Discrete Multivariate Modeling (SySc 551/651) Professor M. Zwick
FINAL EXAM Fall 2018 (12/3/2018)

60 points; 2 hours; closed book, no notes. *Answer question on exam sheets, and put **your name** on them.* You can use $\Gamma(a, b,...) = -a \log a - b \log b - ...$ Do not use red ink.

---

1. [16 points]

(a) True or false? (circle; 2 pts): For the data, $H(Z) - H(Z|AB) = T(A:Z) + T_A(B:Z)$

(b) Let the data be the contingency table below, with known probabilities, a…h. Give an expression for $T(A:Z \mid B_1)$ in terms of parameters, a…h (2 pts).

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

(c) For directed system ABCDZ, I(ABCD:ABCZ→ABCD:BCZ) = H(Z|X) – H(Z|Y), where X and Y are subsets of the IVs. Fill in (2 pts): X = _____ ; Y = _____

(d) Let the probabilities of *micro*states a, b, c, d be .1, .2, .3, .4, respectively, and the probabilities of *macro*states I, II be .3, .7, respectively, where I includes a and b and II includes c and d. Write an expression involving only numbers (probability values) for the entropy of the microstate given the macrostate. (2 pts)

H(microstate | macrostate) =

(e) On the left is an observed probability table (p) for a *directed* system, with sample size N. None of parameters (a...h) is 0. On the right is the calculated table (q) for AB:AZ.

	Z_1		Z_2	
	B_1	B_2	B_1	B_2
A_1	a	b	c	d
A_2	e	f	g	h

	Z_1		Z_2	
	B_1	B_2	B_1	B_2
A_1	q_1	q_2	q_3	q_4
A_2	q_5	q_6	q_7	q_8

(e1) Write an expression in terms of a…h for the amount of constraint that data ABZ *adds* to the constraint that is already captured in the AB:AZ model. (2 pts)

(e2) Write an expression for q_5 in terms of the parameters a through h. (2 pts)

(f) Suppose A is a quantitative variable that needs to be binned. We want to maximize its power to predict Z, but not use degrees of freedom *inefficiently*, which might prevent, for small sample sizes, the use also of B to predict Z. One reasonable *heuristic* is to choose the number of bins for A which maximizes: (circle one; 2 pts)

i. H(Z|A)
ii. H(Z) – H(Z|A)
iii. H(Z|A) * |A|
iv. [H(Z) – H(Z|A)] * |A|
v. H(Z|A) / |A|
vi. [H(Z) – H(Z|A)] / |A|

(g) Suppose the conditional probabilities for model m = AB:AZ:BZ are

A	B	p(AB)	p(Z=1|AB)	p(Z=2|AB)	q_m(Z=1|AB)	q_m(Z=2|AB)
1	1	a	b	c	d	e
1	2	f	g	h	i	j
2	1	k	l	m	n	o
2	2	r	s	t	u	v
		marginals	w	x	y	z

Assume that you know the sample size. In terms of (a…z) in the above table (and any necessary constants), write an expression for transmission T that you would use to test the null hypothesis that the differences between the calculated conditional Z probabilities for (A,B)=(1,1) and the Z marginal probabilities are not statistically significant. (2 pts)

2. [14 points]

(a) For the table given in question 1(e) and for model AB:AZ,

(a1) write the full set of linearly independent constraints in the form M **q** = M **p** by filling in rows in matrix M (starting from the top of M) for the correct number of constraints in this model. In these rows, fill in only the 1's; leave the 0's blank. (The bottom row is the constraint that is true of all models, namely that Σq = Σp.) (2 pts)

Matrix, M

1	1	1	1	1	1	1	1

$$\begin{bmatrix} q_1 \\ q_2 \\ q_3 \\ q_4 \\ q_5 \\ q_6 \\ q_7 \\ q_8 \end{bmatrix} = M \begin{bmatrix} a \\ b \\ c \\ d \\ e \\ f \\ g \\ h \end{bmatrix}$$

(a2) Add one more constraint that is true, that follows from the q(AZ) = p(AZ) constraint (not from the q(AB) = p(AB constraint), as an *additional* row in M. *Circle* this additional row to distinguish it from the rows that answer (a1). Since this constraint is *not* linearly independent of the rows listed for (a1), this added row must be a sum and/or difference of some rows above it. Write the equation for this linear dependence in the form "Row X = Row Y + … – Row Z – … where X is the added row and Y, Z, etc., are rows above it. (2 pts):

To answer the IPF question below, fill in the AB and AZ tables for p and q.

