Discrete Distributions

John S Butler (TU Dublin)

Introduction

Binomial Distribution

A Binomial Distribution is used to describe the probability distribution of n experiments that have two possible outcomes "success" or "failure". The probability of "success" is p, the probability of "failure" is $q = 1 - p$. This gives the general defintion of the distribution as:

$$\Pr(k) = \left(\begin{array}{c} n \\ k \end{array} \right) p^k q^{n-k}, \quad k = 0, 1, 2, ... n,$$

where k is the number of "successes". If the data can be described using a Binomial Distribution then the expected outcome, $E[k]$ of the distribution is given by:

$$E[k] = np,$$

and the variance of the distribution $Var[k]$, is

$$Var[k] = npq.$$

Example

The example we shall use to illustrate the discrete distibutions is the New Zealand vs Ireland World Cup Rugby Quarter Final. Let's assume the probability of "success", Ireland beating New Zealand is

$$p = 0.15,$$

the probability of failure is

$$q = 1 - p = 1 - 0.15 = 0.85.$$

Now imagine that Ireland have to play New Zealand 10 times, this gives the Binomial Distribution

$$\Pr(k) = \left(\begin{array}{c} 10 \\ k \end{array} \right) (0.15)^k (0.85)^{10-k}, \quad k = 0, 1, 2, ... 10,$$

where k is the number of times Ireland beats New Zealand. As it is a Binomial Distribution we can state that the expected number of times Ireland will win is

$$E[k] = 10 \times 0.15 = 1.5,$$

ie Ireland is only expected to win once, the variance of the distribution is

$$Var[k] = 10 \times 0.15 \times 0.85 = 1.35.$$

The plot below shows the Binomial Distribution for the number of times Ireland will beat New Zealand out of 10 games.

```
Games<-0:10 # Number of Games
Pr_ireland_win<-0.15 # Probability of Success
Pr_win<-c(dbinom(0:10,10,Pr_ireland_win)) # Binomial Probability

E_X<- 10*Pr_ireland_win # Expected Value
```

```
Var_X<-10*Pr_ireland_win*(1-Pr_ireland_win) # Variance

df <- data.frame(Games,Pr_win)
## Plot
binomial.p_dist<-ggplot(df, aes(x=Games,y=Pr_win)) + geom_col(width = 0.1)+
  xlab("Number of Irish Wins  out of 10 games")+
  ylab("Pr of Wins")+ggtitle("Binomial Distribution")+
  scale_x_continuous(breaks=0:10)+
  geom_vline(xintercept = E_X, color = "blue", size=1)+
  geom_text(aes(x=E_X, label="E[k]", y=max(Pr_win)),vjust=2,
            colour="blue",  text=element_text(size=18))

binomial.p_dist
```
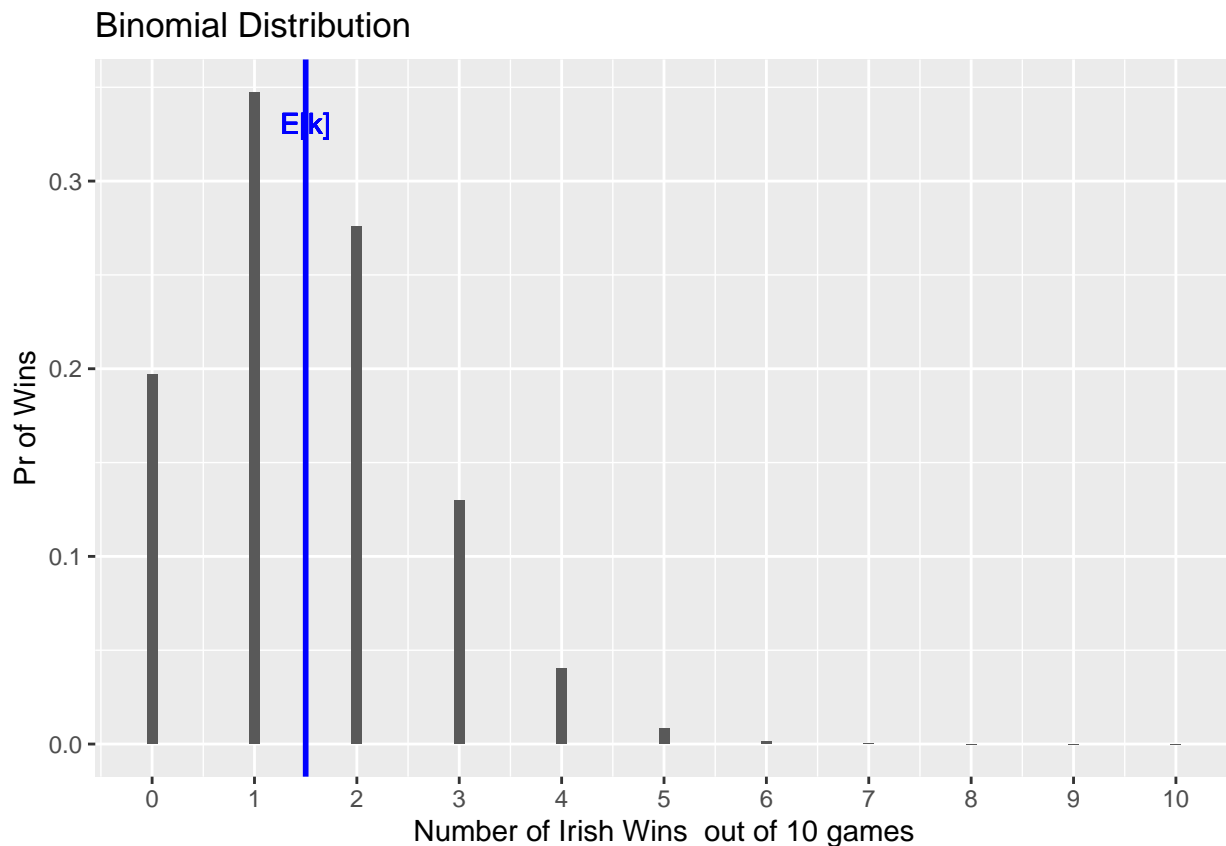

