AMS 550.400 Revised HW SET 1 Due Date: Oct 31

Last Compiled on October 31, 2012

Zhendan Zhu

Problem 1 (10 pts):

In this problem, I used the following code:

cd ~/550400
mkdir ~/550400/honda
cd ~/550400/honda
git init
vi main.txt
git add .
git commit -m "A is done"
vi main.txt
git add .
git commit -m"B is done"
git checkout -b alt
vi main.txt
git add .
git commit -m"X is done"
git checkout master
vi main.txt
git add .
git commit -m "C is done"
git merge alt
vi main.txt
vi main.txt
git add .
git commit -m "D is done"
git log –graph –oneline
git checkout alt
git log –graph –oneline
git push https://github.com/zhendanzhu/honda.git master
git push https://github.com/zhendanzhu/honda.git alt

The graphs are as following:

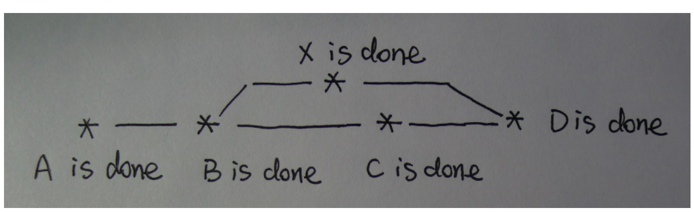


Figure 1: The history graph for master branch

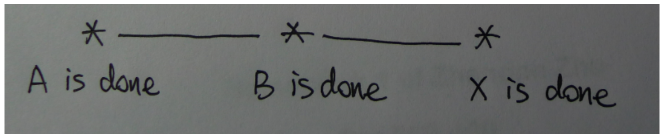


Figure 2: The history graph for alt branch

Problem 2 (10 pts):
mkdir newpoem
cd ˜/newpoem
git config –global user.name ”zhendanzhu”
git config –global user.email zhendanzhu@hotmail.com
git remote add stanza1 git://github.com/nhlee/550400.stanza1.git
git remote add stanza2 git://github.com/nhlee/550400.stanza2.git
git remote add stanza3 git://github.com/nhlee/550400.stanza3.git
git init
git checkout master
git pull stanza1 master
vi main.txt
git add .
git commit -m ” add a title ”
git checkout -b alt1
git pull stanza2
git checkout master
git merge alt1
vi main.txt
git add .
git commit -m ”resolve conflict1”
git checkout -b alt2
git pull stanza3
git checkout master
vi main.txt
git add .
git commit -m ”resolve conflict2”
git remote add origin https://github.com/zhendanzhu/poemmerge.git
git push -u origin master

Problem 3 (40 pts): Consider a team of four students, say, A, B, C and D, who just started working on writing a `latex/beamer` file, say `main.tex`, for a class presentation of their work statement. Assume that they do not wish to coordinate their schedules for a concurrent group meeting (both virtually and physically). Assume that:

- A is in charge of *Introduction*,
- B is of *Problem Statement*,
- C is of *Timeline*,
- D is of *Deliverable* part of the presentation.

In other words, their contributions to `main.tex` do not overlap. Then,

- first, devise a work flow strategy for the team so that they can collaborate asynchronously using `git`,
- next, devise yet another `git` strategy different from your earlier proposal.

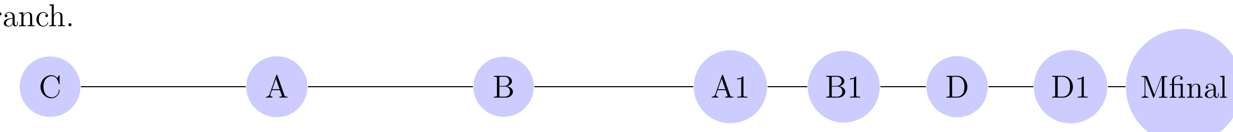
Finally,

- discuss the strength and weakness of each of your proposed strategies in terms of merge conflicts resolution,
- make the final recommendation.

In order to answer this question, *build* a mathematical model, *following* the guideline from IMM. Use Section 1.4 and Section 1.5 of IMM as *role models*. For example, you are to identify which variables are exogenous and which are endogenous. More specifically, among other things, in your model, is the preamble part of `main.tex` an endogenous or exogenous variable? Note also that in addition to this issue, there are other issues that you are to consider. So, *be sure to consult IMM*.

