Laboratorio Economia e Finanza

Marco Delogu

Poisson Prediction Model

Recap

- What we did so far?
  - We studied the key properties of the Poisson Distribution
  - We derived the distribution along along with its expected value and variance
  - Now is time to start working with it with a real example
  - Digression on Interest Rates taking advantage of our recap on Series Expansion
- This lecture builds on the previous ones to develop a forecasting model of football match results

Subject of this lecture

- Get used to one of the most useful discrete statistical distribution
  1. Learn to use Poisson distribution to model real phenomena
     - 1.1 Football matches are characterized by goals.. and goals comes in discrete amount.. it makes sense to take advantage of the Poisson distribution to try to predict outcomes of the football matches
     - 1.2 Before we need some data... luckily it is freely available on internet.. we can take advantage of a simple data scraping procedure available through the Excel Ribbon
  2. Exploit some simple excel function
     - 2.1 We will take advantage of some functions covered in the previous lectures like Vlookup,)
     - 2.2 we are going to determine the probabilities of a match outcome ..Juventus Internazionale..god opportunity to see how well the model works. If you like betting you can use the model to guide your bets
  3. Furthermore we exploit VBA to improve our excel sheet
     - 3.1 Specifically we write a macro and define our first loop

Data (1)

- Follow the link to get the data
- to get the data you must select wide in the relevant table
- Just do a traditional Copy and Paste to import the data in Excel
- Remind that the excel file PoissonPrediction model has alredy the data saved

Data (2)

1. Here we make a very a digression to learn on how “to scrap” data from internet through excel
2. Go to Excel and Select
   - 2.1 Data
   - 2.2 From the Ribbon Select “From Web”
   - 2.3 Here select Table 0
   - 2.4 You have easily get the data from internet... Looks like a nice feature, especially for the ones that are going to look for some data for the the thesis
3. Notice that we easily got the data that we wanted... the world is full of data available, nowadays the problem is in learning to import it
4. If we will not go short about time we may cover this topic more deeply

Data (3)

- As you may know football matches can end up with three possible results
  - Home win Away win Draw
- A peculiar thing of football (shared with many sports) is that playing in the home field gives a substantial advantage to the home team
- Our prediction model takes into account the *homeadvantage*, look at the excel sheet "Data cleaned" on the Excel workbook PoissonPrediction.xlsm
- Notice that the columns
  - L and M report for each Serie A team the amount of goals scored at home (L) and the amount of goals conceded at home conceded (M)
  - S and R report for each Serie A team the amount of goals scored away (S) and the amount of goals conceded away (R)
- These are the key statistics needed to setup our model

Key Stats

In order to feed our model (which we will describe shortly) we need to determine

For each team

- Average goals for at home =l3/h3
- Average goals conceded home=m3/h3
- Average goals for away= r3/n3
- Average goals conceded away=s3/n3

For the league globally

- Average goals for at home (SUM GOALS for HOME/ SUM MATCH HOME) , C25
- Average goals conceded home (SUM GOALS conceded HOME/ SUM MATCH HOME) d25
- Average goals for away (SUM GOALS for away/ SUM MATCH away) , C26
- Average goals conceded away (SUM GOALS conceded Away/ SUM MATCH away) , d26

Model

- **We already agreed that a possible model describing the results of a football match is to rely on the Poisson distribution..Notice that goals come always in discrete amount (0,1,2...)!!**
- What is the key and *only* parameter of the Poisson model?... we already know this from the previous lecture is the mean.... To make any prediction we first need to determine an estimate of the parameter. Once we know this value we can make predictions about matches'outcomes. To do so we make the *assumption* that the phenomena follows a Poisson Distribution
- More clearly
  - Specifically we need to model the goals'distributions. we can use the historical information that we have already gathered to try to predict the future..
  - Let's start with the goals averages about goals..
  - However how many goals a team makes depends also on the defensive strength of the opponent
- Hence the parameter describing the Poisson distribution of goals depends both on the attacking and defensive strenght

