Laboratorio Economia e Finanza

Marco Delogu

Poisson Distribution: Definition and Properties. Interest Rates, Growth Rates

- What we did so far?
  - Review some basic feature of Excel
  - Learnt some of the key excel functions (Among the others vlookup and Match-Index)
  - Introduction to Visual Basic and derivatives
- These slides mostly cover theoretical concepts. However application in Excel are part of the lecture as well
  - Recap of some key concepts: Expected Value and Variance
  - Derivation of the Poisson Distribution and its key properties.
  - Digression on Interest Rates and growth rates
  - Solution of exercises through Excel
  - Those lectures pave the way to the next ones where we will exploit the properties of the Poisson Distribution to build a model to forecast the result of football matches. We write some loop structures in VBA
  - Part of this lecture is adapted from the book "Statistics", Bourghes 1994

Poisson: Key features

- The Poisson distribution takes its name from the french mathematician Siméon Denis Poisson.
- It is a discrete probability distribution. It determines the probability that a given number of events occurrs in a fixed interval of time or space. Events must occur at a constant rate and independently.
- It describes nicely many real phenomena, among them:
  - number of calls received
  - number of vehicles passing on a road in a certain interval of time
  - flaws in a given length of material
- It is useful when for the phenomena in hand the number of possible occurrences is in theory, unlimited
- We start with an *experiment*...and from this experiment we *obtain* the distribution

Empirical Example (1)

1. You don't know what to do and for a week you go to the same road at the same time and count the number of car passing bay..
2. Imagine that you go in the same road, at the same time, for several days and you count the number of vehicles passing by in the road in one minute.
3. Specifically, let assume that in order to make this experiment you went to the same road between 17:50 till 18:00 for 100 days in a row. You don't forget to take note of your observations and eventually you end up with a table $\rightarrow$

Empirical Example (2)

Number of cars per minute(x)	Frequency (f)
0	7
1	34
2	84
3	140
4	176
5	176
6	146
7	104
8	65
9	36
10	18
11	8
12	4
13	1
14	1
$\geq$15	0

Recap (1) Expected Value

- Expected value of a discrete random variable:

$$E\left[X\right] = \sum x_i f_x\left(x\right)$$

where X is discrete with mass points x_1, x_2, $x_3 \ldots . x_j$

- Import the Data in Excel and compute the mean: you should obtain

$$\bar{x} = \frac{0*7+1*34+\ldots .14*1}{1000} \approx 4.997$$

- Another key concept that you already encountered in your studies is the Variance. Let X be a discrete random variable and let μ_x be $E\left(X\right)$ ¡. The variance of X, denoted by σ_x^2 is defined by

$$VAR\left[X\right] = \sum \left(x_j - \mu_x\right)^2 f_x\left(x_j\right).$$

- Remind that if $(g\left(.\right))$ is a function the expected value of the function g of the random variable X is

$$E\left[g\left(x\right)\right] = \sum g\left(x_j\right) f_X\left(x_j\right)$$

- Then if we define $g\left(x\right) = \left(\left(x_j - \mu_x\right)^2\right)$ we have $E\left[g\left(x\right)\right] = E\left[\left(X - \mu_x\right)^2\right] = var\left(X\right)$

Recap (2):Variance

Theorem: if X is a random variable
$var\left(X\right)=E\left[\left(X-\mu_x\right)^2\right]=E\left[X^2\right]-\left(E\left[X\right]\right)^2$ provided $E\left[X^2\right]$ exists

Proof:
By definition
$var\left[X\right]=E\left[\left(X-E\left(X\right)\right)^2\right]$ then
$E\left[\left(X-E\left(X\right)\right)^2\right]=E\left[X^2-2XE\left[X\right]+E\left[X\right]^2\right]$
The E operator is additive. Also the expected value of a constant is a constant.
$Var\left[X\right]=E\left[X^2\right]-E\left[2XE\left[X\right]\right]+E\left[E\left[X\right]^2\right]$
Remind that $E\left(X\right)$ is a constant, therefore the second term becomes equal to $-2\left(E\left[X\right]\right)^2$. The third term, is a constant and equal to:
$\left(E\left[X\right]\right)^2$.
Then

$$VAR\left[X\right]=E\left[X^2\right]-\left(E\left[X\right]\right)^2,$$

Two formulas to determine the Variance. See the Excel workbook *Lecture4*, sheet *exampledata*

