Lecture 11: Statistics in soccer

Skidmore College, MA 276

Goals

- Metrics in soccer
- Expected goals
- Tools: Poisson distribution, Bradley Terry (review)

Review:

```
epl.df <- read.csv('http://www.football-data.co.uk/mmz4281/1314/E0.csv')
epl.df <- epl.df %>%
  select(Date, HomeTeam, AwayTeam, FTHG, FTAG) %>%
  mutate(Outcome = ifelse(FTHG > FTAG, 1, ifelse(FTHG < FTAG, 0, 0.5)))
head(epl.df)
```

```
##       Date   HomeTeam    AwayTeam FTHG FTAG Outcome
## 1 17/08/13    Arsenal Aston Villa    1    3     0.0
## 2 17/08/13  Liverpool       Stoke    1    0     1.0
## 3 17/08/13    Norwich     Everton    2    2     0.5
## 4 17/08/13 Sunderland      Fulham    0    1     0.0
## 5 17/08/13    Swansea  Man United    1    4     0.0
## 6 17/08/13  West Brom Southampton    0    1     0.0
```

Review:

```
homeBT <- BTm(Outcome,
              data.frame(team = HomeTeam, home.adv = 1),
              data.frame(team = AwayTeam, home.adv = 0),
              ~ team + home.adv,
              id = "team", data = epl.df)
head(BTabilities(homeBT))
```

```
##                    ability      s.e.
## Arsenal          0.0000000 0.0000000
## Aston Villa     -1.6500862 0.5230498
## Cardiff         -1.9657610 0.5349346
## Chelsea          0.1446607 0.5382662
## Crystal Palace -1.4101387 0.5170699
## Everton         -0.2706003 0.5213307
```

Review

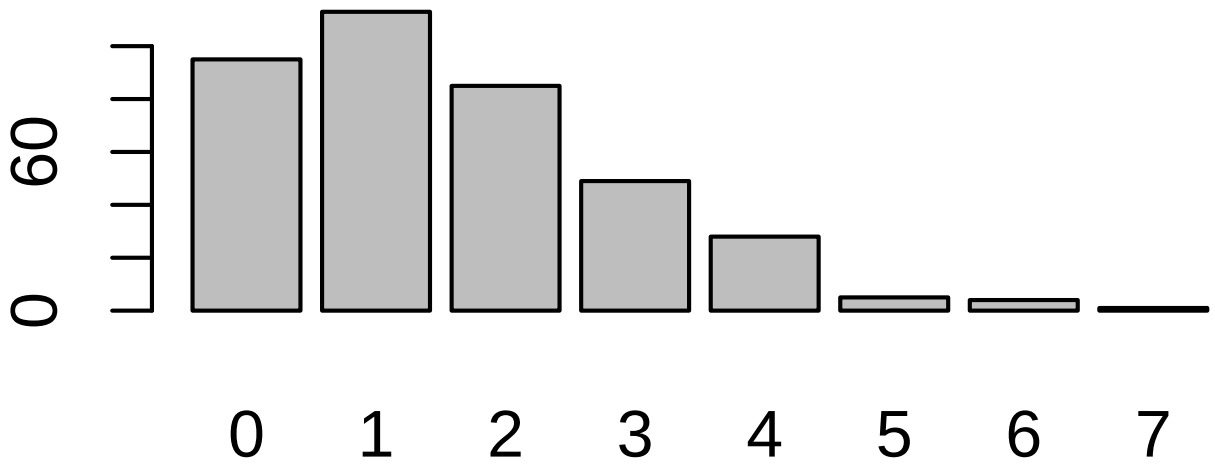
```
tail(homeBT$coeff)
```

```
## teamSunderland    teamSwansea  teamTottenham  teamWest Brom   teamWest Ham
##     -1.6500862     -1.4694329     -0.4615757     -1.5893264     -1.5893264
##       home.adv
##      0.3746677
```

1. Estimate the probability that Aston Villa beats Chelsea on a neutral field
2. Estimate the probability that Aston Villa beats Chelsea when playing at home
3. Let your answer to (1) be $\hat{p}_1$ and your answer to (2) be $\hat{p}_2$. Calculate $\frac{\hat{p}_2/(1-\hat{p}_2))}{\hat{p}_1/(1-\hat{p}_1))}$.

