Recitation 3

Richard Chen

April 16, 2016

Question 1.

An oil company wants to drill in a new location. A preliminary geological study suggests that there is a 20% chance of finding a small amount of oil, a 50% chance of a moderate amount and a 30% chance of a large amount of oil. The company has a choice of either a standard drill that simply burrows deep into the earth or a more sophisticated drill that is capable of horizontal drilling and can therefore extract more but is far more expensive. The following table provides the payoff table in millions of dollars under different states of the world and drilling conditions

	small	medium	large
standard	\$20M	\$30M	\$50M
horizontal	-\$20M	\$40M	\$90M

1. Find the mean payoffs of the two different drilling strategies.
2. Find the variance in payoffs of each strategy.
3. Which strategy would you advocate for and why?
4. How much are you willing to pay for a geological evaluation that would determine with certainty the quantity of oil at the site prior to drilling?

Question 1 Solution

Suppose the data $x_i, i = 1, 2, 3, \cdots, n$. How to compute the sample mean?

$$\bar{x} = \sum_{i=1}^{n} x_i$$

How to compute the sample variance?

$$s_x^2 = \sum_{i=1}^{n} (x_i - \bar{x})^2$$

That is what we did for the first two homeworks. Then how about the population counterpart, in case we know all the possible values and the correponding probabilities?

Suppose all the values a random variable X can take is $\{x_1, x_2, \cdots, x_n\}$, and the corresponding probabilities is $p_1, p_2, \cdots, p_n$. The popolation mean is

$$E[X] = \sum_{i=1}^{n} x_i \times p_i$$

$$\text{Var}(X) = \sum_{i=1}^{n} (x_i - E[X])^2 \times p_i$$

Now we are ready to compute the mean and variance.

1. The mean of the payoff of standard drilling is

$$\mu_1 = 20 \times 0.2 + 30 \times 0.5 + 50 \times 0.3 = 34$$

The mean of the payoff of horizontal drilling is

$$\mu_2 = -20 \times 0.2 + 40 \times 0.5 + 90 \times 0.3 = 43$$

2. The variance of the payoff of standard drilling is

$$s_2^2 = (20 - 34)^2 \times 0.2 + (30 - 34)^2 \times 0.5 + (50 - 34)^2 \times 0.3 = 124$$

The variance of the payoff of horizontal drilling is

$$s_2^2 = (-20 - 34)^2 \times 0.2 + (40 - 34)^2 \times 0.5 + (90 - 34)^2 \times 0.3 = 1461$$

3. It depends on the risk aversion. Although the standard drilling has less payoff than horizontal drilling on average, it enjoys smaller variation.

4. We can evluation the maximal amount of money the company would be willing to investing in the geological investigation. The expected gain in the payoff depends on its original plan:

1) if the company originally is going to adapt the stanard drilling, after the geological evaluation, they would change to horizontal drilling with probability 80% (the probablity of moderate amount and the probability of large amount 50%+30%, if the amount turns out to be small with probability 20%, there will be no motivation for them to change the plan), so the expected gain in payoff is

$$(40 - 30) \times 0.5 + (90 - 50) \times 0.3 = 17$$

in this case, the company would like to pay no more than 17 millions dollars for the geological evaluation.

2) if the company originally is going to adapt the horizontal drilling, after the geological evaluation, they would change to standard drilling with probability 20% (the probablity of small amount, if the amount turns out to be either moderate or large, there will be no motivation for them to change the plan to incur the decrease in the payoff), so the expected gain in payoff is

$$(20 - (-20)) \times 0.2 = 8$$

in this case, the company would like to pay no more than 8 millions dollars for the geological evaluation.

Question 2

Cooper Realty is a small real estate company located in Albany, New York, specializing primarily in residential listings. They have recently become interested in determining the likelihood of one of their listings being sold within a certain number of days. Based on historical data, they produced the following figures based on the past 800 homes sold.

Days Listed until Sold	Under 20	31-90	Over 90	Total
Under $50K	50	40	10	100
$50-$100K	20	150	80	250
$100-$150K	20	280	100	400
Over $150K	10	30	10	50

1. What is the probability that a randomly selected home is listed over 90 days before being sold?
2. What is the probability that a randomly selected initial asking price is under $50K?
3. What is the probability of both of the previous two events happening? Are these two events independent?
4. Assuming that a contract has just been signed to list a home that has an initial asking price less than $100K, what is the probability that the home will take Cooper Realty more than 90 days to sell?

Question 2 Solution

This question asks us to estimate the probability to calculate the marginal probability.

We can approximate the probability of a certain event by the frequency of this event. For example, the coin toss experiment in history.