Projected from p

	B:	0	1
A:	0		
	1		

	Z:	0	1
A:	0		
	1		

Projected from q

	B:	0	1
A:	0		
	1		

	Z:	0	1
A:	0		
	1		

Use IPF to obtain q_5 by writing the sequence of values of q_5 in terms of the parameters, a through h, after first imposing AB and then also AZ. The initial value is $q_5^{initial} = 1/8$.

(a3) By imposing AB, we get $q_5^{AB} = q_5^{initial} *$ _____(2 pts)

(a4) By imposing also AZ, we get (2 pts)

$q_5^{AB:AZ} = q_5^{AB} * $ _____

(b) For model ABC:BCD:CDE, write an algebraic expression for q(ABCDE) in terms of p(ABCDE) and its various projections (2 pts).

(c) Suppose I have a state-based model A_1B_2 for an AB neutral system, where A and B have cardinality 2. Let q_1 be the calculated distribution for A_1B_2. I want to pick a second state to add to this state based model which maximizes the information captured by this second state (or equivalently minimizes the error of the two-state model). The two-state model will be A_1B_2:A_iB_j, for some particular i and j. I have three possible choices for this second state, namely A_1B_1, A_2B_1, and A_2B_2. (Ignore the possibility of choosing the second state from the margins.) Call the distribution I get after choosing the second state q_2. I want to pick that state among the three possible choices that maximizes (circle one; 2 pts)

i. $\Sigma\ p \log [q_1/q_2]$	ii. $\Sigma\ p \log [q_2/q_1]$
iii. $\Sigma\ q_1 \log [q_1/q_2]$	iv. $\Sigma\ q_1 \log [q_2/q_1]$
v. $\Sigma\ q_2 \log [q_1/q_2]$	vi. $\Sigma\ q_2 \log [q_2/q_1]$

(d) True or false (circle; 2 pts): A state that is chosen for a state-based model must have a probability (in the data or a projection) *greater* than the probability it would have if all states (in the data or projection) were equally probable, i.e., only *more* probable, not *less* probable, single states are used in RA state-based models.

3. [18 points]

(a) Suppose I pick a best BIC model, then notice that the 2nd best BIC model has a ΔBIC that is only slightly less than the ΔBIC of the best model. I wonder if I should look at the predictions of this 2nd best BIC model and compare them to the predictions of the best BIC model to see if the two models predict similarly or differently. I should definitely make this comparison if which of the following measures is large compared to 1 (2 pts).

i. $I(m_1 \cup m_2 \rightarrow m_{ind}) / I(m_1 \cap m_2 \rightarrow m_{ind})$ ii. $I(m_1 \cap m_2 \rightarrow m_{ind}) / I(m_1 \cup m_2 \rightarrow m_{ind})$

(b) The Transmission of which model in the Lattice of Structures for ABZ will specify the amount of information that A tells me about Z if I already know B? (2 pts).

(c) After each measure from (i) to (viii), write whether the measure is MD (monotonically decreasing or staying the same), MI, (monotonically increasing or staying the same), or NM (not monotonically increasing *or* decreasing) for every step going down the Lattice of Structures. (8 pts)

i. H	ii. T	iii. L^2
iv. α	v. df (not Δdf)	vi. ΔAIC
vii. $\%\Delta H(DV)(train)$	viii. $\%\Delta H(DV)(test)$	

(d) When the reference is the data, one calculates $\Delta df(m_0 \rightarrow m_j)$ and $L^2(m_j)$ and decides on some α_c. Say that going into the Chi-square table with $\Delta df(m_0 \rightarrow m_j)$ and α_c one gets some L^2_c and suppose that the model $L^2(m_j) > L^2_c$. One would then usually (circle; 2 pts)

i. not reject the null, and consider m_j as possibly a good model
ii. not reject the null, and use the data, m_0, as one's model
iii. reject the null, and consider models higher in the lattice than m_j
iv. reject the null and consider models lower in the lattice than m_j

(e) Occam reports AIC relative to the reference model, but in the literature it is more common to cite 'absolute' AIC which equals $-2\ N \sum p \ln q + 2$ df. Using this absolute AIC measure, a good model is one that has (circle one; 2 pts)

i. minimum AIC ii. maximum AIC iii. AIC = α iv. AIC = $1 - \alpha$

(f) Assume that $q(Z_0|A_iB_j)$ and $q(Z_1|A_iB_j)$ for model m_1 = AB:AZ:BZ are almost equal and that the frequency of A_iB_j is also small, so confidence in predicting Z is low. In this case, it's useful to have a simpler backup model, m_2, and make predictions of Z for A_iB_j from m_2. For a good backup model, $I(m_1 \rightarrow m_2)$ (circle one; 2 pts)

i. should be small ii. should be large iii. is irrelevant

4. [8 points]

Below are results of analyzing some genomic medical data for model IV:AZ:EZ:KZ, where this table has been augmented with three p-value calculations. Z_1 is the diseased state. p and q conditional probabilities are in %. The p-values are as follows:

p_{rule} compares $q(Z|A_jE_kK_l)$ to (50, 50).
$p_{margins}$ compares $q(Z|A_jE_kK_l)$ to the $(q(Z_0), q(Z_1))$ margins, i.e., to (51.8, 48.2).
p_A compares $q(Z|A_jE_kK_l)$ to $q(Z|A_j)$, i.e., to (64.8, 35.2) for j=0 and (28.1, 71.9) for j=1.