Back to the experiment

- The variance of the data is equal to

$$\sigma^2 = \frac{0^2 * 7 + 1^2 * 34 + \ldots\ldots . 14^2 * 1}{1000} - 4.997^2 \ldots$$

- see the excel file Lecture4.xlsx for the calculation
- our data comes from an experiment.. the **mean is almost identical to ithe variance.**.. Soon we see that this is a key property of the Poisson distribution
- In few slides we will see that the Poisson distribution is defined by only one parameter equal to both the mean and the variance
- Let's look with attention to the data
- focusing at the frequencies...

$$\frac{34}{7} \approx \frac{5}{1}, \ \frac{84}{34} \approx \frac{5}{2}\ , \ \frac{140}{84} \approx \frac{5}{3}, \frac{176}{140} \approx \frac{5}{4}, \frac{176}{176} \approx \frac{5}{5}, \frac{146}{176} \approx \frac{5}{6}, \ \frac{104}{146} \approx \frac{5}{7},$$
$$\frac{65}{104} \approx \frac{5}{8}, \frac{36}{65} \approx \frac{5}{9}, \ \frac{18}{36} \approx \frac{5}{10}$$

- Those ratio seem to related among themselves. .
- Take note of the empirical probability of not seeing any car

$$P\left(X = 0\right) = \frac{7}{1000} = 0.007$$

uniss
UNIVERSITÀ DEGLI STUDI DI SASSARI

Empirical Probabilities

- Let's exploit the ratio that seem to define the ratio among two different and consecutives outcomes:

$$P(X=1) = \frac{5}{1} P(X=0)$$

- then...

$$P(X=2) = \frac{5}{2} P(X=1) = \frac{5^2}{2*1} P(X=0)$$

$$P(X=3) = \frac{5}{3} P(X=2) = \frac{5^3}{3*2*1} P(X=0)$$

$$P(X=4) = \frac{5^4}{4*3*2*1} P(X=0)$$

- We *guess* that the probability distribution can be written as follows:

$$P(X=n) = \frac{5^n}{n*(n-1)....2*1} * P(X=0) = \frac{5^n}{n!} P(X=0)$$

- We are not yet done, but we are getting closer. Before moving forward we need another recap: Taylor and MacLaurin series.)

- A Taylor series is a series expansion of a function about a point. A one dimensional Taylor series an expansion of a real function $f(x)$ about a point $x = a$ is given by

$$f(x) = f(a) + f'(a)(x-a) + \frac{f''(a)}{2!}(x-a)^2 + \frac{f'''(a)}{3!}(x-a)^3 + ... + \frac{f^n(a)}{n!}(x-a)^n$$

- Where f' represents the first derivative, f'' the second derivative... and so on... more compactly..

$$f(x) = \sum_{n=0}^{\infty} \frac{f^{(n)}(a)}{n!}(x-a)^n$$

- If $a = 0$ the expansion is known as a Maclaurin series:

$$f(x) = \sum_{n=0}^{\infty} \frac{f^{(n)}(0)}{n!}(x-0)^n$$

- We know compute the McLaurins expansion of e^x. You shortly understand why

Maclaurin expansion of e^x

- Consider e^x, then simply exploiting the definition we have that

$$e^x = e^0 + e^0\,(x-0) + \frac{e^0}{2!} * (x-0)^2 \,........ + \frac{e^0}{n!}\,(x)^n$$

- remind...$e^0 = 1$

$$e^x = 1 + x + \frac{1}{2!}\,(x)^2 + ... + \frac{1}{n!}\,(x)^n$$

- This expansion of e^x will help us to derive the Poisson Probability distribution

Back to Poisson

- We have already showed that

$$P(X=n)=\frac{5^n}{n*(n-1)....2*1}*P(X=0)=\frac{5^n}{n!}P(X=0)$$

- Probabilities must satisfy a key property... they must sum to 1!!!!
- Define

$$P(X=0)=p$$

- then

$$1=p+5p+\frac{5^2p}{2!}+\frac{5^3p}{3!}+\frac{5^4p}{4!}+......$$

Back to Poisson (2)

