Discrete Multivariate Modeling (SySc 551/651) Professor M. Zwick
MIDTERM EXAM Winter 2008 (Feb. 12, 2008)

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

**1. [8 points]**
Consider the contingency below, where $a + b + c + d + e + f + g + h = 1$.

| | $C_1$ | | $C_2$ | |
|---|---|---|---|---|
| | $B_1$ | $B_2$ | $B_1$ | $B_2$ |
| $A_1$ | a | b | e | f |
| $A_2$ | c | d | g | h |

(a) Write an expression for T(A:C) in terms of a…h (2 pts).

(b) Write an expression for $T_B$(A:C) in terms of a…h (2 pts).

(c) Given some ABC data (circle one; 2 pts):

i. $T(A:C) \geq T_B(A:C)$ ii. $T(A:C) \leq T_B(A:C)$ iii. either i or ii could be true

(d) True or false? (circle; 2 pts): For a directed ABZ system, the uncertainty 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) uncertainty, i.e., $u(Z) = T(A:Z) + T_A(B:Z) + u_{AB}(Z)$. (You can, if you wish, try to prove this assertion or its negation, but a proof is not required here.)

**2. [12 points]** Consider the following table. Give *numerical* answers (in bits) for (a)-(g).

| | $C_1$ | | $C_2$ | |
|---|---|---|---|---|
| | $B_1$ | $B_2$ | $B_1$ | $B_2$ |
| $A_1$ | .25 | 0 | 0 | .25 |
| $A_2$ | 0 | .25 | .25 | 0 |

Calculate the uncertainties of the relations in the *Lattice of Relations*:

(a) H(ABC) =_____(2 pts)

(b) H(AB) =_____; H(AC) =_____; H(BC) =_____(2 pts)

(c) H(A) =_____; H(B) =_____; H(C) =_____(1pt)

Calculate now the uncertainties of the models in the *Lattice of Structures*:

(d) H(AB:AC:BC) =_____(3 pts) [In general, one can't directly calculate entropy for models with loops, but *in this particular case*, one can know its value without doing IPF.]

(e) H(AB:BC) =_____; H(AC:AB) =_____; H(BC:AC) =_____(1pt)

(f) H(AB:C) =_____; H(AC:B) =_____; H(BC:A) =_____(1pt)

(g) H(A:B:C) =_____(1 pt)

(h) The above table is for 3-variable "Borromean rings." It is possible to construct a 4-variable analog, where removing any ring leaves the remaining rings not constrained. By analogy to the above table, and getting a hint from the previous H calculations, what would the table for the 4-variable analog look like? Fill in the following table. (1 pt)

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | $D_1$ | | | | $D_2$ | | | |
| | $C_1$ | | $C_2$ | | $C_1$ | | $C_2$ | |
| | $B_1$ | $B_2$ | $B_1$ | $B_2$ | $B_1$ | $B_2$ | $B_1$ | $B_2$ |
| $A_1$ | | | | | | | | |
| $A_2$ | | | | | | | | |

**3. [8 points]**

(a) What is the nearest common ancestor of ABC:D and AB:BC:CD:DA? (2 pts)

(b) What is the nearest common descendant of ABC:D and AB:BC:CD:DA? (2 pts)

(c) Consider a directed system in which A, B, and C are IVs that affect (or predict) Z, the DV. What *two specific structures* would one compare to find out if there is an *interaction effect* between <u>all three</u> IVs in their effect on (prediction of) Z? (We are *not* interested in interaction effects involving only one or two IVs and the DV; we want to know only if there is a "tetradic" effect involving 3 IVs and the DV.) (2 pts)

(d) $\Delta df$(ABC:ABD:ACD:BCD $\rightarrow$ AB:AC:AD:BC:BD:CD) = df(ABC:ABD:ACD:BCD) – df(AB:AC:AD:BC:BD:CD). Let A, B, C, & D have cardinalities of 2, 3, 4, & 5, respectively. Use the log-linear method and compute this $\Delta df$. (2 pts)

$\Delta df$ =

**4. [14 points]** Consider on the left an *observed* contingency data table (p) for a *directed* system, with sample size N and a+b+...+h=1. None of parameters (a...h) is 0. Let the *calculated* table (q) for model AB:BZ be the table on the right.

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

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

(a) Fitting the AB:BZ model involves optimizing an expression subject to a set of constraints. Below in (a1) write the expression that is maximized or minimized (say whether it is maximized or minimized) and in (a2) & (a3) write the set of constraint equations *in terms of $q_1$ to $q_8$ and observed data a to h*. (The general constraint $\sum q_j = 1$ that is not specific to any model will be taken for granted and should not be specified.) Be sure that your constraint equations are linearly independent, i.e., not redundant.

(a1) minimize/maximize (*circle* one of these words; and write the expression that is optimized in terms of $q_1$ to $q_8$ and/or a through h) (2 pts)

(a2) subject to____ (state *how many* linearly independent) AB constraints; also give the constraint equation(s) below (2 pts):

(a3) And ____ BZ constraints (state how many, if any, <u>additional</u> BZ constraints there are, beyond the AB constraint(s)); give the constraint equation(s) below. (2 pts)

(a4) Write the matrix M, for which the full set of constraints can be summarized as equation $\mathbf{M\,q = M\,p}$, where **p** is the column vector with components a…h and **q** is the column vector with components $q_1 \ldots q_8$. (2 pts)

(b1) Solve for $q_4$ algebraically in terms of a…h. (2 pts)

(b2) Calculate $q_4$ not algebraically but using IPF. Obtain the value of $q_4$ in terms of numerical constants and parameters a through h in three steps: at the initialization of IPF (b2.1), after AB is imposed (b2.2), and then after BZ is imposed (b2.3). At each step indicate clearly the correction factor that multiplies the previous estimate of $q_4$ (4 pts)

(b2.1) $q_4^{initial} =$

(b2.2) After AB is imposed, $q_4^{AB} = q_4^{initial} \bullet$ _____

=

(b2.3) After BZ is next imposed, $q_4^{AB:BZ} = q_4^{AB} \bullet$ _____

=

**5. [4 points (or 6 points if you get extra credit)]**
(a) Suppose one has an AB probability table, where |A|=|B|=2, and that in this 2x2 table there is one *systematic zero*, say $(A_1, B_2)$, whose probability *must be* zero (like the probability of pregnant males). True or false? (circle; 2 pts): A:B:$A_1B_2$, i.e., the independence model with the additional systematic zero constraint, necessarily has a T=0.

(b) Verbally state the algorithm for finding loops in a variable-based structure. (2 pts)

(c) EXTRA CREDIT SPECULATIVE RESEARCH QUESTION (2 pts)
Propose an analog of this loop-detecting algorithm for state-based models like A:B:$A_1B_2$, and show whether this proposed algorithm find loops in this model or not?

[Hint#1: replace variable-based components by all of their states. You could also see if IPF needs more than one iteration.]

**6. [14 points]**

(a) The following are some calculations for a neutral system with *reference* = *top* (data). Recall that $L^2$(model) = 1.3863 N T(model); $\alpha$ here is the probability(Type I error):

| model | $L^2$ | df | $\alpha$ |
|---|---|---|---|
| ABC | 0.00 | 7 | 1.000 |
| AB:AC:BC | 0.76 | 6 | 0.382 |
| AB:AC | 10.58 | 5 | 0.005 |
| AB:BC | 51.71 | 5 | 0.000 |
| AC:BC | 1.31 | 5 | 0.518 |
| AB:C | 61.00 | 4 | 0.000 |
| AC:B | 10.61 | 4 | 0.031 |
| BC:A | 51.74 | 4 | 0.000 |
| A:B:C | 61.03 | 3 | 0.000 |

The best model is (circle one; 2 pts):

i. AB:AC:BC because it has the smallest $L^2$ aside from the data itself
ii. AB:AC because it is the most complex model with $\alpha < .05$
iii. AB:BC because it is the most complex model with the lowest recorded $\alpha$
iv. AC:BC because it is the simplest model with a high $\alpha$
v. A:B:C because it has the biggest $L^2$

(b) For data on directed system ABZ, where variables are binary and the *bottom is the reference*, here are results of a variable-based analysis.

| | $\Delta L^2$ | $\Delta$df | $\alpha$ |
|---|---|---|---|
| ABZ | 4.99 | | .17 |
| AB:AZ:BZ | 2.84 | | .24 |
| AB:BZ | 2.43 | | .12 |
| AB:AZ | 0.30 | | .58 |
| AB:Z | 0.00 | | 1.0 |

Here are results for state-based analysis for one additional model, where component $A_1B_1Z$ means that $q(A_1, B_1, Z_1) = p(A_1, B_1, Z_1)$ and $q(A_1, B_1, Z_2) = p(A_1, B_1, Z_2)$.

| | | | |
|---|---|---|---|
| AB:Z:$A_1B_1Z$ | 4.84 | | .03 |

(b1) Fill in the values of all six $\Delta$df values above (3 pts).