For marginal probability, we need the enlarge the table and get an new one:

Days Listed until Sold	Under 20	31-90	Over 90	Total
Under \$50K	50	40	10	100
\$50-\$100K	20	150	80	250
\$100-\$150K	20	280	100	400
Over \$150K	10	30	10	50
Total	100	500	200	800

1. What is the probability that a randomly selected home is listed over 90 days before being sold?
$$\frac{200}{800} = 25\%$$
2. What is the probability that a randomly selected initial asking price is under \$50K?
$$\frac{100}{800} = 12.5\%$$
3. What is the probability of both of the previous two events happening? Are these two events independent?
$$\frac{10}{800} = 1.25\%$$
4. Assuming that a contract has just been signed to list a home that has an initial asking price less than \$100K, what is the probability that the home will take Cooper Realty more than 90 days to sell?
$$\frac{10+80}{100+250} = \frac{9}{35} \approx 25.71\%$$

Question 3

The following joint probability distribution is based on survey data collected by a major financial publication in 2002. The columns represent the percentage of retirement income invested in the stock market (X) by age division (Y).

		X: 10%	X: 30%	X: 60%
Y	$<$ 30 yrs old	0.04	0.05	0.01
Y	30-50 yrs old	0.05	0.23	0.19
Y	$>$ 50 yrs old	0.10	0.26	0.07

1. Among persons investing 60% of retirement income in the stock market, what proportion are younger than 30?
2. Among persons younger than 30, what percentage invest more than 10% of their retirement income in the stock market?
3. Re-express the information in the table as the conditional probability of X *given* Y.
4. What is the conditional expectation of X given that $Y = \{\text{older than } 50\}$?
5. Is Y and X positively or negatively correlated?

Question 3 Solution

We can get the enlarged table

		X: 10%	30%	60%	
Y	< 30 yrs old	0.04	0.05	0.01	0.10
	30-50 yrs old	0.05	0.23	0.19	0.47
	> 50 yrs old	0.10	0.26	0.07	0.43
		0.19	0.54	0.27	

1. Among persons investing 60% of retirement income in the stock market, what proportion are younger than 30?

$$0.01/(0.01 + 0.19 + 0.07) = \frac{1}{27} = 0.037$$

2. Among persons younger than 30, what percentage invest more than 10% of their retirement income in the stock market?

$$(0.05 + 0.01)/(0.04 + 0.05 + 0.01) = \frac{6}{10} = 0.6$$

3. Re-express the information in the table as the conditional probability of X *given* Y.

		$X \mid Y = y$: 10%	30%	60%
y	< 30 yrs old	0.4	0.5	0.1
	30-50 yrs old	0.106	0.49	0.404
	> 50 yrs old	0.232	0.605	0.163

4. What is the conditional expectation of X given that $Y = \{\text{older than } 50\}$?

$$E(X \mid Y = \{\text{older than } 50\}) = 10\%(0.232) + 30\%(0.605) + 60\%(0.163) = 30.25\%$$

5. Anything informative about Y and X positively or negatively correlated?

$$E(X \mid Y = \{\text{30-50 yrs}\}) = 10\%(0.106) + 30\%(0.49) + 60\%(0.404) = 40\%$$

$$E(X \mid Y = \{\text{younger than } 30\}) = 10\%(0.4) + 30\%(0.5) + 60\%(0.1) = 25\%$$

Question 4

In a sample of 100,000 emails you find that 550 are spam.

Your next email contains the word “bigger.” From historical experience, you know half of all spam email contains the word “bigger” and only 2% of non-spam emails contain it. Find the probability that this new email is spam.

Question 4 Solution

Let S denote the event that an email is a spam, NS is that the email is not a spam. Let B denote the event that an email contains the word "bigger".

Let mathematical formulate this question:

1)What we want the compute:

$P(S|B)$ the probability of an email is a spam given that the email contains a work "bigger".

2) What we do know:

$P(S) = \frac{550}{100000}$ is the probability that any given email is a spam;

$P(B|S) = \frac{1}{2}$ is the probability that an email contains "bigger" given it is a spam;

$P(B|NS) = \frac{1}{50}$ is the probability that an email does not contain the word "bigger" given it is not a spam.

$$\begin{aligned}
\mathrm{P}(S|B) &= \frac{\mathrm{P}(S \text{ and } B)}{P(B)} \qquad \text{(Byes theorem)} \\
&= \frac{P(B|S)P(S)}{P(B \text{ and } S) + P(B \text{ and } NS)} \\
&= \frac{P(B|S)P(S)}{P(B|S)P(S) + P(B|NS)P(NS)} \\
&= \frac{\frac{1}{2} \times \frac{550}{100000}}{\frac{1}{2} \times \frac{550}{100000} + \frac{1}{50} \times \frac{99450}{100000}} \\
&\approx 12.15\%
\end{aligned}$$