- things are getting interesting... in fact collecting the RHS for p

$$1 = p\left(1 + 5 + \frac{5^2}{2!} + \frac{5^3}{3!} + \frac{5^4}{4!} + ...\right) + ...$$

- the RHS now looks similar to something.. to the McLaurin expansion of e^x with $x = 5$ series

$$\left(1 + 5 + \frac{5^2}{2!} + \frac{5^3}{3!} + \frac{5^4}{4!} + ...\right) = e^5.....$$

- then solving for p we obtain

$$p = P\left(0\right) = e^{-5}.....$$

- Then in our example where the mean is equal to 5 we have..

When a Bank claims that the interest rate is 10% the statement is still ambiguous... In fact we need to know how the interest is compounded

Compounding Frequency	\$100 dollars at the end of the year
Annually ($m = 1$)	110
Semiannually ($m = 2$)	110.25
Quarterly ($m = 4$)	110.38
Monthly ($m = 12$)	110.47
Weekly ($m = 52$)	110.51
Daily ($m = 365$)	110.52

Compounding frequency: defines the units in which an interest rate is measured. Also, we can convert one compounding frequency into an equivalent rate with a different compounding frequency. Example invest $\$100 dollars$

- Yearly: $\$100 * 1.1 = \$110$
- Semiannual: $\$100 * 1.05 * 1.05 = \$110.25$
- Quarterly: $\$100 * 1.05^4 = \$110.25$
- and so on..

Notice that from the previous table we see that 10.25% with annual compounding is equivalent to 10% with semiannual compounding. How to interpret to the difference between one compounding frequency to another.. It is like we are using a different unit of measurement, like between kilometers and miles

Compounding (2)

Let assume that an amount A is invested for n years at an interest rate R per annum. If the rate is compounded once per annum the value of the investment is equal to:

$$A\left(1+R\right)^{n}$$

Otherwise if the rate is compounded m times per annum the terminal value of the investement is simply:

$$A\left(1+\frac{R}{m}\right)^{nm}$$

Quantitative finance is done mostly with continous compounding.. Namely let m tends to infinite (namely the interest rate is compounded instantaneously). We are going to show that in such a case the value of our investment is equal to:

$$Ae^{Rn}$$

Using the same hypothesis of our table we have, $R = 1$, that the value of our investment is equal to:

$$100e^{0.1} = 110.52$$

Notice that, (to two decimal places), this is the same value with daily computing

Conversion (1)

Let define with R_c the interest rate with continous compounding... we can easily determine the equivalent rate with compounding m times per annum. Notice that:

$$Ae^{R_c n} = A\left(1 + \frac{R_m}{m}\right)^{mn}$$

$$e^{R_c n} = \left(1 + \frac{R_m}{m}\right)^{mn}$$

Take logs in both sides

$$R_c n = mn\left(ln\left(1 + \frac{R_m}{m}\right)\right)$$

Conversion (2)

Solve for R_m

$$\frac{R_c}{m} = ln\left(1 + \frac{R_m}{m}\right)$$

then

$$e^{\frac{R_c}{m}} = \left(1 + \frac{R_m}{m}\right) \rightarrow m\left(e^{\frac{R_c}{m}} - 1\right) = R_m$$

Let assume that the interst rate is quoted as 10% per annum with semiannual compounding. What is equivalent rate with continous compounding:

$$R_c = mln\left(1 + \frac{R_m}{m}\right)$$

$$R_c = 2 * ln\left(1 + \frac{0.1}{2}\right) = 0.09758 \text{ in \% terms } 9.758\%$$

Exercise: Write in VBA a function that takes as input m and R and gives you the equivalent rate with continuous compounding

Conversion Ex:(2)

Let assume that I borrowed money at a rate equal to 8% with continuous compounding. Interest rate is paid quarterly. How much do I pay each 3 months? First we determine the equivalent rate with quarterly compounding:

$$R_m = m\left(e^{\frac{R_c}{m}} - 1\right)$$

$$R_m = 4\left(e^{\frac{0.08}{4}} - 1\right) = 0.0808$$

or 8.08% per annum. So we pay 0.0808*1000 as interests, which is equal to 80.8 dollars. Hence, each quarter I pay 20.2 0 dollars. **Exercise**: write in VBA a Function that computes the equivalent rate once you know the rate per annum with continuous compounding.