(b2) The best model is (circle one; 2 pts):

i. ABZ ii. AB:AZ:BZ iii. AB:BZ
iv. AB:AZ v. AB:Z vi. AB:Z:$A_1B_1Z$

(b3) This model is a good one because (circle all choices that are desirable properties of the model that you selected in (b2); 2 pts)

i. $\Delta L^2$ is low ii. $\Delta L^2$ is high iii. $\Delta$df is low iv. $\Delta$df is high
v. $\alpha$ is low vi. $\alpha$ is high

(c) True or false (circle one; 2 pts): If I am given a data frequency table, and know one of either T(model) or H(model) or $L^2$(model) (where $L^2$ is measured from the top), I can obtain the other two measures as simple linear expressions of the measure that I know.

(d) Suppose I go <u>down</u> the lattice of structures. After each measure ((i) to (iii)), write whether the measure is MD (monotonically decreasing or staying the same), MI, (monotonically increasing or staying the same), or NM (not monotonic, i.e., could either increase or decrease) as I go down the lattice (3 pts):

(i) transmission, T
(ii) alpha (for null hypothesis that the model fits the data)
(iii) degrees of freedom (# of parameters; i.e., # of constraint equations)

Discrete Multivariate Modeling (SySc 551/651) Professor M. Zwick
MIDTERM EXAM Winter 2009 (Feb. 10, 2009)

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

**1. [12 points]**
The following is observed data for a neutral system, ABC.

| | C: | 0 | | 1 | |
|---|---|---|---|---|---|
| | B: | 0 | 1 | 0 | 1 |
| A: | 0 | a | b | c | d |
| | 1 | e | f | g | h |

(a) Write an expression for T(AB:AC) in terms of parameters a…h, *using entropy expressions.* (2 pts)

(b) Write an expression for T(AB:AC) in terms of parameters a…h *using the p log(p/q) form of T*; give only *the first two terms* of this sum. (2 pts)

(c) Write an expression, in terms of parameters a…h, using the p log(p/q) form for the T needed to test the hypothesis that the data (i) is uniform in its (projected) A distribution and that (ii) B and C are independent of one another; give only *the first two terms* of this sum. (2 pts)

(d) The size of the exclusively triadic region of the 3-circles uncertainty diagram for the data (McGill and Quastler's "A function"; also known as the "interaction function") is given by (circle one; 1 pt)

i. -H(A) -H(B) -H(C) + H(A,B) + H(A,C) + H(B,C) - H(A,B,C)
ii. -H(A) -H(B) -H(C) + H(A,B) + H(A,C) + H(B,C) + H(A,B,C)
iii. -H(A) -H(B) -H(C) - H(A,B) - H(A,C) - H(B,C) - H(A,B,C)
iv. -H(A) -H(B) -H(C) - H(A,B) - H(A,C) - H(B,C) + H(A,B,C)

(e) Given *any* ABC data (circle one; 2 pts):

i. $T(A:B) \geq T_C(A:B)$ ii. $T(A:B) \leq T_C(A:B)$ iii. either i or ii could be true

(f) True or false (circle one; 1 pt): The A function = $T(A:B) - T_C(A:B)$.

(g) True or false (circle one; 2 pts) The A function = $I(ABC \rightarrow AB:AC:BC)$, where I is information distance.

**2. [34 points]**
On the left is *observed* data (p) for a directed system, with sample size N and probabilities a..h (all non-zero). On the right is the *calculated* table (q) for AB:AZ:BZ.

p(ABZ)

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

$q_{AB:AZ:BZ}(ABZ)$

| | Z: | 0 | | 1 | |
|---|---|---|---|---|---|
| | B: | 0 | 1 | 0 | 1 |
| A: | 0 | $q_1$ | $q_2$ | $q_3$ | $q_4$ |
| | 1 | $q_5$ | $q_6$ | $q_7$ | $q_8$ |

(a) Give a numerical answer (2 pts): df(AB:AZ:BZ) =

(b) Fill in the tables for AB, AZ, and BZ that are projected from p and from q (2 pts).

Projected from p

| | B: | 0 | 1 |
|---|---|---|---|
| A: | 0 | | |
| | 1 | | |

Projected from q

| | B: | 0 | 1 |
|---|---|---|---|
| A: | 0 | | |
| | 1 | | |

| | Z: | 0 | 1 |
|---|---|---|---|
| A: | 0 | | |
| | 1 | | |

| | Z: | 0 | 1 |
|---|---|---|---|
| A: | 0 | | |
| | 1 | | |

| | Z: | 0 | 1 |
|---|---|---|---|
| B: | 0 | | |
| | 1 | | |

| | Z: | 0 | 1 |
|---|---|---|---|
| B: | 0 | | |
| | 1 | | |

(c) In the (left side) tables projected from p, circle the cells -- the *smallest number* of them -- that provide sufficient & necessary parameters for the AB:AZ:BZ model. Do not circle any cell that is redundant, i.e., that can be evaluated from other circled cells. (*Hint: circle as many non-redundant cells as possible first for the AB table, then AZ, then BZ.* (3 pts)

(d) In terms of parameters a…h and variables $q_1 \ldots q_8$, write the (smallest) set of constraint equations that define model AB:AZ:BZ (3 pts).

(e) Write these constraints in the form M **q** = M **p**, where p and q are column vectors, by filling in, under 'Matrix, M' *only rows sufficient and necessary* to define this matrix for AB:AZ:BZ. Fill in values *only where the value is 1*; leave the matrix element blank when its value is 0. The same matrix is 'M' is understood to be on the right. (3 pts)

Matrix, M

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | | | | | | | |
| | | | | | | | |
| | | | | | | | |
| | | | | | | | |
| | | | | | | | |
| | | | | | | | |
| | | | | | | | |
| | | | | | | | |

| | | | |
|---|---|---|---|
| $q_1$ | | | a |
| $q_2$ | | | b |
| $q_3$ | | | c |
| $q_4$ | = | M | d |
| $q_5$ | | | e |
| $q_6$ | | | f |
| $q_7$ | | | g |
| $q_8$ | | | h |

(f) In fitting AB:AZ:BZ, what quantity – express this explicitly in terms of q's and/or parameters a…h -- is maximized subject to these constraints? (2 pts)

(g) The solution for the q's can be obtained by (circle all that are true; 2 pts)

i. $q_{AB:AZ:BZ}(ABZ) = p(AB)\, p(AZ)\, p(BZ)$
ii. $q_{AB:AZ:BZ}(ABZ) = p(AB)\, p(AZ)\, p(BZ) / [\, p(A)\, p(B)\, p(Z)\, ]$
iii. iteration (IPF) that imposes each projection once
iv. iteration (IPF) that cycles through projections more than once

(h) Use IPF to obtain $q_5$ by writing the sequence of values of $q_5$ in terms of the parameters, a through h, and values of $q_5$ earlier in the sequence.

(h1) The initial value is (1 pt):

$q_5^{initial} =$

(h2) After imposing AB, we get (1 pt)

$q_5^{AB} = q_5^{initial} *$ _____

(h3) After imposing also AZ, we get (3 pts) [Hint: one needs to know another $q_j^{AB}$ in addition to $q_5^{AB}$.]

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

(i) The q distribution can be used to predict Z, knowing A and B, from the conditional distribution of Z, given AB, specified by (circle all that are true; 2 pts)

i. $p(ABZ)/q_{AB:AZ:BZ}(AB)$
ii. $q_{AB:AZ:BZ}(ABZ) / p(AB)$
iii. $q_{AB:AZ:BZ}(ABZ) / q_{AB:AZ:BZ}(AB)$
iv. none of the above

(j) True or false (circle one; 2 pts):
$H(AB:AZ:BZ) = H(AB) + H(AZ) + H(BZ) - H(A) - H(B) - H(Z)$.

(k) To evaluate the model AB:AZ:BZ relative to the independence model as the reference, one needs to know an information distance and a $\Delta df$.

