Homework #3 – Due Monday, Oct. 13

COR1-GB.1305 – Statistics and Data Analysis

Problem 1 Empirical Rule for Sum of Two Dice

Suppose that you throw two dice. Each die can come up as 1, 2, 3, 4, 5 or 6, and the results from the two dice are independent of each other. We are interested in the random variable X, the sum of the two numbers that land face up. The possible values for X are $2, 3, \ldots, 12$.

(a) *Make a table giving the probability distribution of X. Explain briefly how you did the calculations.*

The first and second rolls are each equally likely to be any of the numbers from 1–6. We can compute a table of the values of X for each pair (first roll, second roll):

	1	2	3	4	5	6
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

Each of the 36 sample points is equally likely. There is one sample point with $X = 2$, two sample points with $X = 3$, etc. Thus, we have the pdf

x	2	3	4	5	6	7	8	9	10	11	12
$p(x)$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

(b) *Show that $E(X) = 7$ and $\operatorname{var}(X) = 210 = 26 = 5.833$.*

$$\begin{aligned}
\mathrm{E}[X] &= (1/36)(2) + (2/36)(3) + (3/36)(4) + (4/36)(5) + (5/36)(6) + (6/36)(7) \\
&\quad + (5/36)(8) + (4/36)(9) + (3/36)(10) + (2/36)(11) + (1/36)(12) \\
&= 7. \\
\operatorname{var}[X] &= (1/36)(2-7)^2 + (2/36)(3-7)^2 + (3/36)(4-7)^2 + (4/36)(5-7)^2 + (5/36)(6-7)^2 \\
&\quad + (6/36)(7-7)^2 + (5/36)(8-7)^2 + (4/36)(9-7)^2 + (3/36)(10-7)^2 \\
&\quad\quad + (2/36)(11-7)^2 + (1/36)(12-7)^2 \\
&= \frac{210}{36} \\
&= 5.833.
\end{aligned}$$

(c) *Although the distribution of X is not a normal distribution, a graph of it would look somewhat bell-shaped. (This is not a coincidence. The more dice you toss, the closer the distribution of* *the sum comes to a normal distribution. More on this later in the course.) For now, let's see how well the empirical rule works. Show that the probability that the z-score for X is between* -1 *and* 1 *is* $24/36 = 0.667$*. Show that the probability that the z-score for X is between* -2 *and* 2 *is* $34/36 = 0.944$.

We have that

$$\mu = E[X] = 7$$
$$\sigma = \sqrt{\text{var}(X)} = \sqrt{210/36} = 2.415.$$

We can compute the z scores for the different values of x using the formula $z = (x - \mu)/\sigma$.

x	2	3	4	5	6	7	8	9	10	11	12
z	-2.07	-1.66	-1.24	-0.83	-0.41	0.00	0.41	0.83	1.24	1.66	2.07

With this table, we can see that

$$\begin{aligned} P(-1 < Z < 1) &= P(X=5) + P(X=6) + P(X=7) + P(X=8) + P(X=9) \\ &= \tfrac{4}{36} + \tfrac{5}{36} + \tfrac{6}{36} + \tfrac{5}{36} + \tfrac{4}{36} \\ &= \tfrac{24}{36} \\ &= 0.6667. \end{aligned}$$

Also,

$$\begin{aligned} P(-2 < Z < 2) &= 1 - \{P(X=2) + P(X=12)\} \\ &= 1 - (\tfrac{1}{36} + \tfrac{1}{36}) \\ &= \tfrac{34}{36} \\ &= 0.9444. \end{aligned}$$

. . . .

Problem 2 Roulette Doubling (Martingale) System

Roulette wheels in casinos in the US have 38 numbers, of which two are green (0 and 00), 18 are black and 18 are red. A bet on black pays at even money, $1 : 1$ *odds (though these odds are clearly not fair due to the green numbers.) Each of the 38 numbers is equally likely to occur on any given spin of the wheel, and results from successive spins are independent. (Casinos expend considerable effort to ensure that these properties hold; otherwise, gamblers would have opportunities for arbitrage.)*

Let's consider a doubling system, at a table with a \$100 minimum bet, and a \$100,000 maximum bet. In terms of \$100 chips, that's a 1-chip minimum and a 1000-chip maximum. To start the system, bet 1 chip on Black. If you win, you're up \$100 and that's the end of the system. If you lose, bet 2 chips on Black. If you win at this point, you're once again up \$100 (having lost one chip and then won two chips), and that's the end of the system. If you lose, bet 4 chips on Black. Continue doubling if you lose. Once you finally win (no matter how long this takes), you will be up \$100, since successive powers of 2 add up to one less than the next power of two, for example, $1 + 2 + 4 + 8 = 15 = 16 - 1$. *So as long as you can keep playing until the first time Black is rolled, you will win your \$100.*

So, what's the catch? Unfortunately, the table maximum of 1000 chips eventually becomes a problem, since if you lose 10 times in a row, you will be down $1+2+4+8+16+32+64+128+256+512 = 1023$ *chips, and you will not be allowed to make the next bet, since a 1024-chip bet would exceed the table limit.*

Suppose that you play this system just once, until either you get your \$100 profit, or you spectacularly go bust with a losing streak of 10 non-Black numbers. Compute the expected net winnings (profit minus loss) for the system, in dollars.