So far we have just assumed that

$$A = \lim_{n \to \infty} \left(1 + \frac{R}{n}\right)^{nt}$$

is equivalent to

$$A = e^{rt}$$

Let's show that the equivalence holds rigourously. We take advantage of our recap about the Taylor Expansion:
Taking logs of the first equation we obtain

$$ln\left(A\right) = ln\left(\lim_{n \to \infty} \left(1 + \frac{R}{n}\right)^{nt}\right)$$

Continity allows to swap the limit

$$ln\left(A\right) = \lim_{n \to \infty} ln\left(1 + \frac{R}{n}\right)^{nt}$$

then, exploiting the property of the natural logarithm

$$ln\left(A\right) = \lim_{n \to \infty} \left(nt\right) * ln\left(1 + \frac{R}{n}\right)$$

Take the Taylor Series of the RHS of the above equation at $R = 0$.

1. $f\left(R=0\right)=ln\left(1+\frac{0}{n}\right)=0$
2. $f'\left(R=0\right)*\left(R-0\right)=\left(\frac{1}{1+\frac{R}{n}}\right)\frac{1}{n}\left(R-0\right)=$
$\left(\frac{1}{1+\frac{R}{n}}\right)\frac{1}{n}\left(R-0\right)=\frac{R}{n}$
3. $f''\left(R=0\right)*\left(R-0\right)\frac{1}{2!}=\left(-\left(\frac{1}{n+R}\right)^2\left(R-0\right)^2\frac{1}{2*1}\right)=-\frac{R^2}{2n^2}$
4. $f'''\left(R=0\right)*\left(R-0\right)\frac{1}{3!}=\left(2*\left(\frac{1}{n+R}\right)^3\left(R-0\right)^3\frac{1}{3*2}\right)=\frac{R^3}{3n^3}$
5. $f''''\left(R=0\right)*\left(R-0\right)\frac{1}{4!}=$
$\left(-2*3*\left(\frac{1}{n+R}\right)^4\left(R-0\right)^4\frac{1}{4*3*2}\right)=-\frac{R^4}{4n^4}$

and so on for the remaining terms of the expansion

Derivation (4)

$$ln\left(A\right) = \lim_{n\to\infty}\left(nt\right)\left(\frac{R}{n} - \frac{R^2}{2n^2} + \frac{R^3}{3n^3} - \frac{R^4}{4n^4}\right)$$

Multiply for n and take advantage of the linearity of the limit operator

$$ln\left(A\right) = \left(\lim_{n\to\infty} Rt\right) - \left(\lim_{n\to\infty}\frac{tR^2}{n}\right) + \left(\lim_{n\to\infty}\frac{tR^3}{n^2}\right) - \left(\lim_{n\to\infty}\frac{tR^4}{n^3}\right)....$$

Notice when $n$approaches to infinite all terms of the RHS except the first one tends to zero. Hence

$$ln\left(A\right) = \left(\lim_{n\to\infty} Rt\right) = Rt$$

Take the exponential in both sides. Then

$$A = e^{Rt}$$

The simplest differential equation

- Assume that a variable is growing at some constant rate g. Hence:

$$\frac{\dot{x}}{x} = g$$

- What does it imply about the level of x?
- Notice that

$\frac{dln(x(t))}{d_t} = g$.... to “undo” derivatives we do integrals...$dln\left(x\left(t\right)\right) = gdt$

- Integrate

$$\int dln\left(x_t\right) d_t = \int gdt \implies ln\left(x\left(t\right)\right) = gt + C$$

Initial Condition

- where C is a constant of integration
- take the exponential

$$x(t) = e^{gt}e^{C} = e^{gt}\bar{C}$$

- where

$$\bar{C} = e^{C}$$

- Assume that we know the value of $x(0)$ at t_0, namely

$$x(0) = x_0$$

- Hence

$$x(0) = e^{g*0}e^{C} = e^{C} = \bar{C}$$

Poisson: Theoretical Probabilities

x	P(X=x)
0	e^{-5}
1	$5 * e^{-5}$
2	$\frac{5^2}{2!}e^{-5}$
3	$\frac{5^3}{3!}e^{-5}$
4	$\frac{5^4}{4!}e^{-5}$
5	$\frac{5^5}{5!}e^{-5}$
6	$\frac{5^6}{6!}e^{-5}$
7	$\frac{5^7}{7!}e^{-5}$