(k1) The appropriate information distance is (circle one; 2 pts)

i. T(AB:Z) – T(AB:AZ:BZ)
ii. T(AB:AZ:BZ)- T(AB:Z)
iii. T(ABZ) – T(AB:AZ:BZ)
iv. T(AB:AZ:BZ) – T(ABZ)

(k2) Fill in a numerical value (2 pts): The appropriate $\Delta df =$

(l) The information *lost* in model AB:AZ:BZ is (circle one; 2 pts)

i. T(AB:AZ:BZ)
ii. T(AB:Z) – T(AB:AZ:BZ)
iii. T(ABZ) – T(AB:AZ:BZ)
iv. T(AB:AZ:BZ)- T(AB:Z)
v. none of the above

(m) The information *captured* in model AB:AZ:BZ is (circle one; 2 pts)

i. T(AB:AZ:BZ)
ii. T(AB:Z) – T(AB:AZ:BZ)
iii. T(ABZ) – T(AB:AZ:BZ)
iv. T(AB:AZ:BZ)- T(AB:Z)
v. none of the above

**3. [14 points]**
(a) Using the convention that directed systems have an IV component which includes all the IVs, a directed system that has two or more ("predicting") components involving the same DV necessarily has a loop. True or false? (circle; 2 pts)

(b) Consider a directed system in which A, B, and C are IVs, and Z is a DV. How many *general* structures are there (include the data)? Exemplify each general structure with one specific structure. Circle all structures which do *not* have loops. (4 pts)

(c) Let $\Delta df(ABCD \rightarrow ABC:ABD:ACD:BCD) = df(ABCD) - df(ABC:ABD:ACD:BCD)$. Assume A, B, C, & D have cardinalities of 2, 3, 4, & 5, respectively. Use the log-linear method of computing degrees of freedom to compute this $\Delta df$. (2 pts)

$\Delta df =$ (give a numerical answer)

(d) (1 pt) For the data table shown in question 2, df(ABZ) =

(e) A distribution p(ABZ) can be written as $p(AB)*p(Z|AB)$. We might use for this the notation $AB\ {}_{AB}Z$ where the subscript means 'conditioned on.' p(AB) and p(Z|AB) can be independently specified, so df(ABZ) = df(AB) + df(Z|AB). Given that |A|=|B|=|Z|=2,

(e1) how many parameters does it take to specify AB? ___ (1 pt)

(e2) how many parameters does it take to specify p(Z|AB)? ___ Explain, i.e., justify your answer.(2 pts)

(f) In 1(a), you gave the degrees of freedom for model AB:AZ:BZ. Now consider a different model that also gets a calculated ABZ from specified AB, AZ, and BZ distributions, but differs in one respect. The AB distribution used in this new model is *not* the projected AB distribution from ABZ, but is instead the distribution for A:B. One might use the following notation for this model: $AB^{A:B}$:AZ:BZ, where the first component means the AB distribution for A:B, not the AB obtained directly from the ABZ data. What do you think is the value of $df(AB^{A:B}:AZ:BZ)$, *and why*? (2 pts)

$df(AB^{A:B}:AZ:BZ) =$

Discrete Multivariate Modeling (SySc 551/651) Professor M. Zwick
MIDTERM EXAM Winter 2010 (Feb. 9, 2010)

60 points; 2 hours; closed book, closed notes. *Answer question on exam sheets, where possible, and put **your name** on them.* Use $\Gamma(a, b,...) = -a \log a - b \log b - ...$

**1. [16 points]**
Let the data be the contingency 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, *expressing T in terms of entropies* (2 pts).

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

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

(d) Two variables are 'directly linked' if and only if they are involved in the same relation. Suppose $T_B$(A:Z) = 0. This means (circle one; 1 pt)

i. A is not directly linked to Z
ii. B is not directly linked to Z
iii. A is not directly linked to B
iv. all of the above
v. none of the above

(e1) The size of the triadic intersection region of the three uncertainty circles for ABZ data, namely McGill and Quastler's "A function," also known as the "interaction function, Q(A, B, Z), is given by (circle one; 1 pt)

i. $-H(A)-H(B)-H(Z)-H(A,B)-H(A,Z)-H(B,Z)-H(A,B,Z)$
ii. $-H(A)-H(B)-H(Z)-H(A,B)-H(A,Z)-H(B,Z)+H(A,B,Z)$
iii. $-H(A)-H(B)-H(Z)+H(A,B)+H(A,Z)+H(B,Z)-H(A,B,Z)$
iv. $-H(A)-H(B)-H(Z)+H(A,B)+H(A,Z)+H(B,Z)+H(A,B,Z)$

(e2) True or false? (circle one; 2 pts): $T(A:Z) - T_B(A:Z) = -Q(A, B, Z)$

(f) Circle whichever statement is true (2 pts):

i. $I(ABZ \rightarrow AB:AZ:BZ) = -Q(A, B, Z)$
ii. $I(ABZ \rightarrow AB:AZ:BZ) = +Q(A, B, Z)$
iii. neither equation i. nor equation ii. is true

(g) True or false? (circle one; 2 pts): If an ABZ relation exhibits non-zero constraint, then there exists at least one dyadic projection that exhibits non-zero constraint.

(h) (You may have to derive this.) (circle one; 2 pts) $T(AB:Z) - T(AB:AZ) =$

i. $H(Z) - H(Z|A)$ ii. $H(Z|A) - H(Z)$
iii. $H(Z) - H(Z|B)$ iv. $H(Z|B) - H(Z)$
v. $H(Z) - H(Z|AB)$ vi. $H(Z|AB) - H(Z)$

**2. [28 points]**
On the left is *observed* data (p) for a directed system, with sample size N and probabilities a..h (all non-zero). On the right is the *calculated* table (q) for AB:AZ:BZ.

p(ABZ)

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

$q_{AB:AZ:BZ}(ABZ)$

| | Z: | 0 | | 1 | |
|---|---|---|---|---|---|
| | B: | 0 | 1 | 0 | 1 |
| A: | 0 | $q_1$ | $q_2$ | $q_3$ | $q_4$ |
| | 1 | $q_5$ | $q_6$ | $q_7$ | $q_8$ |

(a) Show how degrees of freedom can be calculated in two different ways (and obtain the numerical value for df): via the Krippendorff method and the log-linear method (4 pts):

$df(AB:AZ:BZ)_{Krippendorff} =$

$df(AB:AZ:BZ)_{log\text{-}linear}$ =

(b) Fill in the tables for AB, AZ, and BZ that are projected from p and from q (2 pts).

Projected from p

| | B: | 0 | 1 |
|---|---|---|---|
| A: | 0 | | |
| | 1 | | |

| | Z: | 0 | 1 |
|---|---|---|---|
| A: | 0 | | |
| | 1 | | |

| | Z: | 0 | 1 |
|---|---|---|---|
| B: | 0 | | |
| | 1 | | |

Projected from q

| | B: | 0 | 1 |
|---|---|---|---|
| A: | 0 | | |
| | 1 | | |

| | Z: | 0 | 1 |
|---|---|---|---|
| A: | 0 | | |
| | 1 | | |

| | Z: | 0 | 1 |
|---|---|---|---|
| B: | 0 | | |
| | 1 | | |

(c) In the (left side) tables projected from p, circle the *smallest number* of cells that provide sufficient & necessary parameters for the AB:AZ:BZ model. Do this in the following sequence: first circle all non-redundant cells for the AB table; then circle as many non-redundant cells as there are in AZ, given that one already knows AB; then circle as many non-redundant cells as there are in BZ, given that one knows both AB and AZ. Then, in terms of $q_1 \ldots q_8$ and parameters a…h, write the (smallest) set of constraint equations that define model AB:AZ:BZ. (Don't include the structure-independent constraint that all probabilities must sum to 1.) Label the constraint equations by indicating which come from AB, which from AZ, and which from BZ. (4 pts)

(d) Constraint equations specify what a model says we know, or to put it differently, what parameters are needed to define a model. Suppose one had a constraint equation of the form $q_i + q_j + q_k = \alpha + \beta + \gamma.$, where these terms on the right side of the equation are constants. (This is made up; there need not be any equation of this form in the right answer to the previous question.) This equation says that we know, i.e., make use of in the composition step (circle one; 2 pts)

i. the individual values of $\alpha$ and $\beta$ and $\gamma$
ii. only the sum of these, not the individual values
iii. neither of these

(e) Write the constraints in your answer to (c) in the form M **q** = M **p**, where p and q are column vectors, by filling in, under 'Matrix, M' *only rows sufficient and necessary* to define this matrix for AB:AZ:BZ. Fill in values *only where the value is 1*; leave the matrix element blank when its value is 0. Here, include the structure-independent constraint that probabilities must sum to 1. The same 'M' is on the right. (2 pts)

Matrix, M

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | | | | | | | |
| | | | | | | | |
| | | | | | | | |
| | | | | | | | |
| | | | | | | | |
| | | | | | | | |
| | | | | | | | |
| | | | | | | | |

| | | | |
|---|---|---|---|
| $q_1$ | | | a |
| $q_2$ | | | b |
| $q_3$ | | | c |
| $q_4$ | = | M | d |
| $q_5$ | | | e |
| $q_6$ | | | f |
| $q_7$ | | | g |
| $q_8$ | | | h |

(f) In fitting AB:AZ:BZ, what quantity – express this explicitly in terms of q's and/or parameters a…h – is maximized subject to these constraints? (2 pts)

(g) The solution for the q's can be obtained by (circle one; 2 pts)

i. $q_{AB:AZ:BZ}(ABZ) = p(AB)\ p(AZ)\ p(BZ)$
ii. $q_{AB:AZ:BZ}(ABZ) = p(AB)\ p(AZ)\ p(BZ) / [\ p(A)\ p(B)\ p(Z)\ ]$
iii. iteration only (there is no algebraic solution for $q_{AB:AZ:BZ}$)
iv. none of the above: the solution for $q_{AB:AZ:BZ}$ cannot be obtained at all

(h) Is the following equation true or false? (circle one; 2 pts):

$$H(AB:AZ:BZ) = H(AB) + H(AZ) + H(BZ) - H(A) - H(B) - H(Z).$$

(i) To evaluate the model AB:AZ:BZ relative to the *independence model* as the reference, one needs to know an information distance and a Δdf.