Geometric Distribution

A Geometric distribution is used to describe the probability distribution if you do an experiment until you succeed, the experiment has two possible outcomes "success" or "failure". The probability of "success" is p, the probability of "failure" is $q = 1 - p$. This gives the general defintion of the distribution as:

$$\Pr(k) = q^{(k-1)}p, \quad k = 1, 2, ...$$

with the expected outcome of,

$$E[k] = \frac{1}{p},$$

and variacne of

$$Var[k] = \frac{q}{p^2}.$$

2

Example

The example we shall use to illustrate the Geometric distibution is the New Zealand vs Ireland World Cup Rugby Quarter Final. Let's assume the probability of "success", Ireland beating New Zealand is

$$p = 0.15,$$

the probability of failure is

$$q = 1 - p = 1 - 0.15 = 0.85.$$

Now imagine that Ireland have to play New Zealand until Ireland wins this gives the Geometric Distribution

$$\Pr(k) = (0.85)^{k-1}0.15, \quad k = 0, 1, 2, \cdots$$

where k is the number of games played until Ireland beats New Zealand. As it is a Geometic Distribution we can state that the expected number of games played until Ireland will win is

$$E[k] = \frac{1}{0.15} = 6.6667,$$

ie it will take this many games on average for Ireland to win once, the variance of the distribution is

$$Var[k] = \frac{0.85}{0.15^2}.$$

The plot below shows the Geometric Distribution for the number of times Ireland needs to play until they beat New Zealand, the x-axis only show 0 to 20 possible games but do note the Geometric Distribution can calculate the probability of infinite games until Ireland wins.

```
Games<-0:20 # Number of Games
Pr_ireland_win<-0.15 # Probability of a win
E_X<-1/Pr_ireland_win # Expected Value
Var_X<-(1-Pr_ireland_win)/(Pr_ireland_win**2) # Variance
Pr_win<-c(dgeom(Games,Pr_ireland_win))

df <- data.frame(Games,Pr_win)
# Plot
Geometric.p_dist<-ggplot(df, aes(x=Games,y=Pr_win)) + geom_col(width = 0.1)+
  xlab("Number of Games until Ireland Win")+ylab("Pr(Games until Ireland Win)")+
  ggtitle("Geometric  Distribution")+
  scale_x_continuous(breaks=Games)+
  geom_vline(xintercept = E_X,  color = "blue", size=1)+
  geom_text(aes(x=E_X, label="E[k]", y=max(Pr_win)),vjust=2,
            colour="blue", text=element_text(size=18))

Geometric.p_dist
```

