1	B_2	B_1	B_2	B_1	B_2
A_1								
A_2								

II. STRUCTURES [28 points]

5. [7 points]
(a) For some ABZ data, $T_A(B:Z) = 0$. Circle all models where T(model) is also 0. (2 pts)

i. ABZ	ii. AB:AZ:BZ	iii. AB:AZ	iv. AB:BZ	v. AZ:BZ
vi. AB:Z	vii. AZ:B	viii. BZ:A	ix. A:B:Z	x. none of these

(b) Consider a directed system in which A, B, and C are IVs, and Z is a DV. List all the specific structures (not general structures merely illustrated by specific structures), i.e., list all the structures that one would need to consider in modeling actual data. *Circle* all structures which do *not* have loops (5 pts).

6. [8 points]
(a) Let CBWAITO mean "cannot be written algebraically in terms of." Circle all choices that are true for models with loops. (4 pts)

i. H(model) CBWAITO the uncertainties of model components
ii. T(model) CBWAITO uncertainties of the model & the data
iii. *calculated* model probabilities CBWAITO *observed* data probabilities
iv. df(model) CBWAITO df values of model relations & their projections

(b) True or false? (circle; 2 pts): Given the convention that directed systems must have an IV component which includes all the IVs, a directed system model that has two or more ("predicting") relations that involve the same DV must necessarily have a loop.

(c) Suppose a problem has few variables, so it is practical to evaluate all models in the Lattice of Structures, and I use the data as my reference in a top-down analysis. I know that model complexity (df) varies monotonically as one goes down the LOS and model error (T) also varies monotonically as one goes down the LOS. True or False? (circle one; 2 pts): The "best model," which would have minimum complexity and minimum error, is easy to choose, since no tradeoff is needed between complexity and error.

7. [8 points] One can sometimes come up with *multiple* good models. It would be handy to have a measure for the similarity or difference between any two models, m_1 and m_2 where this measure would be 0 if the models have nothing in common and 1 if the models were identical. Consider a measure constructed from two quantities, $f(m_1 \cup m_2)$ and $f(m_1 \cap m_2)$, where f is some suitable property of these models gotten by taking the union and intersect.

(a) Which of the following ratios of these quantities varies *directly*, not inversely, with the degree of *similarity* of the two models? (2 pts, circle one)

i. $f(m_1 \cup m_2) / f(m_1 \cap m_2)$ ii. $f(m_1 \cap m_2) / f(m_1 \cup m_2)$
iii. both iv. neither

(b) Actually, one wants *two* different types of measures: (i) a measure of similarity between two structures *without regard to data*, and (ii) a measure of similarity between models, where 'model' means a structure *applied to some actual data*.

(b1) What is a good choice for f for a data-*independent* measure of the similarity or dissimilarity of the two structures? (circle one; 2 pts)

i. H ii. df iii. Neither of these

(b2) What is a good choice for f for a data-*dependent* measure of the similarity or dissimilarity of the two models? (circle one; 2 pts)

i. H ii. df iii. Neither of these

(c) For NCA = nearest common ancestor & NCD = nearest common descendent, circle the statement below that is true (2 pts):

i. $m_1 \cup m_2 =$ NCA; $m_1 \cap m_2 =$ NCD ii. $m_1 \cup m_2 =$ NCD; $m_1 \cap m_2 =$ NCA iii. Neither

8. [5 points]
(a) p(ABZ) can be written as $p(AB)*p(Z|AB)$. We might use for this the notation $AB:_{AB}Z$ where the subscript to the left means 'conditioned on.' Since p(AB) and p(Z|AB) can be independently specified, df(ABZ) = df(AB) + df(Z|AB). Given that |A|=|B|=|Z|=2,

(a1) How many parameters does it take to specify AB? ___ Give a number. (1 pt)

(a2) How many parameters does it take to specify p(Z|AB)? ___ Give a number. (2 pts)

(b) Klir's equivalence classes, called ρ structures, each represented by an ordinary graph which indicates which variables are directly linked by only one relation, where each class has a most complex C-structure and a least complex P structure, is (circle one; 2 pts)

i. a way to search the lattice for loopless models
iii. a way to search the lattice for disjoint searches
iii. a way to search the lattice hierarchically