Problem: Try to reduce the time spent on resolving conflicts with an optimal git workflow strategy. In this problem, there are two endogenous variables: total number of editions and total number of conflicts. There is one exogenous variable: the time we spend on merging. Either increasing total number of editions or total number of merging conflicts will increase the time we spend. In our workflow strategy, we will try to reduce the time we merge files together. When will there be conflicts? Well, if one person has updated a part of the file while the others work on the old edition, there will be a conflict. It will cost a lot of time and energy to deal with the conflicts. The best way is designing a git workflow structure that reduces conflicts as well as total number of editions.

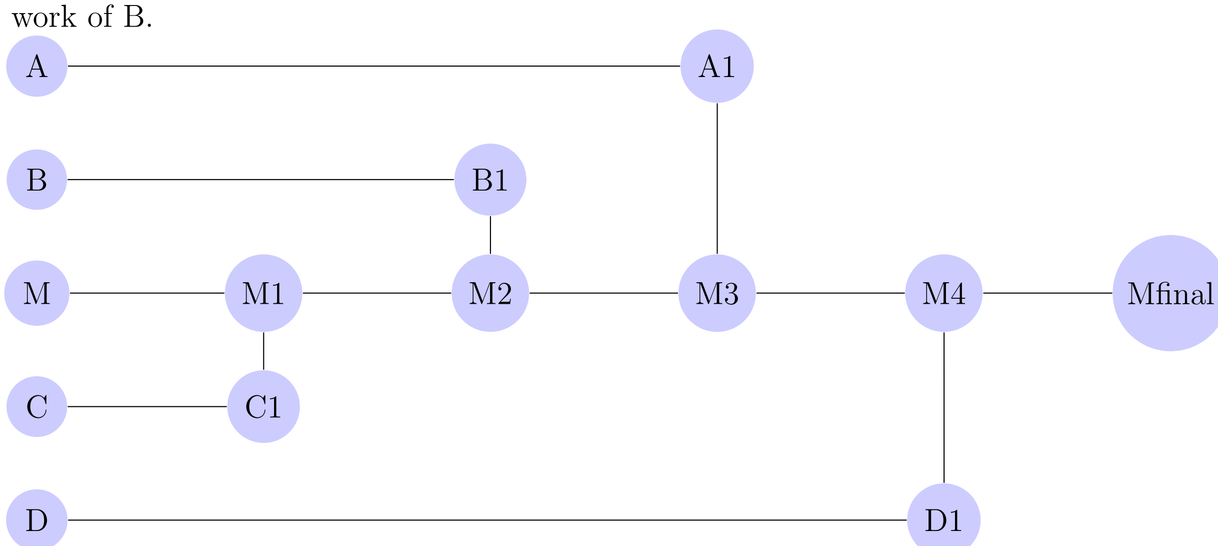
Outline for the model:
Proposal plan 1: As the following graph shows. All the work will be done in master branch, and each person pull it from the master's branch, update the file and push it back to the master's branch.



The strength for this plan is keeping each person's newest edition updated in time. However, the workflow permits only one person work in one period. When he is done, the other person can go on to update the file without overwrite it. If two person work in the same time, it requires many merges during the pulling process. As the previous graph shows, if each person push one original file and update the file for once, there will be in total 7 merges. We'll lose the efficiency of the project.

Proposal plan 2: Since A ,B ,C and D have relatively independent part of the project, each one can work on his/her own branch first, and merge to the master's branch when it is done. Considering different person may work asynchronously. Each time there's a conflict in merging, we can simply keep what it is for our own part and wait for the other one to update their part. Since C is responsible for timeline, it can be done independently in one file. If eventually C finished first, then A, B, D can follow the timeline to do their part and update their work in a fixed schedule. If C haven't finish his part, then A, B, D can work in a free way as long as they pull the latest edition from the github before they start working on their part.

Strength: Each one of the team will have its independent part, so that we can reduce the number of merge conflicts. Thus it will reduce a lot of time in merging part.
Weakness: the merging part will be a headache when B merged his file with master branch if A, C, D still work on the old edition. So someone from A, C, D needs to resolve the conflicts of merge the work of B.



