Math 322: Biostatistics Midterm Examination

Professor Wickerhauser

Due Thursday, 7 March 2024

You may use the textbook, your notes, and any material on the class website, but you may not collaborate with anyone else. You may use a computer or calculator to perform calculations. Please upload your answers to Canvas/GradeScope well in advance of the deadline.

Some formulas:

- The number of ways to choose k numbers from the set $\{1, \ldots, n\}$ is

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}$$

- Bayes' Theorem: $P(B|A)P(A) = P(A|B)P(B) = P(A \cap B)$.
- A Markov chain transition matrix M is aperiodic if and only if every state is aperiodic.
- Maximums: Differentiable function $f = f(x)$ is maximal at $x = x_0$ only if $f'(x_0) = 0$.

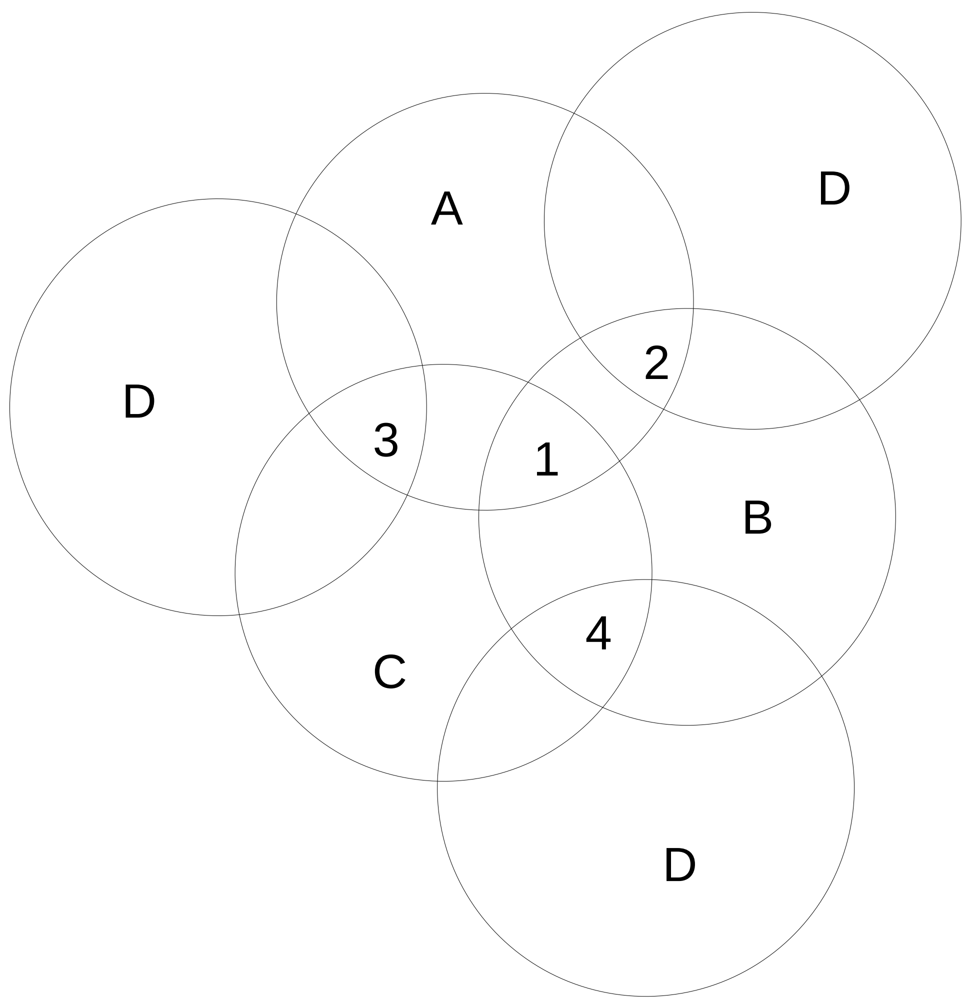
1. (a) Using Venn diagrams, depict four nonempty subsets A, B, C, D satisfying

   - $A \cap B \cap C \neq \emptyset$,
   - $A \cap B \cap D \neq \emptyset$,
   - $A \cap C \cap D \neq \emptyset$,
   - $B \cap C \cap D \neq \emptyset$,
   - $A \cap B \cap C \cap D = \emptyset$,

   (b) Find four subsets A, B, C, D of the set $X = \{1, 2, 3, 4\}$ that meet all the conditions of part a.

   Solution: (a) See the figure below.

   (b) One example is $A = \{1, 2, 3\}$, $B = \{1, 2, 4\}$, $C = \{1, 3, 4\}$, and $D = \{2, 3, 4\}$.



Region 1: $A \cap B \cap C \neq \varnothing$
Region 2: $A \cap B \cap D \neq \varnothing$
Region 3: $A \cap C \cap D \neq \varnothing$
Region 4: $B \cap C \cap D \neq \varnothing$

2. Rosencrantz and Guildenstern decide to play either Tennis or Squash. Since Rosencrantz wins 70% of their Tennis matches, but Guildenstern wins 60% of their Squash matches, they flip a fair coin to choose the game. If it comes up Heads, they play Tennis, otherwise they play Squash. After the match, the loser pays for dinner.