- The key parameter in fitting a Poisson distribution is the **mean value** (which coincide with the variance for this distribution). We denote it with the Greek letter λ. It tells us the average number of occurrences in a specified period (in our case the average number of cars passing in a minute). In general the density of Poisson distribution (remind that is a discrete probability function) is given by

$$P\left(X=x\right)=\frac{e^{-\lambda}\lambda^{x}}{x!},\left(x=0,1,2,3.....\right)\textbf{and is denoted by}$$
$$X\sim P_{o}\left(\lambda\right)$$

$$\lambda=E\left[X\right]=VAR\left(X\right)$$

- The distribution arises when the events being counted occur:
  - independently
  - randomly in time or space
  - uniformly (that is, the mean number of events in an interval is directly proportional to the length of the Interval)

Poisson Distribution Excel (1)

- From our experiment.. we derived the Poisson distribution with parameter $\lambda = 5$
- In excel using built in the Poisson function we can recover the probability mass distribution for each possible event (1,2 ...3 ...n)
- Notice that there is little difference between the theory distribution and the numbers generated by our data
- This is interesting...
- Take an excel sheet and generate Poisson data with $\lambda = 4$
  - Plot the density distribution
  - Plot the cumulative distribution
- Remind that Poisson is a discrete distribution and... Excel knows it!

Poisson Distribution in Excel (2)

- Let's generate a Poisson Distribution in Excel with $\lambda = 4$.
- You must use the function named... *Poisson*... the arguments of the function are :

$$= Poisson(\mathbf{x,mean,\ cumulative})$$

- where x is the value for which we want to know the probability, *mean*, is our λ the parameter characterizing distribution. Cumulative can take two values, True of False. With True you are going to get the value defining the mass distribution. On the contrary with false you are going to obtain the value of the cumulative distribution
- Notice that Excel accepts real number for the x argument. However the value is truncated! (namely, if you write 3.6) excel considers a 3!. Poisson is a discrete!
- For instance if we want to know the probability $P(X = 4)$ we have to write $POISSON(4, 4, FALSE)$
  - If you put true you obtain the cumulative distribution..remind
  - Look at the EXCEL file Lecture4.xlsm, where we used the function
  - We plot both the density and the cumulative distribution function.
- In the following we are going to solve some exercises to learn the usefulness of the distribution

Exercise (1)

- Consider a random variable X follows a Poisson distribution with mean 3.4. Determine $P(X = 6)$
- **Solution**:
  - We can write the distribution as follows $X \sim P_o\,(3.4)$ then,
  - $P\,(X = 6) = \frac{e^{-\lambda}\lambda^6}{6!} = \frac{e^{-3.4}3.4^6}{6!} = 0.071604409 \approx 0.072$
  - It means that if our phenomena is described by a probability distribution with mean equal to 3.4 the probability of getting a 6 is equal to the 7.2 %
  - Solve the excercise using the Poisson function in Excel, $Poisson(6, 3.4, FALSE)$

Exercise (2)

- The number of industrial injuries per working week in a particular factory is known to follow a Poisson distribution with **mean 0.5**. Determine the probability that in a particular week there will be:
  - less than 2 accidents
  - more than 2 accidents
  - in a three week period there will be no accidents
- ***Solution:***
  - Let define by A, the random variable, the "number of accidents in one week", so $A \sim Po\,(0.5)$
  - **(i)** then

$$P\,(A<2) = P(A=0) + P\,(A=1) = e^{-0.5} + \frac{e^{-0.5}0.5}{1!} = \frac{3}{2}e^{-0.5} \approx 0.9098.$$

Hence the probability that the number of accidents is lower than 2 in a given week is equal to 90%