(i1) The appropriate information distance equals (circle one; 2 pts)

i. T(AB:Z) – T(AB:AZ:BZ)
ii. T(AB:AZ:BZ)- T(AB:Z)
iii. T(ABZ) – T(AB:AZ:BZ)
iv. T(AB:AZ:BZ) – T(ABZ)

(i2) Fill in a *numerical* value (2 pts): The appropriate Δdf =

(j) The information *lost* in model AB:AZ:BZ is (circle one; 2 pts)

i. T(AB:AZ:BZ)
ii. T(AB:Z) – T(AB:AZ:BZ)
iii. T(ABZ) – T(AB:AZ:BZ)
iv. T(AB:AZ:BZ)- T(AB:Z)

v. none of the above

(k) The information *captured* in model AB:AZ:BZ is (circle one; 2 pts)

i. T(AB:AZ:BZ) ii. T(AB:Z) – T(AB:AZ:BZ)
iii. T(ABZ) – T(AB:AZ:BZ) iv. T(AB:AZ:BZ)- T(AB:Z)
v. none of the above

**3. [8 points]**

(a1) True or false? (circle one; 1 pt) If a structure has a loop, then a child of this structure will necessarily have a loop.

(a2) True or false? (circle one; 1 pt) If a structure has a loop, then a parent of this structure will necessarily have a loop.

(b) Consider two models, $m_1$ and $m_2$. Let all the different relations involved in model $m_1$ and $m_2$ be written as $r_1$ and $r_2$, respectively. (For example, if $m_1$ = ABC:CD, $r_1$ includes ABC, AB, AC, BC, A, B, C, CD, and D.) Let $m_3$ be the nearest common *ancestor* (parent or grandparent or …) and let $r_3$ be the relations in $m_3$, and let $m_4$ be the nearest common *descendant* (child or grandchild or …) and let $r_4$ be the relations in $m_4$. $r_1 \cap r_2$ means what the two models have in common (their "intersect," i.e., logical "and"). $r_1 \cup r_2$ includes what is present in either model (their "union," i.e., logical "or"). Circle all equations that are true (2 pts)

i. $r_3 = r_1 \cap r_2$ ii. $r_3 = r_1 \cup r_2$ iii. $r_4 = r_1 \cap r_2$ iv. $r_4 = r_1 \cup r_2$

(c) Calculating $\Delta df(m_i \rightarrow m_j)$ without actually calculating either $df(m_i)$ or $df(m_j)$ individually (circle one; 1 pt)

i. can be done by the Krippendorff method of df calculation
ii. can be done by the log-linear method of df calculation
iii. can be done by both methods
iv. cannot be done by either method

(d) Klir's equivalence classes of structures, called ρ structures, each represented by an ordinary graph which indicates which variables are directly linked to which other variables, where each equivalence class has a most complex C-structure and a least complex P structure, is (circle one; 1 pt)

i. a way to do a bottom up search ii. a way to do a top down search
iii. a way to do disjoint searches iv. a way to search the lattice hierarchically

(e) (2 pts) State an analog of the loop-detecting algorithm *for state-based models* like $A:B:A_1B_2$, and show whether this proposed algorithm find loops in this model or not? [Hint: replace variable-based components by all of their states.]

**4. [8 points]**
Given a data distribution as follows

| | $y_0$ | $y_1$ |
|---|---|---|
| $x_0$ | .1 | .2 |
| $x_1$ | .3 | .4 |

(a1) I want to consider the state-based model that specifies that $p(x_1,y_1) = .4$ . Write an expression for transmission, T, the error in this model, in terms of the Γ function and numerical constants. (2 pts)

T =

(a2) What is the value of Δdf = df(data) – df(for this state-based model)? (1 pt)

Δdf =

(b) Define two subsystems for the above table: (1) the diagonal states $(x_0, y_0)$ and $(x_1,y_1)$ and (2) the off-diagonal states, $(x_1,y_0)$ and $(x_0,y_1)$. Call these the D (diagonal) and O (off-diagonal) subsystems. Using the Γ function and the numerical constants in the table, write an expression for the total uncertainty of the table as a sum of between-subsystem uncertainty plus (average) within-subsystem uncertainty. (2 pts)

(c) Suppose I want to test the hypothesis that the table is "really" uniform, i.e., that the deviations of its probability values from .25 are just due to sampling error.

(c1) The T that should be used to test this hypothesis is (write it in terms of the Γ function and the numerical constants in the table) (2 pts)

T =

(c2) The model corresponding to this hypothesis has df = (circle one; 1 pt)

0 1 2 3 4 5 6 7 8

Discrete Multivariate Modeling (SySc 551/651) Professor M. Zwick
FINAL EXAM Winter 2008 (3/18/08)

60 points; 2 hours; closed book. *Answer question on exam sheets, and put **your name** on them.* $L^2 = LR$. You can use $\Gamma(a, b,...) = -a \log a - b \log b - ...$ Do not use red ink.

---

1. [8 points] Consider the following *observed* contingency data table (the p's) for a *directed* system, with sample size N and a+b+...+h=1. None of parameters (a...h) is 0.

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

The calculated table (the q's) for model AB:BZ is

| | Z: | 0 | | 1 | |
|---|---|---|---|---|---|
| | B: | 0 | 1 | 0 | 1 |
| A: | 0 | i | j | k | l |
| | 1 | m | n | o | r |

(a) Write an equation in terms of (a...h) for $q_{AB:BZ}(1,0,1)$. (1 pt)

$q_{AB:BZ}(1,0,1) = o =$

(b) Write an equation in terms of (a...r) for T(AB:BZ). (1 pt)

T(AB:BZ) =

(c1) T(AB:Z) - T(AB:BZ) is the amount of information (1 pt; circle one):

i. captured in model AB:Z ii. captured in model AB:BZ
ii. lost in model AB:Z iv. lost in model AB:BZ

(c2) T(AB:Z)-T(AB:BZ) equals the uncertainty difference (circle one; 1 pt),

i. u(Z) - u(Z|A) iii. u(Z) – u(Z|B) v. none of these
ii. u(Z|A) – u(Z) iv. u(Z|B) – u(Z)

(d) IPF maximizes $H = -\sum q \log q$ subject to the model constraints. For model AB:BZ, using parameters i…r, write the expression for H which is maximized and a complete but minimal set of constraint equations. (Don't include any redundant constraints.) (4 pts)

2. [20 points]
(a1) Given the following RA analysis for a neutral system, with reference = top,

| | $L^2$ | Δdf | α | Info |
|---|---|---|---|---|
| ABCD | 0.00 | 0 | 1.00 | 1.00 |
| ABC:ABD:ACD:BCD | 0.00 | 1 | 1.00 | 1.00 |
| ABC:ABD:BCD | 0.00 | 2 | 1.00 | 1.00 |
| ADB:BCD:AC | 1.33 | 3 | 0.73 | 1.00 |
| BCD:AB:AC:AD | 2.45 | 4 | 0.66 | 1.00 |
| AB:AC:AD:BC:BD:CD | 29.09 | 5 | 0.00 | 0.99 |
| AB:AC:BC:BD:CD | 74.27 | 6 | 0.00 | 0.97 |
| AC:AD:BC:BD | 193.25 | 7 | 0.00 | 0.91 |
| AC:BC:BD | 404.83 | 8 | 0.00 | 0.83 |
| BC:BD:A | 973.57 | 9 | 0.00 | 0.59 |
| BD:A:C | 1574.54 | 10 | 0.00 | 0.33 |
| A:B:C:D | 2366.00 | 11 | 0.00 | 0.00 |

one consideration is to find a model that is as simple as possible but still retains some significant fraction of the information in the data. A second consideration is that the difference between the model and the data not be statistically significant. For the present case, looking at the above output, (circle one; 2 pts)

i. these two considerations agree, pointing to the same good model
ii. these two considerations disagree, pointing to different good models

(a2) If you answered i. for (a1), what is the good model that these considerations agree on. If you answered ii., which different models do the two considerations point to? (2 pts)

(b) The Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC) are measures of the goodness of a model based on its error and complexity. These measures are linear combinations of these two criteria, with form error + k * complexity, where k is different for AIC and BIC. Alpha (α), the p(Type I error), also depends on both error and complexity.

