MATH1180
Homework 5
Due Friday, March 3, 2017

Please (re)read Sections 4.1-4.4 and Section 4.5.1 of the Baglivo textbook.

Please submit solutions to the following problems. You *must* use graph paper for problem 8. You may download a graph paper PDF file from the course website. A PDF version of graph paper is available at the course website.

When submitting homework, please remember the following:
- Show all work leading to each solution.
- *Staple* all sheets together. A paper clip is not acceptable.
- Do not submit crossed-out or sloppy work.
- Do not submit ripped or torn pages.
- Be sure to submit your own work.

PROBLEM 1. A new diagnostic test has been designed to detect chromosomal fetal abnormalities. The test is positive for abnormalities in 75% of cases where the fetus has chromosomal abnormalities and in 30% of cases where the fetus does not have chromosomal abnormalities.

Suppose that in a certain population of pregnant women, 1% of fetuses have chromosomal abnormalities.

(a) Write the information provided in terms of probabilities.
(b) Find the predictive value of a positive test.
(c) Find the predictive value of a negative test.

PROBLEM 2 (AF). The Prostate Specific Antigen (PSA) blood test is used to screen for prostate cancer in men over 40 years of age. PSA is a protein produced by the prostate gland. The higher a man's PSA level, the more likely it is that cancer is present. Using a cutoff of 4 (values greater than or equal to 4 indicate a positive test), the sensitivity is 0.86 but the specificity is only 0.33.

Suppose that 12% of those who took the PSA test truly had prostate cancer. Find the predictive value of a positive test and the predictive value of a negative test. Use 4 decimal places of accuracy in your computations and final answers.

PROBLEM 3 (BM). A recent report from the US Health Resources and Services Administration states that co-infection with the human immunodeficiency virus (HIV) and the hepatitis C virus (HCV) is on the rise. Only 1.8% of Americans have HCV, but 25% of Americans with HIV also have HCV. Further, 10% of Americans with HCV also have HIV.

(a) Write the information provided in terms of probabilities. (That is, write each piece of information provided either as "P(??) = ??" or as "P(?? | ??) = ??" for appropriately defined events.)
(b) Find the probability that a randomly chosen American has both HIV and HCV.
(c) Find the probability that a randomly chosen American has either HIV or HCV (or both).

PROBLEM 4 (RG). A hospital uses two tests to classify blood. Every blood sample is tested using both tests. Test A correctly identifies the blood type 73% of the time. Test B correctly identifies the blood type 82% of the time. At least one of the two tests correctly identifies the blood types 90% of the time.

(*a*) Write the information provided in terms of probabilities. (That is, write each piece of information provided either as "P(??) = ??" or as "P(?? | ??) = ??" for appropriately defined events.)
(*b*) Find the probability that both tests A and B correctly identify the blood type.
(*c*) If test A is correct, what is the probability that test B is correct?
(*d*) If test B is correct, what is the probability that test A is correct?
(*e*) Are the events "test A is correct" and "test B is correct" independent, positively associated, or negatively associated? Why?

PROBLEM 5 (SK). Approximately 100 children's products are recalled every year. Children's clothing could be recalled, for example, for drawstrings that are too long or pose a hazard, or material that fails to meet federal flammability standards. Let X be the number of recalls in a given month and suppose that the following table gives the probability distribution of X:

x	0	1	2	3	4	5	6
$P(X = x)$	0.005	0.185	0.275	0.305	0.200	0.020	0.010

(*a*) Find the mean and standard deviation of the number of products recalled in a given month. That is, find $E(X)$ and $SD(X)$.
(*b*) Suppose it is known that there were at least 3 recalls in a given month. What is the probability that at least 5 will occur in that month?
(*c*) If the number of recalls in a given month is above one standard deviation from the mean, the federal government issues a special warning directed toward parents. What is the probability that a special warning will be issued during a given month?

PROBLEM 6 (BIOMETRIKA 42:412–416). Between 1943 and 1946, the eyesight of more than seven thousand women aged 30–39 employed in the British weapons industry was tested. For each woman the unaided distance vision of the right eye (X) and the unaided distance vision of the left eye (Y) were recorded using the four-point scale 1, 2, 3, 4, where 1 is the highest grade and 4 is the lowest grade.

