## Quizz 2 10/12/2015

1. What do you know about x2(chi-squared) test?

<http://www2.lv.psu.edu/jxm57/irp/chisquar.html>

Chi-square is a statistical test commonly used to compare observed data with data we would expect to obtain according to a specific hypothesis. The chi-square test is always testing what scientists call the **null hypothesis,** which states that there is no significant difference between the expected and observed result.

x2 = sum(observed-expected)^2/expected

df: k-c

k is number of categories

c is number of estimated parameters for the distribution+1

<http://www.itl.nist.gov/div898/handbook/eda/section3/eda35f.htm>

2. Given a 2 by 3 grid (which has 6 blocks and 17 edges), shortest route to visit all edges

(assuming edge length is 1).

<http://ie454.cankaya.edu.tr/uploads/files/Chp-03%20044-064.pdf>

The answer is 19. This is a variation of Chinese Postman Problem (CPP).

CPP: a postman needs to deliver mail to a number of streets from the postoffice and come back to the postoffice in the end with shortest distance.

A traversable graph is one that can be drawn without taking a pen from the paper and without retracing the same edge. In such a case the graph is said to have an Eulerian trail.

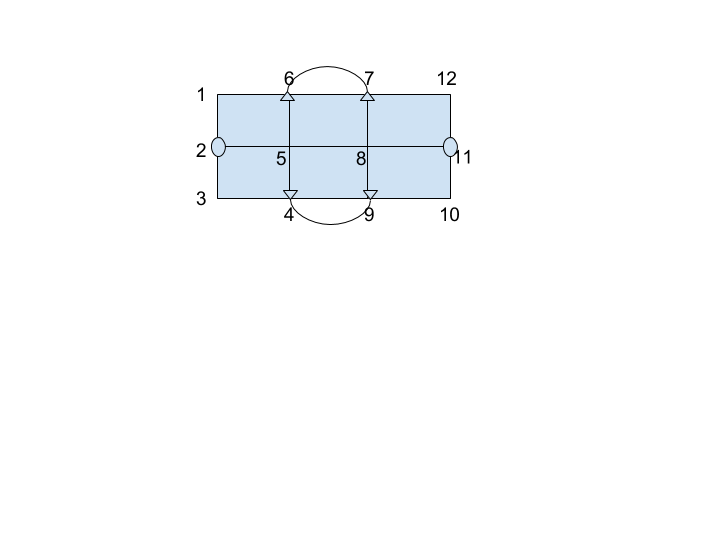
An Eulerian trail uses all the edges of a graph. For a graph to be Eulerian all the vertices must be of even order.

If a graph has two odd vertices then the graph is said to be semi-Eulerian. A trail can be drawn starting at one of the odd vertices and finishing at the other odd vertex.

If there are odd vertices, we need to pair them (connect them with possible routes). For example, if there are two odd vertices, there’s only one way of pairing them together; if there are 4 odd vertices, it would be 3 ways of pairing; if there are 6 odd vertices, it would be 5x3=15 ways of pairing; generally, if there are n odd vertices, there would be (n-1)x(n-3)x(n-5)...1 ways of pairing.

If there are weights between each vertices, then we can select the shortest weights of possible pairing, and then add them to the original weights to get the shortest distance for the postman.

For the number of edges, we need to sum the half of order for each vertices (after pairing), then plus one (because the postman needs to come back to the post office).



For our problem, the postman doesn’t need to come back to the start point, and the start point can be any vertices.

The vertices with odd orders are 2, 4, 6, 7, 9, 11.

To find the shortest distance, we just need to pair the shortest odd vertices. That is, connect 6 and 7, 4 and 9. In this way, we just add two more edges. We leave 2 and 11 unchanged. That is, we can start from 2 and end at 11. One shortest route would be 2->5->6->1->2->3->4->5->8->9->4->9->10->11->8->7->6->7->12->11. Total distance is 17+2=19.

3. X, Y are iid N(0; 1), calculate P(X|X + Y > 0), try not use density function of joint

distribution.

By Bayes’ Theorem, , where is the pdf of . Let . Since , . Hence, . Note , where is the cdf of . Therefore, .

(We could use density function of joint distribution to confirm this result. By hypothesis, we have , where is the pdf of . Let . Then the determinant of Jacobian matrix : is computed to be . Hence, . Note

, which is the same as the above answer.)

4. You have a six sided dice, you can keep rolling the dice and you get the dollars equal to the

mount of the sum. However, if at some point, the sum is a square number, you must stop and

will get zero dollars. (1) If at some point, your sum is 35, should you stop or keep rolling?

(2) in (1), if you choose to continue and this is your strategy: you will keep rolling until you

exceed 43, what is the most probable amount of dollar you win when you stop? (3) Is there

a best strategy for this game, any number that you should stop ?

<http://math.stackexchange.com/questions/1176195/would-you-ever-stop-rolling-the-die>

<http://math.stackexchange.com/questions/977679/toss-a-fair-die-until-the-cumulative-sum-is-a-perfect-square-expected-value>

(1) If the next is 1, then we will get 0; if the next is not 1, then we can safely keep rolling until we get 43 (at least). So the expectation is at least ⅙\*0+⅚\*43 ~=35.8. So we should keep rolling.