Relative attacking- and defensive-strength

- For each team we determine both home- and away-attacking strength
  - we divide the average goals made at home and away by each team with the respective league average. You will obtain a value either lower or larger than 1. When this value is larger than one it means that the teams scores the value less 1 percentage more/less than the league average. For instance for Juventus the home attacking strength is given by $V4/\$C\$25$, away attacking strength is given by $X4/\$C\$26$
- Likewise we determine the home- and away- defensive strength for all other teams
  - we divide the average goals conceded at home by each team with the respective league average. In such a case, when 1 less this value is positive means that the team concede x percentage points more goals than the league average

Goal Expectancy

- Our model takes into account the fact that Juventus is likely to score more goals against Spal than against Inter...
  - Inter defense is stronger than Spal defense...
  - Also Spal attack is weaker than Inter one... hence we may expect Juventus to concede less against Spal than Inter
  - $GoalExpHome = Hattstrengh * Awaydefstrenght * avgoalshome$
- Essentially this simple model determine the Poisson mean for the home team in a way that it accounts for
  - the average goals scored in the league
  - the home attacking strength
  - the away defense strength
- In a similar fashion we determine the expected value of the goals scored by the away team
  - $GoalExpAway = Aattckstrenght * Hdefenstrenght * AverageGoalAway$

Poisson Distribution (1)

- By doing so we have determined the "expected mean" of goals produced for both home and away team....
- Goals come in discrete amount.... then we can prediction exploiting the assumption that goals follow a Poisson distribution
  - Let $GoalExpHome = GH_m$ and $GA_m = GoalExpAway$
- We have already gained a profound knowledge of the Poisson distribution... Let GH define the random variable capturing the number of goals scored by the home team. Then

$$P(GH = x) = \frac{\lambda^x e^{-\lambda}}{x!}, \text{ In our case } \lambda = GH_m \text{ then}$$

$$P(GH = x) = \frac{GH_m^x e^{-GH_m}}{x!}, \text{ let } x = 2 \text{ then } P(GH = 2) = \frac{GH_m^2 e^{-GH_m}}{2!}$$

So for instance for Juventus against Internazionale, this value) ≈ 1.56. Then the probability that Juventus scores two goals is equal to:

$$P(GH = 2) = \frac{1.56^2 e^{-1.56}}{2!} = 0.25 = 25\%$$

Poisson Distribution (2)

In a similar fashion we can compute the probabilities for the away team. Let GA define the random variable capturing the number of goals scored by the away team. Then

$$P\left(GA = x\right) = \frac{\lambda^x e^{-\lambda}}{x!}, \text{ In our case } \lambda = GA_m \text{ then}$$

$$P\left(GH = x\right) = \frac{GH_m^x e^{-GH_m}}{x!}, \text{ let } x = 2 \text{ then } P\left(GH = 2\right) = \frac{GH_m^2 e^{-GH_m}}{2!}$$

So for instance for Juventus against Internazionale, this value is equal to ≈ 1.30. Then the probability that Internazionale scores two goals at Juventus Stadium, according to our simple model is equal to:

$$P\left(GA = 2\right) = \frac{1.30^2 e^{-1.30}}{2!} == 22\%$$

Building a Simple Prediction Model

- Consider the Excel sheet *Forecast* of the Excel file PoissonPrediction.xlsm . We want to exploit our model to produce a table that tells us the probability of each possible result. Notice
  - A football match can end up 0-0, 1-0.... 0-1... we are going to consider all "likely" results. We consider till each team makes less than 10 goals... Hence the final result that we consider is 10-10
- We already determined the probability that both Juventus and Internazionale makes two goals..It is now easy to determine the probability that the match end up 2-2...
- Key assumption: independence (remind the previous lecture). Then
- The probability that the match ends 2-2 is equal to:
  - $P(GH = 2) \wedge P(GA = 2)$ ... INDEPENDENCE (it is not absurd to assume that the two variables are indipendent)... $P(GH = 2) * P(GA = 2)$
  - Hence $\frac{GH_m^2 e^{-GA_m}}{2!} * \frac{GA_m^2 e^{-GA_m}}{2!}$ ....we can then easily calculate this probability: $0.25 * 0.22 = 5.6\%$
- We can extend this reasoning for all the possible results (look at the excel sheet Forecast!)