- **Solution (ii)**
    - Notice that $P\,(A > 2) = 1 - P(A \leq 2)$ (remind that you can determine the probabilities through tables or by using the excel function). However at the exam you will be asked to determine the solution through the formula)
    - Notice that
    - $P(A \leq 2) = 1 - [P(A = 0) + P(A = 1) + P(A = 2)]$
    - then $1 - \left[e^{-0.5} + \frac{e^{-0.5}0.5}{1!} + \frac{e^{-0.5}0.5^2}{2!}\right] = 1 - 1.625e^{-0.5} \approx 0.0144$
    - Hence there is a probability equal to 1.44 % that there will be more than two accidents in this firm in a given week
- **Solution Point (iii)**:
    - Assume that the events are **independent**, than the probability that there will be no accidents in 3 weeks is equal to:
    - $P\,(0 \text{ in } 3 \text{ weeks}) = \left(e^{-0.5}\right) * \left(e^{-0.5}\right) * \left(e^{-0.5}\right) = 0.223$ equal to 22.3%
    - Think how you could have exploited the Excel function to solve the two exercises (look at the file Lecture4.xlsx)

- Theorem **(no formal proof)**.

if $A \sim P_o(a)$ and $B \sim Po(b)$ are independent Poisson random variables then the random variable $C = (A + B) \sim P_0(a + b)$

- Namely summing up Poisson, independent, random variables you obtain another Poisson random variable with *mean equal to the sum of the generating Poisson random variables*
- We don't get into the formal proof, however it is useful to understand why
- Consider
  - $A \sim P_o(2)$ and $B \sim P_o(2)$.
  - Consider the distribution $C = A + B$
- We know that $P(A = 0) = e^{-2}$ and $P(B = 0) = e^{-3}$,
  - then $P(C = 0) = P(A = 0) * P(B = 0) = e^{-5}$
- $P(C = 1) = P(A = 0) * P(B = 1) + P(A = 1) * P(B = 0)$....which is equal to
  - $e^{-2} * 3 * e^{-3} + 2 * e^{-2} * e^{-3} = 5e^{-5}$
- Exercise compute $P(C = 2)$....think about it... are your computation confirming the theorem?

The number of misprints on a page of the Daily Mercury has a Poisson distribution with mean 1.2. Find the probability that the number of errors

1. on page four is 2
2. on page 3 is less than 3
3. on the first ten page totals 5
4. on all forty pages adds up to at least 3

The probability distribution governing the number of errors on one page is $E \sim P_o\,(1.2)$

- **Solution (i)**
  - $P\,(E=2) = \frac{e^{-1.2}(1.2)^2}{2!} \approx 0.217$
  - It means that the probability of finding two misprints at page 2 is equal to 21.7%
- **Solution (ii)**
  - From Excel or Tables, we easily obtain that $P\,(E<3) = P\,(E \leq 2) = 0.8975$
  - We obtain that the probability of finding less than 3 misprints at page 3 is equal to 89.75%

The probability distribution governing the number of errors on one page is $E \sim P_o\,(1.2)$

- **Solution (iii)**
  - Assume independence... then it is quite easy..
  - Let E_{10} be the number of errors. Errors made in one page are independent then... $E_{10} \sim Po\,(12)$,as $E_{10} = E + E + E ... E$,then
  - $E_{10} \sim P_o\,(1.2 + 1.2 + 1.2 .... + 1.2) = P_o\,(12)$
  - then it is quite easy to obtain $P\,(E_{10} = 5) = \frac{e^{-12} 12^5}{5!} \approx 0.0127$
- **Solution (iiii)**
  - Similarly $E_{40} \sim P_o\,(48)$,then
  - $P\,(E_{40} > 3) = 1 - P\,(E_{40} \leq 2) = 1 - 1201 e^{-48} \approx 1$