Sidebar: What is this shape?

```
barplot(tally(epl.df$FTHG))
```



Sidebar: Poisson distribution

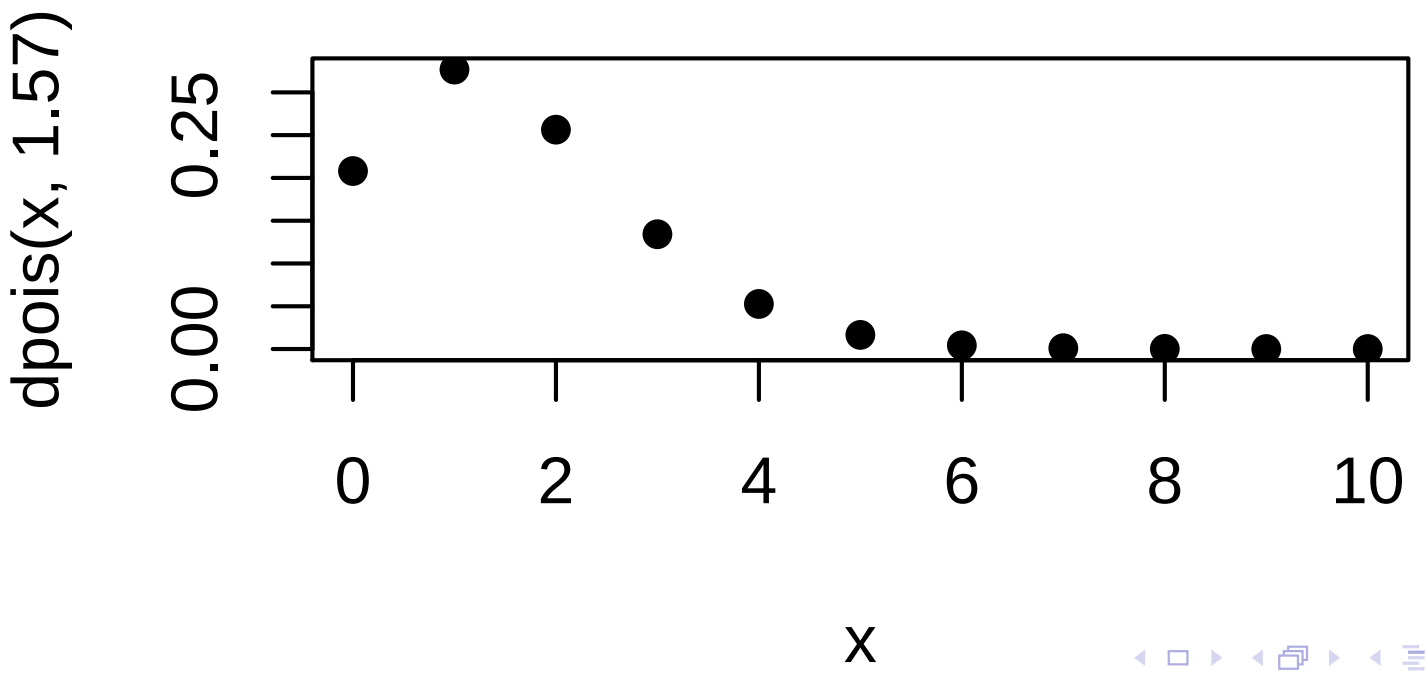
P(x events in interval) $= \frac{\lambda^x e^{-\lambda}}{x!}$ for $x = 0, 1, ...$

```
mean(epl.df$FTHG); var(epl.df$FTHG)
```

```
## [1] 1.573684
```

```
## [1] 1.896931
```

```
## Lambda ~ 1.57
x <- seq(0, 10, 1)
plot(x, dpois(x, 1.57), pch = 16)
```



Soccer analytics

Summarize: *Introduction to Analytics in . . . Soccer* (link)

Those who don't understand the past are . . .

Summarize: *What analytics can teach us about the beautiful game* (link)

Expected goals

Summarize: *Expected goals 2.0* (link)

Expected goals & predictability

Summarize: *Best predictor of future performance is expected goals* (link)

Expected goals, repeatability of finishing skill

Summarize: *Repeatability of finishing skill* (link)

Randomness and expected goals

Summarize: *12 shots good, 2 shots better* (link)