Details on the Excel Sheet

- Using the Vlookup function we can retrieve the data on attacking and defensive strength for both Juventus and Internazionale (remind to pay attention to differntiate between home and away team(look at the excel sheet). Hence we can easily determine the result for any other couple
- Afterwards we build a simple table to calculate the probability for each possible result... some examples
  - $P(GH = 2) * P(GA = 0)$, which is the probability that the match ends up 2-0 for Juventus. Our model predicts that this event has a probability equal to 6.9%
  - $P(GH = 0) * P(GA = 2)$, which is the probability that the match ends up 0-2 for Internazionale. Our model predicts that this event has a probability equal to 4.8%
  - $P(GH = 0) * P(GA = 0)$ which is the probability that the match ends up 0-0 . Our model predicts that this event has a probability equal to 5.7%
  - You can notice that the most likely result predicted by our model is a draw with 1 gol per team. This event is predicted with a probability almost equal to 11%

Probabilities

- Summing up all probabilities for the events where Juventus is predicted to win the match we obtain $P(Juventus = win)$=43.4%
- Likewise
  - $P\left(Internazionale = win\right) = 31\%$
  - $P(draw) = 49\%$
- Now we have several information. Exploiting the information in the table try to calculate
  - Probability that the amount of goal scored is lower than 2, 3, 1, 4,5 (UNDER)
  - Probability that the amount of goal scored is larger than 1,2,3,4,5 (OVER)

Betting

- Assume that you are risk neutral and that you deeply trust your model.. Then
- You bet 1 euro to the victory for Milan if and only you expect to earn a value larger than 2.29 euro... why
- $1 = p * x$... We have already determined the probability that Juventus win the match and it is equal to
- $1 = 0.48 * x \rightarrow x = 2.29$
- You shall bet 1 euro to the victory of Juventus if and only if you receive, whether this outcome success, a value larger or equal to 2.29 dollars
- Do the same computation for Internazionale victory and draw.... and also for the UNDER and OVER probabilities that you previously determined

Extension

- This is a very simple way to model football results. However it is not so different from the model used by professionals
  - if you are interested look for zero inflation models. It seems that Poisson underestimates the zeroes..
  - We use all data available... it makes sense to use information of several months ago?... would be better to use a decay factor?
  - We don't use information concerning the state of the team
- However
  - very simple model which does result not so different to the ones that you will get in any betting website
  - good application of the Poisson distribution
- Put the attention on the modelization of the mean

- When we write some code there are some instructions that we want to be repeated for a previously defined amount of time..
- The basic way to repeat certain instruction several time in Visual Basic is trough the **for, next construct.**
  - For i=1 to 10
    - instructions
  - next i
- The above code start setting the variable i=1.
- Afterwards the instructions are exectued
- Then i becomes equal to i+1
- The instruction are executed till i os not larger than 10

Example

- Using the “basic” features of Excel we built a table which provides the probability of the results of Milan Bologna
- We started from 0-0 considering all intermediate results till 10-10
- Subsequently we computed the probability of each possible results as the probability that Milan scored x gol time the probability that Bologna scored 0 gol
- We saw that it is not difficult to build this table in Excel... Just a bit of attention with the *dollars*
- Now we build the same table with Visual Basic. This is the right training to learn something more about loops.. .