(b1) (Let HDM = hierarchically descendant models, i.e., models, one of which is a child, grandchild, etc. of another.) Circle all true choices in the following list (2 pts):

i. α can be used to compare only HDM
ii. AIC and BIC can be used to compare only HDM
iii. α is also linear in error & complexity, with just a different k than AIC or BIC
iv. for ref=bottom, AIC accepts higher ordinality relations than BIC

(b2) 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 = $\alpha$ iv. AIC = $1 - \alpha$

(b3) In this absolute AIC expression, the error term in the error-complexity tradeoff is not obvious, but one can add to AIC any constant depending only on the data but not on the models under evaluation. What can you add to AIC that makes the error term in it plainly visible? What does AIC then become? (2 pts)

(b4) Occam outputs $\Delta$AIC = AIC(reference) – AIC(model). For the table in (a1), give numerical answers for the $\Delta$AIC values below. Circle the model that $\Delta$AIC favors. (2 pts)

$\Delta$AIC(ADB:BCD:AC) = _____

$\Delta$AIC(BCD:AB:AC:AD) = _____

(c) In reconstruction, a null hypothesis typically asserts that a model (circle one; 2 pts)

i. is the same as a reference model ii. is different from a reference model

(d1) True or false? (circle one; 1 pt) When the reference model is the top (the saturated model), one should choose models for which $\alpha$ is small, conventionally less than 0.05.

(d2) True or false? (circle one; 1 pt) When the reference = bottom (the independence model), one should choose models for which $\alpha$ is small, conventionally less than 0.05.

(e1) In a top-down search, when the reference model is the top, if one has made a Type I error, the consequences have probably been that one has ended up (circle one; 1 pt)

i. a model unnecessarily complex ii. a model too simple

(e2) In a top-down search, when the reference model is the top, if one has made a Type II error, the consequences have probably been that one has ended up with (circle one; 1 pt)

i. a model unnecessarily complex ii. a model too simple

(e3) In a bottom-up search, when the reference is the bottom, if one has made a Type I error, the consequences have probably been that one has ended up (circle one; 1 pt)

i. choosing a model whose predictive interactions are statistically unjustified
ii. abandoning a model whose predictive interactions are statistically justified

(e4) In a bottom-up search, when the reference is the bottom, if one has made a Type II error, the consequences have probably been that one has ended up (circle one; 1 pt)

i. choosing a model whose predictive interactions are statistically unjustified
ii. abandoning a model whose predictive interactions are statistically justified

3. [18 points]
(a) Consider system ABZ, with |A| = |B| = |Z| = 2.

| | Z: | 0 | | 1 | |
|---|---|---|---|---|---|
| | B: | 0 | 1 | 0 | 1 |
| A: | 0 | | | | |
| | 1 | | | | |

(a1) df(ABZ) = _____ (1 pt)

(a2) If $q = p_1 * p_2$, then $df(q) = df(p_1) + df(p_2)$; e.g., df(AB:CD) = df(AB)+df(CD). For ABZ, we can write *p(ABZ) = p(Z|AB) p(AB)*. Therefore, df(ABZ) = df(Z|AB) + df(AB). What is df(Z|AB), i.e., if table entries are *conditional* probabilities $p(Z_k | A_i, B_j)$, and not *joint* probabilities $p(A_i, B_j, Z_k)$, how many numbers does it take to specify such a table of the conditional distribution p(Z|AB)? Hint: Consider what the df of AB is.

df(Z|AB) = _____ (1 pt)

(a3) Consider the Bayesian Network model $q_{BN}(ABZ) = p(A)p(B)\ p(Z|AB)$. p(AB) in the italicized equation in (a2) is here replaced by p(A)p(B), i.e., by q(A:B). $q_{BN}(ABZ)$ is the same as q for which RA model in the Lattice of Specific Structures (circle one; 2 pt)

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 the above

(a4) df( $q_{BN}(ABZ)$ ) = _____ (1 pt)

(a5) The df for (RA) model AB:AZ:BZ is: df(AB:AZ:BZ) = _____ (1 pt)

(b) I want to predict Z, using RA models ABZ and AB:AZ:BZ and above BN model.

(b1) True or false? (circle one; 2 pts): With knowledge of A and B, AB:AZ:BZ will reduce the uncertainty of Z at least as much as and possibly more then ABZ.

(b2) The uncertainty reduction in Z can be derived *algebraically* from the data (i.e., without iterative calculations) for (circle one; 2 pts)

i. ABZ but not for AB:AZ:BZ iii. both models
ii. AB:AZ:BZ but not for ABZ iv. neither model

(b3) With a reference model of AB:Z, if I want to use AB:AZ:BZ to predict Z, I want an alpha (probability of error if I reject the null hypothesis) to be (circle one; 2 pts)

i. small ii. large

(b4) True or false? (circle one; 2 pts): Given A and B, the above BN model will generate the same predictions of Z as RA model ABZ.

(c1) The 'partialled relatedness' ("A") function defined by McGill & Quastler is here

$T_B(A:Z) - T(A:Z) = T_A(B:Z) - T(B:Z) = T_Z(A:B) - T(A:B)$
$= -H(A) - H(B) - H(Z) + H(AB) + H(AZ) + H(BZ) - H(ABZ).$

The value of this function is (circle; 2 pts): i. $\geq 0$ ii. $\leq 0$ iii. $\geq 0$ *or* $\leq 0$

(c2) (circle; 2 pts) $H(ABZ) - H(AB:AZ:BZ) =$: i. $\geq 0$ ii. $\leq 0$ iii. $\geq 0$ *or* $\leq 0$

4. [14 points]
(a) True or false? (circle; 2 pts): In state-based RA, only a state with a *higher* probability than expected can be a model parameter, not one with a *lower* probability than expected.

(b) True or false (circle one; 2 pts): When compressing a function with RA, i.e., treating the function value as if it were a frequency, BIC is a good measure to pick a model with.

(c) For set-theoretic RA, (circle all statements that are true, 3 pts)

i. models with loops don't require iteration
ii. complexity is also measured by df
iii. deterministic systems (mappings) cannot be decomposed
iv. stochastic directed systems (relations) cannot be decomposed
v. composition uses minimum, not maximum, entropy
vi. statistical significance is not applicable

(d) Is ABC = {001, 010, 100} decomposable without loss? ___Yes ___No. If 'No', the extra tuple(s) from the first decomposition is/are _____ If 'Yes,' the simplest no-loss decomposed structure is _____ Show your work. (3 pts)

(e) Consider the following AB, BC, and AC relations.

| | $B_0$ | $B_1$ |
|---|---|---|
| $A_0$ | 0 | .7 |
| $A_1$ | .3 | 0 |

| | $C_0$ | $C_1$ |
|---|---|---|
| $B_0$ | 0 | .3 |
| $B_1$ | .7 | 0 |

| | $C_0$ | $C_1$ |
|---|---|---|
| $A_0$ | .4 | .3 |
| $A_1$ | .3 | 0 |

(e1) Are these relations pair-wise consistent in their overlaps? Yes or no? (circle; 1 pt)

(e2) Treat AB, AC, BC *set-theoretically*, so that if a probability > 0, then the tuple is present; if it is 0, the tuple is absent. Consider the composition of these three relations. What result do you get? Discuss and explain your findings briefly. (3 pts)

Discrete Multivariate Modeling (SySc 551/651) Professor M. Zwick
FINAL EXAM Winter 2009 (3/17/09)

60 points; 2 hours; closed book. *Answer question on exam sheets, and put **your name** on them.* $L^2 = LR$. You can use $\Gamma(a, b,...) = -a \log a - b \log b - ...$ Do not use red ink.

---

1. [4 points]
(a) "Complexity" usually means df. Suppose one was considering, in a top-down search of neutral system ABCD, whether one should descend from ABC:ABD:CD to ABC:ABD or ABC:AD:BD:CD. Both of these latter two models are one-step descendant models of ABC:ABD:CD. All four variables have cardinality 2. What is the relationship between the df values for the two descendant models? (circle one; 2 pts)

i. df(ABC:ABD) < df(ABC:AD:BD:CD)
ii. df(ABC:ABD) = df(ABC:AD:BD:CD)
iii. df(ABC:ABD) > df(ABC:AD:BD:CD)

(b) There is a way to integrate error and complexity known as minimum description length (MDL), in which complexity is not simply the number of parameters used to specify the model, but also involves the functional form of the parameters. For example, $z = a x + b y$ and $z = x^a \sin(b y)$ might have different MDL complexities even though both models have two parameters, a and b. If one considered the two RA models ABC:ABD and ABC:AD:BD:CD, what, besides df, *might* be relevant to the complexity of these two models, and thus might be considered in quantifying complexity? (2 pts)

2. [28 points]
(a1) Transmissions and conditional transmissions are *measures* that one calculates from data, but can also be thought of as indicating the *goodness of models*. Let $T_A(B:Z) = 0$. This is a measure calculated from the data, but it also tells us which models have zero error. Given this specific equation, circle all models that have T(model) = 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 | |

(a2) Draw the Krippendorff structure (boxes & lines) for the *simplest* structure that you circled. (1 pt)

(b) Suppose one has an AB frequency table, where |A|=|B|=2, and in this 2x2 table there is one systematic zero, i.e., an $(A_i, B_j)$ state whose frequency *must* be zero (like a frequency of pregnant males). True or false? (circle; 2 pts): The independence model with the *additional* systematic zero constraint necessarily has T=0.