The following table is the joint right eye-left eye distribution for a woman chosen at random from the population of women studied by the British:

	$y = 1$	$y = 2$	$y = 3$	$y = 4$
$x = 1$	0.203	0.036	0.017	0.009
$x = 2$	0.031	0.202	0.058	0.010
$x = 3$	0.016	0.048	0.237	0.027
$x = 4$	0.005	0.011	0.024	0.066

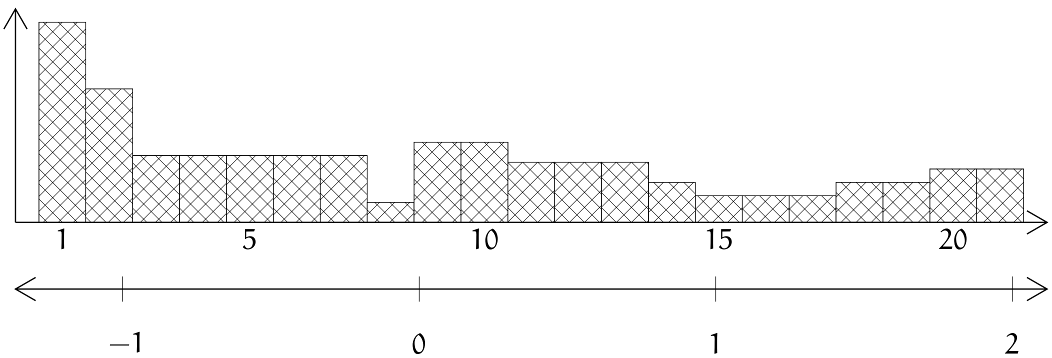
(*a*) Construct a chart showing the probability distribution of X (the score for the *right* eye):

x	1	2	3	4
$P(X = x)$	???	???	???	???

Use your chart to find $E(X)$ and $SD(X)$.
(*b*) Repeat part (*a*) for the probability distribution of Y (the score for the *left* eye).
(*c*) Explicitly compare your answers to parts (*a*) and (*b*) in the context of the eyesight study.
(*d*) Find the following probabilities: $P(X > Y)$, $P(X = Y)$, and $P(X < Y)$. Explicitly interpret these probabilities in the context of the eyesight study.

PROBLEM 7. Let X be the number of days a patient stays at a hospital (before being released or transferred) and suppose that the following table gives the probability distribution of X:

x	1	2	3	4	5	6	7	8	9	10	11
$P(X=x)$	0.150	0.100	0.050	0.050	0.050	0.050	0.050	0.015	0.060	0.060	0.045
x	12	13	14	15	16	17	18	19	20	21	
$P(X=x)$	0.045	0.045	0.030	0.020	0.020	0.020	0.030	0.030	0.040	0.040	



For this random variable, the mean is 8.610 days and the standard deviation is 6.324 days.

(a) Find the probability that X is within 1.5 standard deviations of its mean.

(b) Explicitly compare your answer to part (a) to the appropriate Chebyshev lower bound.

PROBLEM 8. Researchers examined the medical records of the more than 1200 asthmatic children seen regularly at a large inner-city health center. For each child, the researchers determined the number of scheduled quarterly checkups the child had actually kept in the last year (X), and the number of emergency department visits and hospitalizations the child had in the last year (Y).

The following table is the joint checkup-emergency distribution for a child chosen at random from this population of children.

	$y=0$	$y=1$	$y=2$
$x=0$	0.0025	0.0050	0.0425
$x=1$	0.0080	0.0300	0.0620
$x=2$	0.0350	0.0450	0.0200
$x=3$	0.1860	0.0840	0.0300
$x=4$	0.3510	0.0900	0.0090

(a) Compute $E(X)$, and interpret its value in the context of the asthma study.

(b) Compute $E(Y)$, and interpret its value in the context of the asthma study.

(c) The following table gives $(x, E(Y|X=x))$ for each possible x:

x	0	1	2	3	4
$E(Y\|X=x)$	???	1.54	0.85	0.48	???

(i) Find the missing values. Use 2 decimal places of accuracy.

(ii) *Using graph paper*, construct a plot of the

$$(x, E(Y|X=x)) \text{ pairs, for } x = 0, 1, 2, 3, 4.$$

Connect successive pairs using straight line segments. Explicitly interpret the pattern that you see in the context of the asthma study.

(d) For the subpopulation of children who had no emergency room visits and no hospitalizations in the last year, find the average number of quarterly checkup appointments they actually kept in the last year. Explicitly compare your answer to the one obtained in part (a).

PROBLEM 9 (BM, SK). The random variable X is said to be *discrete* if its values form a finite set of numbers or a sequence of numbers (such as 0, 1, 2, ...), and X is said to be *continuous* if its values form an interval on the real line. In each case, indicate whether the random variable X is discrete or continuous.

(a) X is the number of days last week that a randomly chosen child exercised for at least one hour.

(b) X is the running time to the nearest minute of a movie selected at random from the ones playing at the Chestnut Hill cinema.

(c) X is the number of rain delays during a baseball game played in the month of August.

(d) X is the blood alcohol level of an individual arrested for a DUI in Middlesex County, Massachusetts.

(e) X is the thickness of ice 20 feet from the shoreline of Lake Superior during a random December day.

PROBLEM 10 (SL). Let X be the weight (in pounds) of a randomly chosen NFL lineman. Assume that X has mean 323.8 pounds and standard deviation 11.2 pounds.

(a) Find k so that $\mu - k\sigma = 301.6$ and $\mu + k\sigma = 346.0$.

(b) Use the Chebyshev inequality and your answer to part (a) to find a lower bound for the probability that X takes a value in the interval $[301.6, 346.0]$.