   (a) What is the probability that Rosencrantz pays for dinner?

   (b) If Guildenstern pays for dinner, what is the probability that they played Tennis?

   Solution: First build the conditional probability tree. Here "loses" means "pays for dinner."

   - Tennis (Heads): 0.5
     - Rosencrantz loses: $1 - 0.7 = 0.3$
     - Guildenstern loses: 0.7
   - Squash (Tails): 0.5
     - Rosencrantz loses: 0.6
     - Guildenstern loses: $1 - 0.6 = 0.4$

   (a) Total probability that Rosencrantz loses: $(0.5)(0.3) + (0.5)(0.6) = 0.45$, or 45%.

   (b) Let G be the event "Guildenstern pays" and H be the event "coin is Heads." Then by Bayes' rule,

$$P(H|G) = \frac{P(G|H)P(H)}{P(G)} = \frac{(0.7)(0.5)}{1 - 0.45} = \frac{0.35}{0.55} = \frac{7}{11} \approx 0.636,$$

   where $P(G)$ is 1 minus the probability that Rosencrantz pays (0.45 from part a). □

3. A standard deck of 52 playing cards is thoroughly shuffled so that each card has an equal probability of being anywhere in the deck.

   (a) What is the probability of getting no red cards in a hand of 5 cards? NOTE: there are 26 red cards and 26 black cards in a deck.

   (b) What is the probability of getting exactly 4 red cards in a hand of 5 cards?

   (c) Two players are each dealt 5 card hands from the same deck. What is the probability that one player gets all red cards and the other player gets all black cards?

   You may give your answers as formulas using $\binom{n}{k}$ and $m!$.

   Solution:

   (a) There are $\binom{26}{5}$ ways to have 5 non-red cards, and $\binom{52}{5}$ ways to form a hand of 5 cards, so the probability is $\binom{26}{5} / \binom{52}{5} \approx 0.025$.

   (b) There are $\binom{26}{4}\binom{26}{1}$ ways to have 4 red cards and one black card. Therefore the probability is $\binom{26}{4}\binom{26}{1} / \binom{52}{5} \approx 0.15$.

   (c) There are $\binom{2}{1} = 2$ ways to choose the player that gets the red cards. There are $\binom{26}{5}$ ways choose 5 red cards, out of $\binom{52}{5}$ ways to choose 5 cards from the original deck. There are $\binom{26}{5}$ ways to get 5 black cards from the remaining 47-card deck, out of $\binom{47}{5}$ ways to choose the next 5 cards, so the probability is

$$\frac{\binom{2}{1}\binom{26}{5}\binom{26}{5}}{\binom{52}{5}\binom{47}{5}} = \frac{\binom{2}{1}\binom{26}{5}\binom{26}{5}}{\binom{52}{10}\binom{10}{5}} \approx 0.0022$$

$\square$

4. Suppose that an unknown proportion $p \in [0,1]$ is a random variable with a Beta prior pdf

$$f_{\alpha,\beta}(p) = Cp^{\alpha-1}(1-p)^{\beta-1}$$

with shape parameters $\alpha = 3$ and $\beta = 6$ and normalization constant C. An experiment with success probability p has the following binomial likelihood:

$$L(k|p) = \binom{n}{k} p^k (1-p)^{n-k}$$

Suppose that it produces a count of $k = 7$ successes in $n = 20$ trials.

(a) Find the posterior pdf $P(p|k=7)$ for p. You may ignore normalization constants.

(b) From the pdf of part a, find the MLE, namely the most likely value of p. (Hint: use calculus.)

(c) What was the standard deviation of p before the experiment? What was it after the experiment?

Solution:

(a) By Bayes' theorem, compute

$$P(p|k=7) \propto P(k=7|p)\, f_{3,6}(p) \propto p^{10-1}(1-p)^{19-1}$$

which is proportional to a Beta pdf $f_{10,19}(p)$.

(b) Differentiate P with respect to p and solve

$$0 = P'(p) = 9p^8(1-p)^{18} - 18p^9(1-p)^{17} = 9p^8(1-p)^{17}(1-p-2p) \qquad \Rightarrow p = 0, 1, 1/3.$$

Reject the roots $p = 0$ and $p = 1$, since they correspond to zero likelihood, in favor of the root $p = 1/3$.

(c) From the variance formula for the Beta density:

$$\sigma^2 = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$$

For the prior Beta with $\alpha = 3$, $\beta = 6$:

$$\sigma = \sqrt{\frac{(3)(6)}{(3+6)^2(3+6+1)}} = \sqrt{1/45} = \frac{1}{3\sqrt{5}} \approx 0.149.$$

For the posterior Beta with $\alpha = 10$, $\beta = 19$:

$$\sigma = \sqrt{\frac{(10)(19)}{(10+19)^2(10+19+1)}} = \sqrt{\frac{19}{2523}} \approx 0.087.$$

□

5. Consider the following transition matrix for a 4-state Markov chain:

$$M = \begin{pmatrix} 0.4 & 0.3 & 0.2 & 0.1 \\ 0 & 0.4 & 0.3 & 0.3 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix}.$$

(a) Is M irreducible?