(c) An RA analysis of a 3-variable directed system gave the following results, where the reference model for $\alpha^*$ is the structure marked by * and the reference model for $\alpha^{\#}$ is the structure marked by #.

| Structure | ΔDF | ΔLR | $\alpha^*$ | $\alpha\#$ | %ΔdH(Z) |
|---|---|---|---|---|---|
| ABZ | 8 | 27.00 | 0.0007 | 0.0047 | 9.10 |
| AB:AZ:BZ# | 4 | 11.98 | | 1.0000 | 4.04 |
| AB:AZ | 2 | 8.22 | 0.0164 | - | 2.77 |
| AB:BZ | 2 | 4.20 | 0.1227 | - | 1.41 |
| AB:Z* | 0 | 0.00 | 1.0000 | - | 0.00 |

(c1) Using the provided Chi-square table, the range of values for $\alpha^*$ for model AB:AZ:BZ is from _____ to _____ (fill in *numeric* values; 2 pts).

(c2) What can be said about the relative power of A vs. B to predict Z (circle one; 1 pt)

i. A predicts better than B ii. B predicts better than A
iii. A and B predict equally well iv. neither A nor B predicts Z at all

(c3) Is there a statistically significant triadic interaction effect between A, B, and Z? (circle one; 2 pt)

i. no ii. yes iii. insufficient information to decide

(c4) Given the above table, the best model is (circle one; 2 pts)

i. ABZ ii. AB:AZ:BZ iii. AB:AZ iv. AB:BZ v. AB:Z

(d) In an analysis where the bottom model is the reference, suppose one uses two approaches to select models. In approach I, one uses either AIC or BIC. In approach II, one uses $\alpha$ relative to the reference model – call this $\alpha_{cumulative}$ – or all the $\alpha$s for every step from the reference to the model under consideration – call this (whole set of values) $\alpha_{incremental}$. The search direction is up, so 'going higher' means more complex models.

(d1) In approach I, (circle one; 2 pts)

i. using AIC lets one go higher than using BIC
ii. using BIC lets one go higher than using AIC
iii. these two criteria point to the same model in nearly all cases

(d2) In approach II, (circle one; 2 pts)

i. using $\alpha_{cumulative}$ lets one go higher than using $\alpha_{incremental}$
ii. using $\alpha_{incremental}$ lets one go higher than using $\alpha_{cumulative}$
iii. these two criteria point to the same model in nearly all cases

(d3) Suppose I want to select a model that has the highest %correct in the training data. This (third) approach is (circle one; 2 pts)

i. stupid because it will usually result in overfitting
ii. smart because it will usually produces high %correct(test)
iii. equivalent to using $\alpha_{cumulative}$
iv. equivalent to using AIC

(e) Given (part of) an analysis for a neutral system, with reference = top,

| | $L^2$ | Δdf | α | Info |
|---|---|---|---|---|
| ABCD | 0.00 | 0 | 1.00 | 1.00 |
| ADB:BCD:AC | 1.33 | 3 | 0.73 | 1.00 |
| BCD:AB:AC:AD | 2.45 | 4 | 0.66 | 1.00 |
| A:B:C:D | 2366.00 | 11 | 0.00 | 0.00 |

Occam outputs ΔAIC = AIC(reference) – AIC(model). For the above table, give *numerical* answers for the ΔAIC values below. Also, which of these two models does ΔAIC favor? (circle the favored model) (2 pts)

ΔAIC(ADB:BCD:AC) = _____

ΔAIC(BCD:AB:AC:AD) = _____

(f) Suppose the Occam fit output for model m = AB:AZ:BZ is

| 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 |

(f1) The model prediction rule (whether to predict Z=1 or Z=2) for (A,B)=(1,1) is obtained by comparing (circle one; 2 pts)

i. b & c ii. d & e iii. a & (b+c) iv. a & (d+e)
v. (b+c) to (d+e) vi. none of the above

(f2) In terms of the parameters in the above table (and necessary constants), write an expression for the transmission T that you would use to test the null hypothesis that there really isn't any basis for (A,B)=(1,1) to predict either Z=1 or Z=2, i.e., that the difference in model probabilities for the two states of Z isn't statistically significant). (2 pts)

(g) For modeling when the reference model is the top (data), assume that one cares more about fidelity to the data than one does about the simplicity of the model. For modeling

when the reference model is the bottom (independence), assume that one cares most about positing only interactions that are statistically significant.

(g1) When the reference is the top, (circle one; 2 pts)

i. Type I errors are worse than Type II errors
ii. Type II errors are worse than Type I errors
iii. Type I and Type II errors are equally bad
iv. Neither is bad; we actually want these errors to occur

(g2) When the reference is the bottom, (circle one; 2 pts)

i. Type I errors are worse than Type II errors
ii. Type II errors are worse than Type I errors
iii. Type I and Type II errors are equally bad
iv. Neither is bad; we actually want these errors to occur

3. [16 points]
(a) In set-theoretic RA, for XY:XZ:YZ, the calculated relation is (circle one; 3 pts)

i. $XY \cup XZ \cup YZ$
ii. $XY \cap XZ \cap YZ$
iii. $XY \otimes XZ \otimes YZ$
iv. $(XY \otimes Z) \cup (XZ \otimes Y) \cup (YZ \otimes X)$
v. $(XY \otimes Z) \cap (XZ \otimes Y) \cap (YZ \otimes X)$
vi. none of the above (no closed form)

(b) Both of the following relations (I & II) were offered in a paper by Cordell as examples of 'epistasis', where genotypes A and B jointly effect phenotype, C. In both relations, the genotypes have 3 states (common for diploid organisms that have two copies of each gene). The phenotype has 3 states in relation I, but 2 states in relation II. For each of these two relations, analyze these relations as if they were neutral systems and state below what the simplest structure(s) is (are) that fits the data with no error. It is not necessary to show your work, but if you like, you can attach it as separate worksheets.

| I | | |
|---|---|---|
| A | B | C |
| 1 | 1 | 1 |
| 1 | 2 | 2 |
| 1 | 3 | 2 |
| 2 | 1 | 3 |
| 2 | 2 | 2 |
| 2 | 3 | 2 |
| 3 | 1 | 3 |
| 3 | 2 | 2 |
| 3 | 3 | 2 |

| II | | |
|---|---|---|
| A | B | C |
| 1 | 1 | 1 |
| 1 | 2 | 1 |
| 1 | 3 | 1 |
| 2 | 1 | 1 |
| 2 | 2 | 2 |
| 2 | 3 | 2 |
| 3 | 1 | 1 |
| 3 | 2 | 2 |
| 3 | 3 | 2 |

(b1) The simplest structure(s) for relation I is _____ (2 pts)

(b2) The simplest structure(s) for relation II is _____ (2 pts)

(c) When analyzing a function, not a frequency distribution, using the k-systems technique, circle all outputs of Occam that are no longer meaningful (3 pts):

| | | |
|---|---|---|
| i. H | ii. ΔDF | iii. LR |
| iv. information | v. ΔAIC | vi. ΔBIC |

(d) For the following contingency table, 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

(e) Log-linear parametization of frequencies have the form, $F = \eta\, \tau^X\, \tau^Y\, \tau^{XY}$. In such representations, the absence of 'an effect,' e.g., deviation from uniformity of a margin or presence of an interaction of non-zero strength, shows up as (circle one; 2 pts)

i. all parameters being equal
ii. the parameter(s) for the relevant effect having value 0
iii. the parameter(s) for the relevant effect having value 1
iv. the parameter(s) for the relevant effect having value N (sample size)
v. none of the above

(f) True or false (circle one; 2 pts): In 'identification,' i.e., composition of separate datasets that are not projections of some unitary dataset, 'global inconsistency' of the separate datasets is impossible if all pairs of these datasets have the same margins for those variables and variable combinations that they have in common.