VBA code (1)

- File PoissonPrediction contains the VBA routine loop1. Below I detail you the code

```
Sub loop1()

foglio = forecast2

startcol = 8

startrow = 3

maxgol = 10

'definiamo i numeri di gol nelle colonne, saranno gli input delle funzioni Poisson

For i = 0 To maxgol

Worksheets(foglio).Cells(startrow + i, startcol - 1) = i

Worksheets(foglio).Cells(startrow - 1, startcol + i) = i

Next i
```

The second part of the code is provided in the next slide. Notice that all the text after ' is considered as a comment in VBA

VBA code (2)

```
meanhome = Worksheets(foglio).Cells(2, 2)

meanaway = Worksheets(foglio).Cells(2, 4)

For i = 0 To maxgol

For j = 0 To maxgol

Worksheets(foglio).Cells(startrow + j, startcol + i) =
(WorksheetFunction.Poisson(Worksheets(foglio).Cells(startrow + j, startcol - 1), meanhome,
False)) * _ (WorksheetFunction.Poisson(Worksheets(foglio).Cells(startrow - 1, startcol + i),
meanaway, False))

Next j

Next i

End Sub
```

- Notice that _ is used in VBA to go to the next line!

Explanation (1)

foglio = forecast2

startcol = 8

startrow = 3

maxgol = 10

Here I am setting the name of the worksheet on which I want to work, the other variables define the starting column (D=8) and the starting row (C=3). maxgol instead defines the maximum amount of goal scored that I want to consider.

Explanation (2) First Loop

```
For i = 0 To maxgol

Worksheets(foglio).Cells(startrow + i, startcol - 1) = i

Worksheets(foglio).Cells(startrow - 1, startcol + i) = i

Next i
```

This is our first loop. Notice that we are starting from 0 till 10. Then I am going to move across rows and across columns to report the number of goal scored by both home and away team. To better understand how the code works consider the case whe $i = 3$.

```
Worksheets(foglio).Cells(3+ 3, 8 - 1) = 3 '(which means cell (6,G)=3 )

Worksheets(foglio).Cells(3-1, 8 + 3) = i ('which means cell(2,k)=3)

Next i
```

Explanation 2nd Loop (nested)

- Before entering in the loop we save in two variables the Poisson mean of home and away goals (look at meanhome and mean away)
- Then I have a nested loop, it means in this case that first I do the row and then I move to the columns

```
For i = 0 To maxgol

For j = 0 To maxgol

Worksheets(foglio).Cells(startrow + j, startcol + i) =
(WorksheetFunction.Poisson(Worksheets(foglio).Cells(startrow + j, startcol - 1), meanhome,
False)) * _ (WorksheetFunction.Poisson(Worksheets(foglio).Cells(startrow - 1, startcol + i),
meanaway, False))

Next j

Next i

End Sub
```

Notice that we can use the Poisson function inside the VBA code as a method of the *Worksheetfunction* object. The syntax is the same of the one used in Excel. Further information about this method are available at the following link

Exercise (1)

- Consider the excel file Exercise. Consider the excel sheet exercise
- Write a VBA macro which reports in cell (1,B) of the sheet exercise the sum of the values contained in column A
- Change the code to sum the values of column A only if the value in the cell is equal to 2
- Hints:
  - Look at lecture 1 the last line of the code... vba code to determine the last row with data...
  - Look at lecture2–3... we learnt how to use the if then else construct in Visual Basic

Exercise (2)

- Let assume that you trust our Poisson Prediction Model
- Collect the most recent data on 'Serie A home and away goals
  - Hint: Consider the last Match played by Milan. Assume that you know for sure that Milan scored one goal. What is the probability that Milan actually won the match?
- You did not see the match. However a friend of yours calls you and tells that Milan won..What is the most likely amount of goals that Milan scored? Trusting our simple model...

RECAP

- We saw how to do very simple web scraping with Excel
- We exploited our knowledge of the Poisson distribution to build a prediction model for Football matches
- We applied our model to the Juventus Internazionale Match
- We determined the odds and built using the built in function an excel sheet predicition model
- We also wrote our first loop in Visual Basic.
- We wrote a nested loop to “automatize” the table which we previously built manually inside excel