(b) Is M aperiodic?

(c) Find a stationary distribution for M.

(d) Prove that the stationary distribution of part (c) is unique, or else find another.

Solution: (a) M is reducible (that is, not irreducible) since states 3 and 4 can never reach state 1 or state 2.

(b) M is not aperiodic since states 3 and 4 are both periodic of period 2.

(c) Since M is not ergodic, iteration is not guaranteed to converge to a stationary state. However, $\pi = (\,0 \quad 0 \quad 0.5 \quad 0.5\,)$ is a stationary distribution for M easily found by inspection or by the method in part (d).

(d) The stationary distribution from part (c) happens to be unique even though M is not ergodic. This may be shown by linear algebra. Let $\pi = (\,a \quad b \quad c \quad d\,)$ be any stationary distribution, so that $\pi M = \pi$, or

$$(\,a \quad b \quad c \quad d\,) = (\,a \quad b \quad c \quad d\,) \begin{pmatrix} 0.4 & 0.3 & 0.2 & 0.1 \\ 0 & 0.4 & 0.3 & 0.3 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix}.$$

This gives the linear system of equations

$$\begin{aligned} a &= 0.4a \\ b &= 0.3a + 0.4b \\ c &= 0.2a + 0.3b + d \\ d &= 0.1a + 0.3b + c \end{aligned}$$

which implies that $a = 0$, $b = 0$, and $c = d$. But π is a pdf, so $a + b + c + d = 1$, so $c = d = 0.5$. Thus $\pi = (\,0 \quad 0 \quad 0.5 \quad 0.5\,)$ is the unique stationary distribution. $\square$

6. Suppose that K is a positive constant and $f(x,y) = K(x^2 + 2xy + y^3)$ is the joint pdf for random variables X and Y on the square $0 \le x \le 1$ and $0 \le y \le 1$.

   (a) Find K.

   (b) Find the marginal distributions f_X and f_Y of f.

   (c) Determine, with proof, whether X and Y are independent.

   Solution: (a) Since f is a pdf on $[0,1]$ it must satisfy

$$1 = \int_0^1 \int_0^1 f(x,y)\,dxdy = K\frac{13}{12}.$$

   Conclude that $K = 12/13 \approx 0.923$.

   (b) The marginal distribution functions of f are

$$f_X(x) \stackrel{\text{def}}{=} \int_0^1 f(x,y)\,dy = \frac{12}{13}(x^2 + x + \frac{1}{4}),$$

   defined on $0 \le x \le 1$, and

$$f_Y(y) \stackrel{\text{def}}{=} \int_0^1 f(x,y)\,dx = \frac{12}{13}(y^3 + y + \frac{1}{3}),$$

   defined on $0 \le y \le 1$.

   (c) Set $x = 0$ and $y = 0$ to get $f(0,0) = 0$ while

$$f_X(0)f_Y(0) = \left(\frac{12}{13} \times \frac{1}{4}\right)\left(\frac{12}{13} \times \frac{1}{3}\right) \ne 0.$$

   Conclude that X and Y are not independent. $\square$

7. Suppose that X and Y are discrete random variables that have the following joint pdf:

Joint pdf of X and Y

$f(X,Y)$	$Y=1$	$Y=2$	$Y=3$	$Y=4$	$Y=5$
$X=1$	0.1	0.1	0.1	0.1	0.1
$X=2$	0.1	0.1	0.1	0.1	0.1

(a) Compute the marginal pdfs f_X and f_Y.

(b) Determine, with proof, whether X and Y are independent.

(c) Compute the complete conditional pdfs $f_{X|Y}$ and $f_{Y|X}$.

Solution:

(a) $f_X = \begin{pmatrix} 0.5 \\ 0.5 \end{pmatrix}$, $f_Y = \begin{pmatrix} 0.2 & 0.2 & 0.2 & 0.2 & 0.2 \end{pmatrix}$.

(b) X and Y are independent since $f(i,j) = 0.1 = (0.2)(0.5) = f_X(i)f_Y(j)$ for every $i \in \{1,2\}$ and every $j \in \{1,2,3,4,5\}$.

(c)

$$f_{X|Y} = \begin{pmatrix} 0.5 & 0.5 & 0.5 & 0.5 & 0.5 \\ 0.5 & 0.5 & 0.5 & 0.5 & 0.5 \end{pmatrix}; \qquad f_{Y|X} = \begin{pmatrix} 0.2 & 0.2 & 0.2 & 0.2 & 0.2 \\ 0.2 & 0.2 & 0.2 & 0.2 & 0.2 \end{pmatrix}.$$

□