4. [12 points]
(a) Consider the contingency table (p distribution) below on the left, where a+b+c+d=1. In terms of known parameters a, b, c, d, fill in the values below on the right of the calculated (q) distribution for the one-parameter *state-based* model $X_1Y_2$. (2 pts)

p-distribution

| | $Y_1$ | $Y_2$ |
|---|---|---|
| $X_1$ | a | b |
| $X_2$ | c | d |

q-distribution

| | $Y_1$ | $Y_2$ |
|---|---|---|
| $X_1$ | | |
| $X_2$ | | |

(b) Write an expression for $T(X_1Y_2)$, the transmission of this model, in terms of parameters a, b, c, d. (2 pts)

$T(X_1Y_2) =$

(c) What $\Delta df$ would I use to test the hypothesis that q is the same as p, i.e., that the model fits the data perfectly? (Here the reference model is the top.) (1 pt)

$\Delta df =$

(d) Write an expression, in terms of parameters a, b, c, d and numerical constants (but you can use $T(X_1Y_2)$ from question (b) above which has already been expressed in terms of these parameters), for the information *captured* in the $X_1Y_2$ model relative to the *uniform* distribution (<u>not</u> relative to the *independence* model). (2 pts)

Information captured in $X_1Y_2 =$

(e) Rename the above calculated distribution $q_1$ (instead of just q). I now want to pick a second state to add to the state based model. Now the model will look like $A_1B_2{:}A_iB_j$, for some particular i and j. I have three possible choices for this second state, namely $X_1Y_1$, $X_2Y_1$, and $X_2Y_2$. Call the distribution I get after choosing the second state $q_2$. I want to pick that state among these three 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. $\Sigma\, q_2 \log [q_2/q_1]$ |

(f1) (New question, unrelated to above.) True or false (circle one; 2 pts): For $|A|=|B|=|Z|=3$, the q-distribution for the state-based model $AB{:}Z{:}A_1B_1Z_2$ cannot be calculated analytically but needs to be iteratively calculated, using IPF.

(f2) df for the model in (f1) = _____ (1 pt)

Discrete Multivariate Modeling (SySc 551/651) Professor M. Zwick
FINAL EXAM Winter 2010 (3/16/10)

60 points; 2 hours; closed book. *Answer question on exam sheets, and* ***put your name on page 1.*** $L^2$ = Likelihood Ratio Chi-square. Do not use red ink.

---

**1. [10 points]**

(a) 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

(b) A test is given to detect a disease. A 'negative' test result means that this condition is not detected, i.e., the patient is judged to be free of the disease; a 'positive' result means that the condition is detected, i.e., the patient is judged to have the disease. The actual condition of the patient and the test conclusions are summarized in these frequencies: TN = true negatives, FP = false positives, FN = false negatives, TP = true positives. The frequency marginals of the actual distribution are N = TN + FP, P = FN + TP. Assume that the null hypothesis is 'negative.'

| | Test | (-) | (+) | |
|---|---|---|---|---|
| Actual | (-) | TN | FP | N |
| | (+) | FN | TP | P |

(b1) Which of these is true (circle one; 2 pt)?

i. FP and FN are both Type I errors
ii. FP and FN are both Type II errors
iii. FP are Type I errors; FN are Type II errors
iv. FP are Type II errors; FN are Type I errors
v. none of the above

(b2) Recall that in the communication situation, where S means message sent and R means message received, H(R|S) is called 'noise' and H(S|R) is called 'equivocation.' In the present case, the Actual situation is the message sent, and the Test results are the message received. Which of the following is true (circle one; 2 pts)?

i. FP = 'noise' and FN = 'equivocation'
ii. FP = 'equivocation' and FN = 'noise'
iii. neither: FP and FN do not map 1:1 onto these conditional uncertainties

(c) Let |A|=|B|=|Z|=2.

(c1) What is df(AB:AZ:BZ)? (1 pt)

(c2) Assume that the AB projection of the ABZ data is nearly what I would expect if A and B were independent, and so I want a (non-standard-RA) model where the hypothesis of independence for A and B is included in the model. The model will have a relation written as $AB_{A:B}$ and I could have a model like $AB_{A:B}$:AZ:BZ, which will mean that q(AZ) = p(AZ) and q(BZ) = p(BZ), but q(AB) = p(A)*p(B), *instead of* q(AB) = p(AB).

(c2.1) df($AB_{A:B}$:AZ:BZ) = _____ (2 pts)

(c2.2) Name one standard RA model that has the same df:_____ (1 pt)

**2. [8 points]**

(a) On the left is *observed* data (p) for a directed system, with sample size N and probabilities a..h (all non-zero). On the right is the *calculated* table (q) for AB:AZ:BZ.

p(ABZ)

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

$q_{AB:AZ:BZ}$(ABZ)

| | Z: | 0 | | 1 | |
|---|---|---|---|---|---|
| | B: | 0 | 1 | 0 | 1 |
| A: | 0 | $q_1$ | $q_2$ | $q_3$ | $q_4$ |
| | 1 | $q_5$ | $q_6$ | $q_7$ | $q_8$ |

(a1) Fill in the tables for AB, AZ, and BZ that are projected from p and from q (2 pts).

Projected from p

| | B: | 0 | 1 |
|---|---|---|---|
| A: | 0 | | |
| | 1 | | |

Projected from q

| | B: | 0 | 1 |
|---|---|---|---|
| A: | 0 | | |
| | 1 | | |

| | Z: | 0 | 1 |
|---|---|---|---|
| A: | 0 | | |
| | 1 | | |

| | Z: | 0 | 1 |
|---|---|---|---|
| A: | 0 | | |
| | 1 | | |

| | Z: | 0 | 1 |
|---|---|---|---|
| B: | 0 | | |
| | 1 | | |

| | Z: | 0 | 1 |
|---|---|---|---|
| B: | 0 | | |
| | 1 | | |

Use IPF to obtain $q_7$ by writing the sequence of values of $q_7$ in terms of the parameters, a through h, and values of $q_7$ earlier in the sequence. The initial value is $q_7^{initial} = 1/8$.

(a2) After imposing AB, we get $q_7^{AB} = q_7^{initial} *$ _____(2 pts)

(a3) After imposing also AZ, we get (2 pts) [Hint: one needs to know another $q_j^{AB}$ in addition to $q_7^{AB}$.]

$q_7^{AB:AZ} = q_7^{AB} *$ _____

(b) The q distribution can be used to predict Z, knowing A and B, from the conditional distribution of Z, given AB, specified by (circle all that are true; 2 pts)

i. $p(ABZ)/q_{AB:AZ:BZ}(AB)$ ii. $q_{AB:AZ:BZ}(ABZ) / p(AB)$
iii. $q_{AB:AZ:BZ}(ABZ) / q_{AB:AZ:BZ}(AB)$ iv. none of the above

**3. [28 points]**
(a) In the analysis of directed systems, where one wants to know which independent variable(s) are statistically significant predictors of a dependent variable, it is common to choose the reference to be the independence (*bottom*) model.

(a1) In such an analysis, one is more concerned with preventing (circle one; 2 pts)

i. Type I errors than Type II errors, & one wants α small (e.g., .05)
ii. Type I errors than Type II errors, & one wants α intermediate (e.g., .10 to .35)
iii. Type I errors than Type II errors, & one wants α large (e.g., .75)
iv. Type I errors than Type II errors, & one wants α very large (e.g., near 1.0)
v. Type II errors than Type I errors, & one wants α small (e.g., .05)
vi. Type II errors than Type I errors, & one wants α intermediate (e.g., .10 to .35)
vii. Type II errors than Type I errors, & one wants α large (e.g., .75)
viii. Type II errors than Type I errors, & one wants α very large (e.g., near 1.0)

(a2) Over-fitting, i.e., using a model with too many parameters and consequently doing poorly when generalizing to test data, is made (circle one; 2 pts)

i. less likely ii. more likely iii. neither more nor less likely

by insisting that, in addition to the 'cumulative' step from reference to model being statistically significant, all incremental steps to the model should also be significant.

(a3) If a 'conservative' model acceptance measure means one that doesn't easily allowing models with many parameters, then write 'most (conservative),' 'least (conservative),' and 'intermediate' after each of the following model acceptance measures (2 pts):

α _____; AIC _____ ; BIC _____

(b) Now suppose one instead chooses the *top* as the reference model.

(b1) Suppose also that while one's primary objective is to have a model that adequately fit the data, one also does want a simple model (but not so simple that it clearly differs from the data). One is therefore more concerned with preventing (circle one; 2 pts)

i. Type I errors than Type II errors, & one wants $\alpha$ small (e.g., .05)
ii. Type I errors than Type II errors, & one wants $\alpha$ intermediate (e.g., .10 to .35)
iii. Type I errors than Type II errors, & one wants $\alpha$ large (e.g., .75)
iv. Type I errors than Type II errors, & one wants $\alpha$ very large (e.g., near 1.0)
v. Type II errors than Type I errors, & one wants $\alpha$ small (e.g., .05)
vi. Type II errors than Type I errors, & one wants $\alpha$ intermediate (e.g., .10 to .35)
vii. Type II errors than Type I errors, & one wants $\alpha$ large (e.g., .75)
viii. Type II errors than Type I errors, & one wants $\alpha$ very large (e.g., near 1.0)

(b2) For a model under consideration, $m_j$, the *null hypothesis* is usually (circle one; 2 pts)

i. $m_j$ is the same as $m_0$, i.e., $T(m_j) = 0$
ii. $m_j$ is different from $m_0$, i.e., $T(m_j) \neq 0$

