1. (6 pts) One of the hardest problems is forming an appropriate probability model. E.g., suppose you're working for Verizon deciding how much data capacity your network will need
once it starts selling the iPhone. Suppose that you know that each customer will use 5GB/month. Since a month has about 2.5M seconds, does that mean that your network will
need to provide only 2000KB/s per customer? What might be wrong with this model? How
might you make it better? (This is an open-ended question; any reasonable answer that shows creativity gets full points.)

Answer：

No, we can’t assume our network will provide only 2000KB/s per customer.
Possible answers:
● We can’t assume the customers will use the constant value of 2000KB all the time. They may use less or more than 2000KB.
Improved possible ways:
● We assume customer will use most of the data during the night time or other specific time.
● We set level data limit for customer, according to their request to upgrade the data.

2. (3 pts) One hard problem with statistics is how they should be interpreted. For example, mental health care professionals observe that young men with schizophrenia usually smoke pot (marijuana). Assuming for the sake of argument that this correlation is real, does this mean that pot smoking causes schizophrenia?

Historical note: In 1974, the question of whether cigarette smoking causes lung cancer was answered by forcing some dogs in a lab to smoke and observing that they got cancer more than otherwise identical dogs not forced to smoke.

The tobacco companies were maintaining that the strong correlation between smoking and lung cancer (1/4 of smokers died from cancer) did not demonstrate a causal relation. Maybe there was a common cause for both a desire to smoke and a likelihood to later get cancer. These experiments refuted that claim.

Answer：

We couldn’t assume these two things are correlated

Any reasonable answers will receive full credit.

3. (6 pts) (1.3) Explain under what conditions the following experiments are equivalent to a
random coin toss. What is the probability of heads in the experiment?
(a) Observe a pixel (dot) in a scanned black-and-white document.
(b) Receive a binary signal in a communication system.
(c) Test whether a device is working.
(d) Determine whether your friend Joe is online.
(e) Determine whether a bit error has occurred in a transmission over a noisy communication
channel

Answer:
(a) If the document is black and white, then the dot must be black and white. If the black and
white have the equal chance to appear, then it is equivalent to tossing a fair coin If the black
is head and white is tail, the probability of head is $1/2$.
(b) The signal system will either receive an binary or non-binary signal. If the chance to receive binary signal is equal the chance to receive non-binary signal then this experiment is
equivalent to tossing a fair coin. Assume receiving binary signal as head and non-binary as
tail, the probability is $1/2$.
(c) The machine can only be working or not working. If chances of these two conditions happen are equal which is the same to tossing a fair coin. Assume machine working properly
is head, the probability of the head is $1/2$.
(d) Joe can only be online or not online. If the chance that Joe is online is equal to Joe is not
line, it is same as tossing a fair coin. Assume Joe is online as head and not online as tail,
the probability is $1/2$.
(e) The bit error can either occur or not occur. If chances of these two conditions happen are
equal, it is equivalent to tossing a fair coin. Assume bit error occur as head and not occur as
tail, the probability of head is probability $1/2$.

4. (6 pts) (1.6) A random experiment consists of selecting two balls in succession from an
urn containing two black balls and one white ball.
(a) Specify the sample space for this experiment.
(b) Suppose that the experiment is modified so that the ball is immediately put back into the
urn after the first selection. What is the sample space now?
(c) What is the relative frequency of the outcome (white, white) in a large number of repetitions of the experiment in part a? In part b?
(d) Does the outcome of the second draw from the urn depend in any way on the outcome of
the first draw in either of these experiments?

Answer：
(a) S = {(black, black), (black, white), (white, black)}
(b) S = {(black, black), (black, white), (white, black). (white, white)}
(c) In (a), the (white, white) is impossible therefore the frequency is 0.
In (b), it contained 1 white ball and 2 black balls, therefore the probability to get a white ball is $1/3$, the frequency is $1/3 \times 1/3 = 1/9$.
(d) In (a), the second draw depends on the first draw.
In (b), the second draw doesn't depend on the first draw.

5. (6 pts) (1.11) In order to generate a random sequence of random numbers you take a column of telephone numbers and output a "0" if the last digit in the telephone number is even and a "1" if the digit is odd. Discuss how one could determine if the resulting sequence
is "random." What test would you apply to the relative frequencies of single outcomes? Of
pairs of outcomes?

Answer：
There are two possibilities for the last digit of the telephone.
0: last digit is even
1: last digit is odd
Each with probability is $1/2$, to test and determine if the resulting sequence is random, we can
take a large sample and calculate the result whether the relative frequency of the last digit in
the telephone is $1/10$. For pairs of outcomes, we can check if each outcome has a relative
frequency of $1/100$. To test the relative frequencies, we can use binomial or any other reasonable probability test model.