- Let's prove that is λ, the parameter of the Poisson distribution is actually equal to its expected value.
- We know the formula that defines the expected value
- Let $X \sim P_o(\lambda)$, we know that for this distribution $E[X] = V[X] = \lambda$
- Let's formally show that if a random variable X follows a Poisson distribution then

$$E[X] = \lambda$$

- **Proof**

$$E[X] = \sum_{all\ x} xP(X = x) =$$

$$0e^{-\lambda} + 1 * \left(\lambda e^{-\lambda}\right) + 2 * \left(\frac{\lambda^2 e^{-\lambda}}{2!}\right) + 3 * \left(\frac{\lambda^3 e^{-\lambda}}{3!}\right) \textbf{ recall MacLaurin expansion of } e^x$$

$$\lambda e^{-\lambda}\left(1 + \lambda + \frac{\lambda^2}{2!} + \frac{\lambda^3}{3!} \dots\dots\right) \textbf{ then (remind Maclaurin expansion of } e^\lambda$$

$$E[X] = \lambda e^{\lambda} e^{-\lambda} = \lambda$$

Variance (1)

in the previous slide we showed that the expected variable of a random variable X that follows a Poisson distribution is equal to λ

To show that $Var(X) = \lambda$ we have proceed in a similar fashion. But we must apply one trick

Remind that

$$VAR[X] = E\left[X^2\right] - (E[X])^2$$

We have already showed that if X is Poisson $E[X] = \lambda$. Therefore $(E[X])^2 = \lambda^2$. Hence for our Poisson random variable

$$VAR[X] = E\left[X^2\right] - \lambda^2$$

We determine:

$$E\left[X^2\right] = \sum x_i^2 f_x(x) = \sum_{all\ x} x^2 P(X = x) =$$

$$0^2 e^{-\lambda} + 1^2 * \left(\lambda e^{-\lambda}\right) + 2^2 * \left(\frac{\lambda^2 e^{-\lambda}}{2!}\right) + 3^2 * \left(\frac{\lambda^3 e^{-\lambda}}{3!}\right)$$

Variance (2)

$$0^2 e^{-\lambda} + 1^2 * \left(\lambda e^{-\lambda}\right) + 2^2 * \left(\frac{\lambda^2 e^{-\lambda}}{2!}\right) + 3^2 * \left(\frac{\lambda^3 e^{-\lambda}}{3!}\right)$$

Collecting for $e^{-\lambda}$ we obtain

$$e^{-\lambda}\lambda\left(1^2 + 2\lambda + \frac{3\lambda^2}{2!} + \frac{4\lambda^3}{3!}.....\right)$$

Before going forward we need an intermediate result.. notice that for any a

$$\left(\frac{a}{n!} + \frac{a}{(n-1)!}\right) = \frac{a\left((n-1)!\right) + an!}{n! * (n-1)!} = \frac{a\,(n+1)}{n!}$$

Variance (3)

$$(1+\lambda)\left(1+\lambda+\frac{\lambda^2}{2!}+\frac{\lambda^3}{3!}.....\right)=\left(1+(\lambda+\lambda)+\left(\frac{\lambda^2}{2!}+\lambda^2\right)+\left(\frac{\lambda^3}{3!}+\frac{\lambda^3}{2!}\right)......\right)$$

Exploiting the results that we obtained befowre we can write:

$$\left(1+2\lambda+\left(\frac{3\lambda^2}{2!}\right)+\left(\frac{4\lambda^3}{3!}\right)......\right)=(1+\lambda)\left(1+\lambda+\frac{\lambda^2}{2!}+\frac{\lambda^3}{3!}.....\right)$$

But we already know that

$$\left(1+\lambda+\frac{\lambda^2}{2!}+\frac{\lambda^3}{3!}.....\right)=e^{\lambda}$$

Hence

$$E\left[X^2\right]=\lambda e^{-\lambda}e^{\lambda}\left(1+\lambda\right)$$

$$VAR\left[X\right]=E\left[X^2\right]-\lambda^2=\lambda\left(1+\lambda\right)-\lambda^2=\lambda$$

$$VAR\left[X\right]=\lambda$$

Recap

- We heuristically derived the Poisson distribution
- We reviewed some key concepts:
  - Expected value, Variance, Taylor and Maclaurin series
- We learnt how to recover the property of some real phenomena exploiting key properties of the distribution
- We formally derived the key property of the Poisson distribution, namely that the mean is equal to the variance

Next

- Exploit the theoretical insights of this lecturer to build a simple football prediction model
- Why:
  - Interesting per se... and much more if you like betting
  - It allows us to understand how to apply theoretical concepts to real life problems
  - We take advantage of VBA features to automatize our Excel Workbook