If we play just once, we have

$$\begin{aligned} P(\text{Lose}) &= 1 - P(\text{Win}) \\ &= 1 - \tfrac{18}{38} \\ &= \tfrac{20}{38}. \end{aligned}$$

To go bust with the doubling system, we must lose 10 times in a row. Using independence,

$$\begin{aligned} P(\text{Bust}) &= P(\text{Lose 10 times in a row}) \\ &= \{P(\text{Lose})\}^{10} \\ &= (\tfrac{20}{38})^{10} \\ &= 0.001631 \end{aligned}$$

By the complement rule,

$$\begin{aligned} P(\text{Not Bust}) &= 1 - P(\text{Bust}) \\ &= 1 - 0.001631 \\ &= 0.998369. \end{aligned}$$

Let X be our winnings from playing the doubling system. We either win \$100 or we lose \$102,300 (1023 chips) depending on whether or not we go bust. Thus, X has the probability distribution function (PDF) given in the following table:

x	-102300	100
$p(x)$	0.001631	0.998369

The expected value of X is

$$\begin{aligned} E[X] &= (0.001631)(-102300) + (0.998369)(100) \\ &= -67.01. \end{aligned}$$

That is, we have an expected loss of \$67.01.

. . . .

Problem 3

Suppose a 40-year-old male purchases a \$100,000 10-year term life policy from an insurance company, meaning that the insurance company must pay out \$100,000 if the insured male dies within the next 10 years.

(a) *Use the accompanying life table to determine the insurance company's expected payout on this policy. (Hint: Remember that your universe here is the set of males 40 and older). The age intervals in the table contain all ages from the lower limit up to (but not including) the upper limit.*

First, we compute

$$\begin{aligned} P(\text{dies before age 50}|\text{dies after age 40}) &= \frac{4502}{4502 + 10330 + 19954 + 28538 + 29721} \\ &= 0.04839. \end{aligned}$$

Let X be the payout. Then X has the PDF

x	0	100000
$p(x)$	0.95161	0.04839

The expected payout is

$$\begin{aligned} \mathrm{E}[X] &= (0.95161)(0) + (0.04839)(100000) \\ &= 4839. \end{aligned}$$

That is, \$4,839.

(b) *What would be the expected payout if the same policy were taken out by a 50-year-old male?*

For a 50-year-old male, we have

$$\begin{aligned} P(\text{dies before age 60}|\text{dies after age 50}) &= \frac{10330}{10330 + 19954 + 28538 + 29721} \\ &= 0.11667. \end{aligned}$$

The expected payout is

$$(0.11667)(100000) = 11667.$$

In dollars: \$11,667.

Number of Deaths at Various Ages
Out of 100,000 American Males Born Alive

Age Interval	0–10	1–10	10–20	20–30	30–40	40–50	50–60	60–70	70–80	80 and Over	Total
Number of Deaths	1,527	495	927	1,901	2,105	4,502	10,330	19,954	28,538	29,721	100,000

. . . .

Problem 4

Logic analyzers come off the assembly line with a 3% defective rate. You must ship 17 of these analyzers tomorrow. In this problem we will determine how many analyzers to schedule for production today in order to be reasonably sure that 17 or more of the scheduled machines will work.

Note that if we schedule n machines, and if we let X be the number of machines that work, then X is a binomial random variable with n trials and success probability $p = 0.97$.

(a) *Suppose that we schedule 17 machines. Find $P($all 17 will work$)$.*

$$\begin{aligned} P(X = 17) &= {}_{17}C_{17}\,(.97)^{17}(.03)^0 \\ &= (.97)^{17} \\ &= .5958 \end{aligned}$$

The chance that all 17 will work is %59.58.

(b) *Suppose that we schedule 19 machines. Find $P($at least 17 work$)$.*

$$\begin{aligned} P(X = 17) &= {}_{19}C_{17}\,(.97)^{17}(.03)^2 \\ &= \frac{19 \cdot 18}{2 \cdot 1}(.97)^{17}(.03)^2 \\ &= 171(.97)^{17}(.03)^2 \\ &= 0.0917 \end{aligned}$$

$$\begin{aligned} P(X = 18) &= {}_{19}C_{18}\,(.97)^{18}(.03)^1 \\ &= 19(.97)^{18}(.03)^1 \\ &= 0.3294 \end{aligned}$$

$$\begin{aligned} P(X = 19) &= {}_{19}C_{19}\,(.97)^{19}(.03)^0 \\ &= 1(.97)^{19}(.03)^0 \\ &= 0.5606 \end{aligned}$$

Thus,

$$\begin{aligned} P(X \geq 17) &= P(X = 17) + P(X = 18) + P(X = 19) \\ &= 0.0917 + 0.3294 + 0.5606 \\ &= 0.9817 \end{aligned}$$

The chance that at least 17 will work is 98.17%.