Final Recommendation: plan 2 will be better. Compared with plan 1, plan 2 has same number of editions and less merging conflicts, which will save time effectively. On the other hand, each member of the team can keep track of the others work (individual branch) while have a clean idea of the main project (master branch).

Problem 4 (aka. Fair Play, 40 pts): Answer the following question:

Is the tennis game fair?

Note that unlike Problem 3, this question is vaguely stated. This is intensional, whence to begin, you will first need to clarify what exactly your question is. You may use the class discussion on this particular problem, but you *may not* directly refer to our discussion. Instead, formulate the model carefully but concisely in your own words.

Original Problem:
Is the game fair? A general standard for this is if the roles of the competitors are reversed, their probability of winning does not change. Our original problem can be broken down to several parts: whether the player who is first to serve will be in advantage? What will be the chances for the first game server to win the match given the probability of winning rate for each ball? And to what extent will the advantage be?

Outline for the model: Based on the winning rate of each game, we can infer the probability that the first server to win the match. According to tennis rule, one player delivers the ball to start the game, called server; and one player receives the ball, called receiver. We simplified the rule so that any person who wins two straight points will win the game. The person who wins 6 games is the winner in a match. The possible conditions for a game are presented as following:

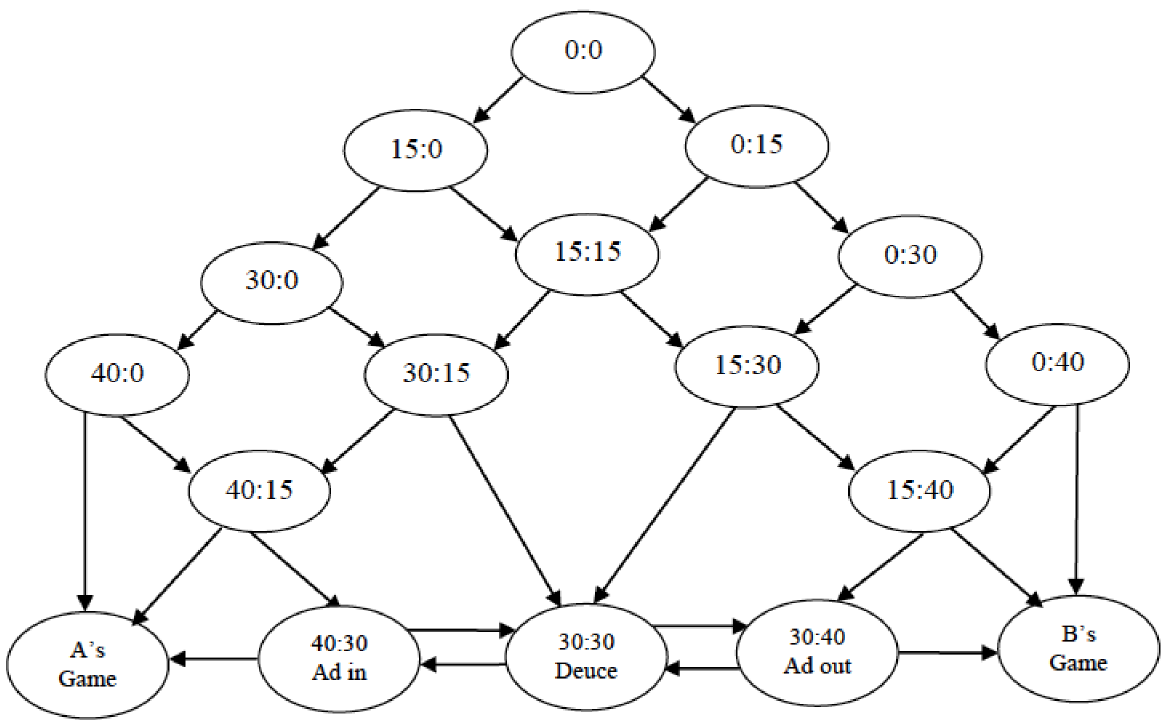


Figure 3: The graph for possible score results

Condition: For both players, the chances for the server to win is P, and the probability for the receiver is $1 - P$.