IV				Data obs p(Z | IV)		Model calc q(Z | IV)				correct			
A	E	K	freq	Z_0	Z_1	Z_0	Z_1	rule	p_{rule}	#	%	p_{margin}	p_A
0	0	0	4	0.0	100.0	12.2	87.8	1	0.131	4	100.0	0.113	0.028
0	0	1	8	12.5	87.5	12.4	87.6	1	0.033	7	87.5	0.026	0.002
0	0	2	4	25.0	75.0	29.4	70.6	1	0.409	3	75.0	0.369	0.138
0	1	0	31	64.5	35.5	61.6	38.4	0	0.198	20	64.5	0.277	0.707
0	1	1	37	62.2	37.8	61.9	38.1	0	0.147	23	62.2	0.219	0.714
0	1	2	23	78.3	21.7	82.7	17.3	0	0.002	18	78.3	0.003	0.072
0	2	0	66	63.6	36.4	64.0	36.0	0	0.023	42	63.6	0.047	0.894
0	2	1	61	65.6	34.4	64.4	35.7	0	0.025	40	65.6	0.050	0.942
0	2	2	33	84.8	15.2	84.2	15.8	0	0.000	28	84.8	0.000	0.020
1	0	0	1	0.0	100.0	2.6	97.4	1	0.343	1	100.0	0.325	0.571
1	0	1	7	14.3	85.7	2.6	97.4	1	0.012	6	85.7	0.009	0.134
1	0	2	2	0.0	100.0	7.4	92.6	1	0.228	2	100.0	0.208	0.514
1	1	0	13	30.8	69.2	23.4	76.6	1	0.055	9	69.2	0.041	0.709
1	1	1	24	16.7	83.3	23.7	76.3	1	0.010	20	83.3	0.006	0.633
1	1	2	11	54.5	45.5	47.8	52.2	1	0.884	5	45.5	0.789	0.146
1	2	0	32	21.9	78.1	25.4	74.6	1	0.005	25	78.1	0.003	0.732
1	2	1	39	25.6	74.4	25.6	74.4	1	0.002	29	74.4	0.001	0.735
1	2	2	17	52.9	47.1	50.4	49.6	0	0.973	9	52.9	0.908	0.040
			413	51.8	48.2	51.8	48.2	0		291	70.5		
-	-	-				50.0	50.0						
0	-	-				64.8	35.2						
1	-	-				28.1	71.9						

In (a)-(d), mark a state with X only if its deviation from its reference *is statistically significant*, using the criterion of $p \leq 0.05$. In the tables below (2 pts each),
(a) In columns labeled 1, mark with X all IV states whose prediction *rules* are significant.
(b) In columns labeled 2, mark with X all IV states where $q(Z|IV) \neq$ margins.
(c) In columns labeled 3, mark with X all IV states that ***increase*** the risk of disease *beyond what is predicted by A alone*.
(d) In columns labeled 4 mark with X all IV states that ***decrease*** the risk of disease *beyond what is predicted by A alone.*

A	**E**	**K**	**1**	**2**	**3**	**4**
0	0	0				
0	0	1				
0	0	2				
0	1	0				
0	1	1				
0	1	2				
0	2	0				
0	2	1				
0	2	2				

A	**E**	**K**	**1**	**2**	**3**	**4**
1	0	0				
1	0	1				
1	0	2				
1	1	0				
1	1	1				
1	1	2				
1	2	0				
1	2	1				
1	2	2				

5. [4 points]

(a) For the relation below, what is the *simplest* structure that this relation can be decomposed into with no error? If it can be decomposed into *multiple* structures all equally simple, state all the structures. If it cannot be decomposed at all, say so. (2 pts)

X	Y	Z
0	0	0
1	1	1

(b) The above relation is an example of the smallest number of tuples (2) that a relation between binary variables can have. (If the relation had only one tuple, the variables wouldn't in fact be variables but constants). The *maximum* number of tuples that a relation between binary variables X, Y, and Z can have and still have non-zero constraint is 7. A relation having this number of tuples is (circle one; 2 pts):

i. decomposable ii. non-decomposable iii. either; it depends on the relation