Geometric Distribution
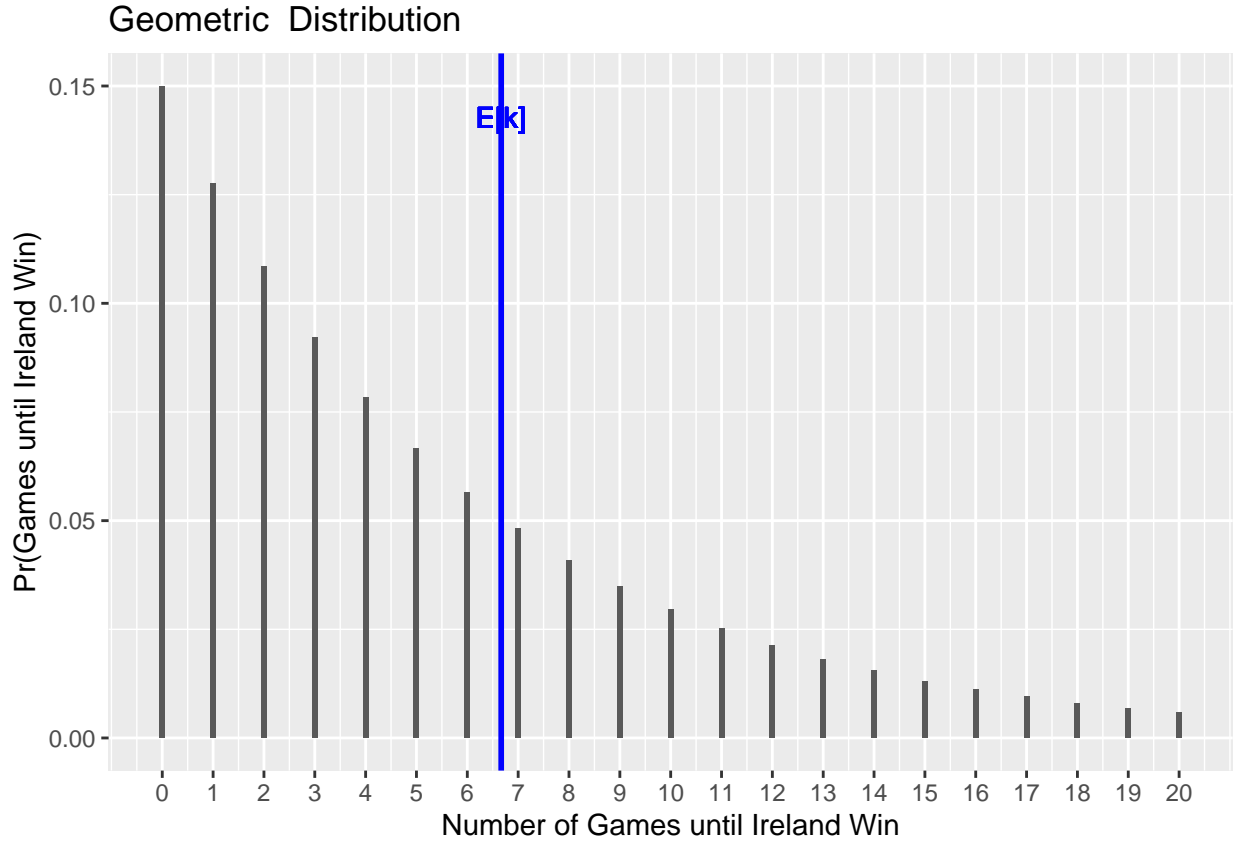
Poisson Distribution

A Poisson distribution is used to describe the probability distribution of the average number of events in a specific time period or distance. The distribution is described by the average, λ,

$$\Pr(k) = \frac{\lambda^k e^{-\lambda}}{k!}, \quad k = 0, 1, 2, ...$$

where k is the number of times the event happens. The expected value of the Poisson Distribution is

$$E[k] = \lambda,$$

with the variance

$$Var[k] = \lambda.$$

Example

The example we shall use to illustrate the Poisson distribution is the New Zealand vs Ireland World Cup Rugby Quarter Final. Let's assume that over a 10 year period Ireland win on average $\lambda = 2.2$ games. This gives the Poisson distribution

$$\Pr(k) = \frac{2.2^k e^{-2.2}}{k!}, \quad k = 0, 1, 2, ...$$

where k is the number of games Ireland beats New Zealand in a 10 year period that Ireland will win is

$$E[k] = 2.2,$$

the variance of the distribution is

$$Var[k] = 2.2.$$

The plot below shows the Poisson distribution for the number of times Ireland will beat New Zealand given an average number of wins $\lambda = 2.2$ over a 15 year period, the x-axis only show 0 to 20 possible wins, but do note the Poisson distribution can calculate the probability of infinite wins.

```r
Games<-0:20 # Number of Games
Lambda_ireland_win<-2.2 # Lambda
Pr_win<-c(dpois(Games,Lambda_ireland_win)) # Poisson Probabilities
E_X<-Lambda_ireland_win # Expected Outcome

df <- data.frame(Games,Pr_win)
## Plot
Poisson.p_dist<-ggplot(df, aes(x=Games,y=Pr_win)) +
  geom_col(width = 0.1)+xlab("Number of Irish Wins in a 15 year period")+
  ylab("Pr of Irish Wins")+ggtitle("Poisson Distribution")+
  geom_vline(xintercept = E_X,  color = "blue", size=1)+
  geom_text(aes(x=E_X, label="E[k]", y=max(Pr_win)),vjust=2,
            colour="blue",text=element_text(size=18))

Poisson.p_dist
```