. . . .

Problem 5

If X is a binomial random variable with $n = 100$ and $p = 0.7$, find the mean and standard deviation of X.

$$
\begin{aligned}
E[X] &= np = (100)(0.7) = 70, \\
\mathrm{sd}[X] &= \sqrt{np(1-p)} = \sqrt{(100)(0.7)(0.3)} = \sqrt{21} \approx 4.58.
\end{aligned}
$$

. . . .

Problem 6

A multiple-choice quiz has 10 questions. Each question has five possible answers, of which only one is correct.

(a) *What is the probability that sheer guesswork will yield at least 9 correct answers?*

The number of correct answers, X, is binomial with $n = 10$ and $p = 1/5 = 0.20$.

$$\begin{aligned} P(X \geq 9) &= P(X = 9) + P(X = 10) \\ &= {}_{10}C_9\,(.20)^9(.80)^1 + {}_{10}C_{10}\,(.20)^{10}(.80)^0 \\ &= .000004096 + .0000001024 \\ &= .0000041984 \end{aligned}$$

(b) *What is the expected number of correct answers by sheer guesswork?*

$$E[X] = (10)(.2) = 2.$$

(c) *Suppose that 10 points are awarded for a correctly answered question. How many points should be deducted for an incorrectly answered question, so that for a student guessing randomly, the expected score on a question is zero? (Most standardized tests use this met hod to set penalties for guessing).*

If we let c denote the cost for guessing incorrectly, and if Y denotes the expected score for a question on which a student guesses randomly, then Y has the PDF:

y	$-c$	10
$p(y)$	0.80	0.20

For the expected score to be zero, we must have

$$\mathrm{E}[Y] = (0.80)(-c) + (0.20)(10) = 0,$$

so that

$$c = \frac{(0.20)(10)}{(0.80)} = 2.5.$$

(d) *If a student is able to correctly eliminate one option as a possible correct answer but is still guessing randomly, what happens to his/her expected score for that question? Use your answer to (c) as the number of points being deducted for an incorrect answer.*

In this case Y has the PDF

y	-2.5	10
$p(y)$	0.75	0.25

The expected score is

$$\mathrm{E}[Y] = (0.75)(-2.5) + (0.25)(10) = 0.625.$$

. . . .

Problem 7

The No-Tell Motel has 10 bedrooms. From past experience, the manager knows that 20% of the people who make room reservations don't show up. The manager accepts 15 reservations. If a customer with a reservation shows up and the motel has run out of rooms, it is the motel's policy to pay $100 as compensation to the customer. What is the expected value of the compensation that the motel must pay?

Let X denote the number of customers that show up. This is binomial with $n = 15$ and $p = 0.80$. The manager will have to pay compensation if $X > 10$. We can compute

$$\begin{aligned}
P(X = 11) &= {}_{15}C_{11}\,(.80)^{11}(.20)^4 \\
&= \frac{15 \cdot 14 \cdot 13 \cdot 12}{4 \cdot 3 \cdot 2 \cdot 1}(.80)^{11}(.20)^4 \\
&= 0.1876 \\
P(X = 12) &= {}_{15}C_{12}\,(.80)^{12}(.20)^3 \\
&= \frac{15 \cdot 14 \cdot 13}{3 \cdot 2 \cdot 1}(.80)^{12}(.20)^3 \\
&= 0.2501 \\
P(X = 13) &= {}_{15}C_{13}\,(.80)^{13}(.20)^2 \\
&= \frac{15 \cdot 14}{2 \cdot 1}(.80)^{13}(.20)^2 \\
&= 0.2309 \\
P(X = 14) &= {}_{15}C_{14}\,(.80)^{14}(.20)^1 \\
&= \frac{15}{1}(.80)^{14}(.20)^1 \\
&= 0.1319 \\
P(X = 15) &= {}_{15}C_{15}\,(.80)^{15}(.20)^0 \\
&= (.80)^{15}(.20)^0 \\
&= 0.0352.
\end{aligned}$$

The probability that the manager will not pay compensation is

$$\begin{aligned}
P(X \le 10) &= 1 - (0.1876 + 0.2501 + 0.2309 + 0.1319 + 0.0352) \\
&= 0.1643.
\end{aligned}$$

Let Y denote the amount of compensation that the manager has to pay. We can use the probabilities above to compute the PDF of Y:

y	0	100	200	300	400	500
$p(y)$	0.1643	0.1876	0.2501	0.2309	0.1319	0.0352

The expected value of Y is

$$\begin{aligned}
\mathrm{E}[Y] &= (0.1643)(0) + (0.1876)(100) + (0.2501)(200) + (0.2309)(300) + (0.1319)(400) + (0.0352)(500) \\
&= 208.41.
\end{aligned}$$

The expected compensation is $208.41.

. . . .