(2) We need to consider the most probable case for this question. Now we have 35, the most probable case for next rolling is not getting zero, that is, we get 2,3,4,5,6 for the next rolling. They are equally likely. If we get 2, then we can roll at least twice more without any risk. Now, we consider all the cases of cumulative sum of two rolling. That is,

1st: 1, 2, 3, 4, ,5, 6

2nd: 1, 2, 3, 4, 5, 6

2, 3, 4, 5, 6, 7

3, 4, 5, 6, 7, 8

4, 5, 6, 7, 8, 9

5, 6, 7, 8, 9, 10

6, 7, 8, 9, 10, 11

7, 8, 9, 10, 11, 12

If we first get 2 after 35, then we can do at least two rolling with most probable sum to be 7+2=9.

If we get 3, there would be two cases:

1 rolling: 6 -> 3+6=9

1 rolling: 1, 2, 3, 4, 5

2nd rolling: 1, 2, 3, 4, 5, 6 => most probable is 6+3=9 and 7+3=10.

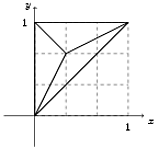
So the total most probable case is 9.

The 4, 5, 6 cases are similar, and they all include 9 as one of the most probable case. Because 2, 3, 4, 5, 6 are equally likely for the first rolling after 35. So the most probable amount would be 35+9=44.

(3) The only time we might consider stopping is when kk is nearly a perfect square. Suppose k≈n2(more specifically, take n2−6≤k<n2).

If we stop, our score is always at least n2−6. If we try to continue past n2, and then stop when we nearly reach (n+1)2, the probability that we actually make it past n2 is at most 5/6, so the expected score is at best 5/6((n+1)2−1). So stopping now is better than going through (n+1)2 whenever n2−6>5/6(n2+2n), and possibly sooner. This inequality holds whenever n≥13.

5. Break the unit interval at xx and yy where x<yxy. Our lengths are then xx, y−xyx, and 1−y1y. It's not hard to show that they all have probability 1/313 of being the shortest. In any case, our joint PDF is given by f(x,y)=6fxy6 (since xx and yy remain uniform random variables on 1/616th of the square [0,1]×[0,1]0101). Each triangle in the diagram below corresponds to the domain of the PDF for one of the three cases.



I'll take care of the case when xx is shortest, that is, x≤y−xxyx and x≤1−yx1y. This is the leftmost triangle. Since we're assuming xx is least, we are looking for

E[x]=∫1/30∫1−x2x6xdydx=1/9Ex0132x1x6xdydx19

The cases when y−xyx and 1−y1y are shortest are similar.

<http://math.stackexchange.com/questions/13959/if-a-1-meter-rope-is-cut-at-two-uniformly-randomly-chosen-points-what-is-the-av>

6. At a party, N people throw their hats (all hats are different) into the center of room. The

hats are mixed up and each people randomly selects one. Let Y be the number of people who

select their own hats. Now ask (1) what is the expectation of Y ? (2) what is the variance

of Y ? Now, the picking hats game rule is extended. For each hats pick round, the people

choosing their own hats quit the game, while others (those picked wrong hats) put their

selected hats back in the center of room, mix them up, and then reselect. Also, suppose that

this game continues until each individual has his own hat. Suppose N individuals initially

join the game, let R(N) be the number of rounds that are run and S(N) be the total number

of selections made by the these N individuals, (N > 1). (3) Find the expectation of R(N).

(4) Find the expectation of S(N). (5) Find the expected number of false selections made by

one of the N people.

1. Let be the indicator variable that equals to when the th person select his own hat and equals to otherwise. Then it is obvious that . Since , .
2. Since , we need to calculate . Note , because and . Hence, we have .

7. Consider linear regression of Y on features X1, X2: Model1-(Y;X1), R2 = 0.1; Model2-

(Y;X2), R2 = 0.2; Model3-(Y;X1;X2), calculate the range of R2 of Model3.

Let , and be the coefficients of determination for model1, model2 and model3, respectively. According to an equation on page 44 of our Econometrics book, , where is the partial correlation between and controlling for .

8. Given a function for a fair coin, write a function for a biased coin that returns heads with

probability 1/n (n is a param).

flip the coin k times ( where k is the minimum number of bits required to represent n in binary form). if ith flip is head, make ith bit 1, 0 otherwise. Let the new number formed is x, then

if(x == n-1)

return head;

if(x >= n)

repeat the process;

return tail;

An equivalent solution in C:

// return true with probability 1/n

boolean flip(int n)

{

// random is random in range 0..(range-1)

int random=0, range=1;

while( true ){

while( range<n ){

random = random\*2 + fairFlip()?1:0;

range \*= 2;

}

if( random<n ){

return random==0;

} else {

random -= n;

range -= n;

}

}

}

<https://www.ocf.berkeley.edu/~wwu/cgi-bin/yabb/YaBB.cgi?board=riddles_cs;action=display;num=1302721993>

9. Denote expected number of bridges crossed to reach island , then , from which we get . When , we have , from which we get . Observe that we can change this equation a little bit, such that . Let , then is a geometric sequence. So . Hence,

(an alternative solution: the bridges can be treated as states in a Markov Chain, or MC. Denote as the time of absorption of the MC, and denote as the state at time , where . Then the original question is actually asking to compute . Let . By the first step analysis, we have for , and

. Solving this set of equations yields .)