Formulate the Problem: Whether or not the game is fair depends on the server's winning rate P on each ball. For each player, the chance to win a game is the same, it equals $Q = P^2/(P^2+(1-P)^2)$. The chance to lose the game is $(1 - P)^2/(P^2 + (1 - P)^2)$. If P>1/2, then the winning rate for the server in each game is bigger than 1/2. Given the rule for winning a tennis match is the one who wins the first 6 games will win. The final score can be "6:0", "6:1" , "6:2", etc. The total chance for the first server to win is

$$\sum_{k=0}^{5} \binom{k+5}{6} Q^6 (1 - Q)^k.$$

We are going to apply the historical data to back test the estimated rate. The best way is comparing the winning rate of each player as first-server with the winning rate as first-receiver against the same player. Then we are going to test whether there's significant difference between these two rates by two sample T-test.

The value for this function is show as following

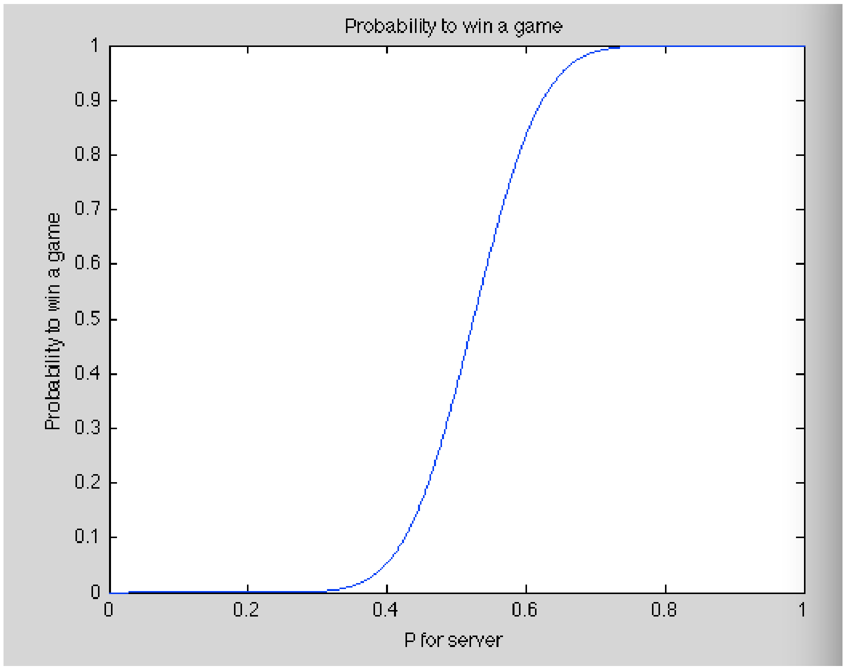


Figure 4: The graph for possible winning rate

Code is as following:

```
b=zeros(1,1000); c=zeros(1,5);
for Q=1:1000;
 for q=1:5;
     p=(Q/1000)^2/((1-Q/1000)^2+(Q/1000)^2);
     c(q+1)=c(q)+binopdf(5,q+5,p)*p;
     b(1,Q)=c(q)+p^6;
     q=q+1;
     c(1)=0;
 end;
  Q=Q+1;
end

a=[0.001:0.001:1];
plot(a,b)
title('Probability to win a game');
xlabel('P for server');
ylabel('Probability to win a game');
```

Figure 5: Matlab code

Is it useful? [?] From the model, we can infer that the fairness of tennis depends on the capability of each player. The stronger the server is, the more advantage he will possess. Especially, when $P = 1/2$, the game is absolutely fair. Both receiver and server will have equal chance to win each game, thus the same probability to win the match. On the other hand, if the receiver is strong enough that the chance for him to win each point exceeds 1/2, then the game will be favorable to him. However, there are several assumptions that needs to be verified to make sure the conclusion is correct. One assumption is that winning rate in each game is constant. As we know, the result of pervious game may effect the psychological state of the players and make a difference on the next game. So the first server may have more advantage if he wins his serving game. With the historical test result, we can make the final conclusion.

References

[1] E. Bender. *An Introduction to Mathematical Modeling.* Dover Publications, 1978. 6