(b3) One calculates $df(m_0 \rightarrow m_j)$ and $L^2(m_j)$ and decides on some $\alpha_c$. Say that going into the Chi-square table with $df(m_0 \rightarrow m_j)$ and $\alpha_c$ one gets some $L^2_c$ and that the model $L^2(m_j) > L^2_c$. One would then say that (circle; 2 pts)

i. $m_j = m_0$, so $m_j$ is accepted as a good model
ii. $m_j = m_0$, so one uses the data, $m_0$, as one's model
iii. $m_j \neq m_0$, so one tries a model higher in the lattice than $m_j$
iv. $m_j \neq m_0$, so one tries a model lower in the lattice than $m_j$

(b4) 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 monotonic, i.e., either increasing or decreasing) for every step going <u>down</u> the lattice of structures (sequentially from ancestor to descendant) (8 pts):

| | |
|---|---|
| i. uncertainty, H | ii. transmission, T |
| iii. $L^2$ | iv. $\alpha$ |
| v. df (not $\Delta$df) | vi. $\Delta$AIC |
| vii. %correct(training data) | viii. %correct(test data) |

(c) In seeking a balance between minimal error and minimal complexity, the measure that includes a sample size dependent factor in its *complexity term* is (circle one; 2 pts)

i. $\alpha$ ii. AIC iii. BIC iv. none of these

(d1) Let $Inf_{model}$ = the normalized information for model $m_j$, which is 1 for the data, $m_0$, and 0 for the independence model, $m_{ind}$. Write an expression for $Inf_{model}$ in terms of some or all of the following quantities: $H(m_0)$, $H(m_j)$, $H(m_{ind})$. (circle one; 2 pts)

(d2) True or false (circle one; 2 pts): The %reduction of uncertainty of the DV for any model is given by the equation $\%\Delta H_{model}(Z) = \%\Delta H_{data}(Z) * Inf_{model}$.
(e) True or false (circle one; 2 pts): For a good predictive model, $\%\Delta H(Z)$ needs to have a high magnitude comparable to the magnitudes usually desired for correlation measures like $R^2$, e.g., at least 50% or more.

**4. [6 points]**
(a) (4 pts) Consider the set-theoretic mapping below. ABCZ is (circle one):

i. decomposable (without loss) ii. not decomposable (without loss)

*If you chose "decomposable,"* the simplest equivalent structure is _____ and its *predicting* components are (circle one):

i. all deterministic ii. all stochastic iii. some deterministic, some not

*If you chose "not decomposable,"* the number of extra tuples that the first decomposition adds is _____

| A | B | C | Z |
|---|---|---|---|
| 0 | 0 | 0 | 1 |
| 0 | 0 | 1 | 1 |
| 0 | 1 | 0 | 0 |
| 0 | 1 | 1 | 1 |
| 1 | 0 | 0 | 1 |
| 1 | 0 | 1 | 0 |
| 1 | 1 | 0 | 0 |
| 1 | 1 | 1 | 0 |

(b) Suppose one has a frequency distribution, and one decides to *approximate* this information-theoretic (IRA) problem as a set-theoretic (SRA) one by replacing the distribution with a set-theoretic relation, where a state is in the relation if its frequency is equal or greater than some minimum threshold value. For example, one might consider a state in the set-theoretic relation if its frequency is greater or equal to, say, 5, but not in the relation if its frequency is less than 5. Even though reducing the full range of frequency values to a simple binary "either it occurs (as frequently or more frequently

than 5 times) or it doesn't occur (occurs less frequently than 5 times)" loses information, one might do such an approximation because (circle one; 2 pts)

i. the Lattice of Structures is smaller for SRA than for IRA
ii. df can be calculated faster for SRA than for IRA
iii. $\alpha$ is more easily calculated for SRA than for IRA
iv. models with loops do not require iteration in SRA

**5. [8 points]**
(a) Log-linear parametization of frequencies have the form, $F_{ij} = \eta\ \tau_i^X\ \tau_j^Y\ \tau_{ij}^{XY}$. In such representations, the absence of 'an effect,' e.g., deviation from uniformity of a margin or presence of an interaction of non-zero strength, shows up as (circle one; 2 pts)

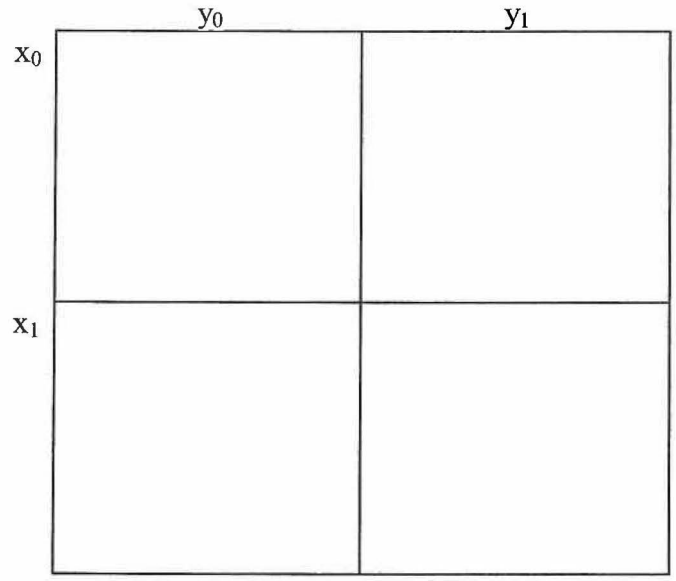
i. all $\tau$ parameters being equal
ii. the $\tau$ parameter(s) for the relevant effect having value 0
iii. the $\tau$ parameter(s) for the relevant effect having value 1
iv. the $\tau$ parameter(s) for the relevant effect having value N (sample size)
v. none of the above

(b) Consider the following distribution:

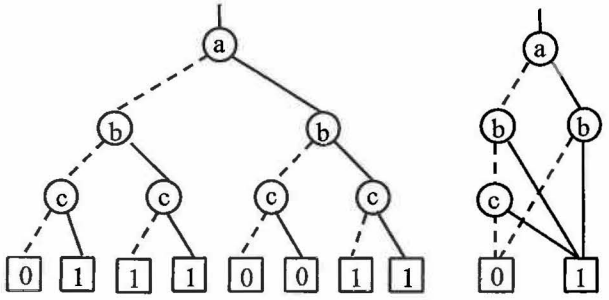
| | $Y_0$ | $Y_1$ | |
|---|---|---|---|
| $X_0$ | .1 | .2 | .3 |
| $X_1$ | .3 | .4 | .7 |
| | .4 | .6 | |

Do composition for model X:Y by imposing the X and Y margins in a *totally new way* as follows: do "Fourier composition" (or "back-projection," not to be confused with "back-propagation," used in neural nets), which sets $q_{X:Y}(x,y) = [\ p(x)/|Y|\ ] + [\ p(y)/|X|\ ] - K$. This method *distributes* a marginal value of an X or a Y equally to all cells that contribute to this marginal value (and at the end of the process subtracts the same constant from each cell to make the sum of the q's equal to 1). Note that this equation looks like a regression equation, where scaled p(x) and p(y) values now add rather than multiply, given that x and y are independent.

Start with cells having zero probabilities (not uniform!). Then take each probability in the X margin (i.e., first .3, then .7), divide this probability by the number of Y states, i.e., the number of columns, and *add* the result to every cell in the row that contributes to this marginal probability (i.e., add .15 to the $x_0y_0$ and $x_0y_1$ cells, and add .35 to the $x_1y_0$ and $x_1y_1$ cells). Do also the corresponding steps for the Y margin. Then subtract a constant, K, from every cell, choosing the constant so that the total calculated probability in the table is 1. (Note that there is no multiple iterations here over the set of projections.) Show the calculations and their results in the table on the next page, and say here what is interesting about the results. (3 pts)




(c) ('Binary decision diagrams') Consider the set-theoretic relation R = {001, 010, 011, 110, 111}. It can be represented by the tree on the left side of the figure below, where lines coming out of (and down from) a variable indicate its possible values, where a dashed line means 0 and a solid line means 1. In the square boxes at the bottom of the tree, 1 indicates that the tuple is in the relation; 0 indicates it is not. So reading down the tree on its left side, the three dashed lines culminating in a boxed 1 mean that 001 (i.e., a=0, b=0, c=0) is in the relation. 000 is not. Etc. By information-preserving operations one can transform the tree on the left to the graph on the right. Reading this graph downwards in the same way, we see that 001 is in the relation, while 000 is not; also 01* is in the relation, as is 11*, but 10* is not. The right hand graph can be considered to specify a 'compressed' relation, $R_C$ = {001, 01*, 11*}.




(c1) Suppose I don't know what the original R distribution was, but I am just given $R_c$. To get a set of abc tuples, I could 'decompress' $R_C$ by replacing 01* by {010, 011} and 11* by {110, 111}. What would be the theoretical justification for making these replacements? (2 pts)

(c2) The graph ($R_C$) is a set-theoretic analog of information-theoretic (circle one; 1 pt)

i. k-systems analysis ii. Bayesian networks
iii. latent variable modeling iv. state-based modeling