MA615 The problem of misprint

Si Chen

September 16, 2018

1 Calculate p_k

Given the stated mean value of 2 misprints per page, assume the distribution of M, the number of misprints on a page, is $Poisson(\lambda = 2)$

Conditioned on k, a limit to be determined, the probability p_k of more than k misprints on a page is:

$p_k = P(M > k) = 1 - P(M \leq k) = 1$ - ppois(k, lambda $= 2$).

2 Calculate $P(T \leq n|k)$

Probability of T no more than n pages with more than k misprints is:

$P(T \leq n) =$ pbinom(n, 50, p_k).

3 Visualize $P(n \leq k)$

For a given k, the probability mass and cumulateive distribution functions of errors per page can be visualized as follows:

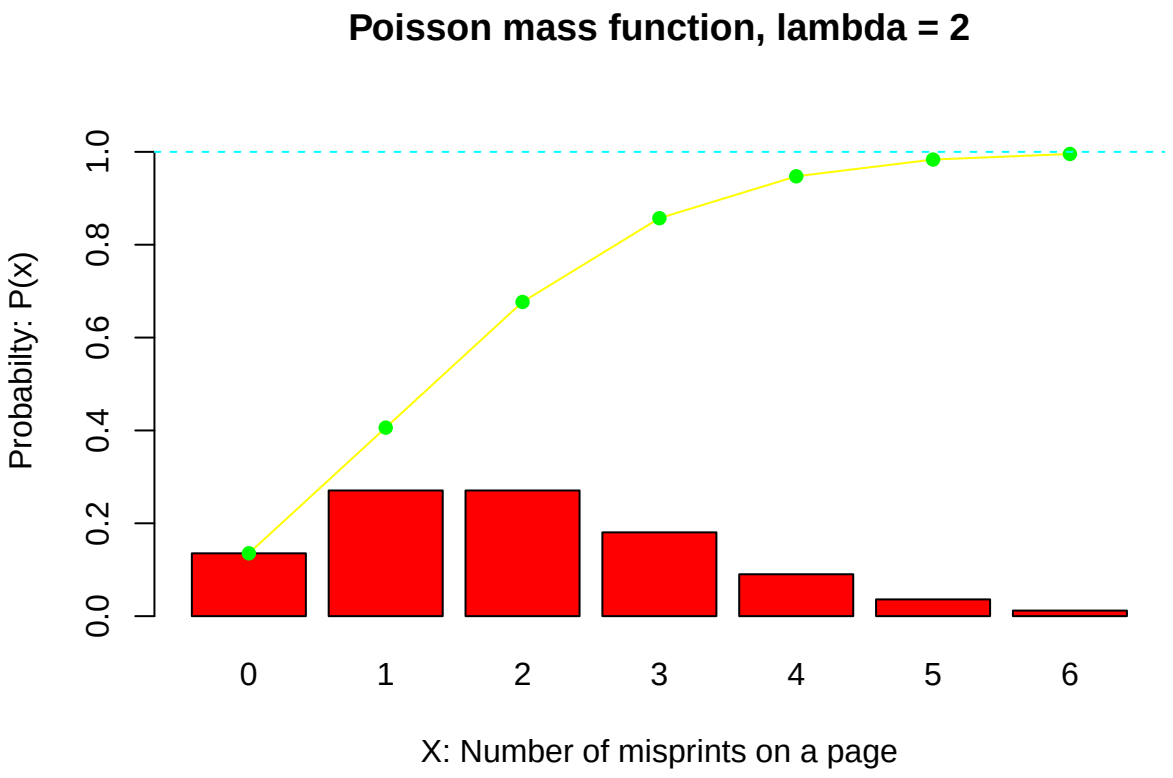
```
options(digits = 3)
options(scipen = 999)

## produce a dataframe containing probabilities
prob <- as.data.frame(dpois(x = 0:6, lambda = 2))
prob <- cbind(prob, cumsum(prob)) ## these probabilities = ppois(0:10,lambda=2)
prob <- cbind(prob, 1 - prob[,2])

colnames(prob) <- c("prob", "cprob", "tail_prob")

## barplot with mass function and CDF
df.bar <- barplot(prob$prob,
                  main="Poisson mass function, lambda = 2",
                  names.arg = c("0","1","2","3","4","5","6"),
                  xlab = "X: Number of misprints on a page",
                  ylab = "Probabilty: P(x)",ylim = c(0,1.1), col="red")

lines(x=df.bar, y=prob$cprob, col="yellow")
points(x=df.bar, y=prob$cprob, pch=16, col="green")
abline(h=1, lty=2, col = "cyan")
```




```
library(knitr)
library(kableExtra)

# suoppose there are usually no more than 6 misprints in one page for the book

m <- c(0:6)

# for each page, it may have 0,1,2,3,4,5,6 misprints.

misprint <- ppois(m,2,lower.tail = FALSE)
misprint
```

```
## [1] 0.86466 0.59399 0.32332 0.14288 0.05265 0.01656 0.00453
```

```
# use binomial function to count for every 10 page:10,20,30,40,50
# define each page range

errorpage10 <- dbinom(1,10,misprint)
errorpage20 <- dbinom(11,20,misprint)
errorpage30 <- dbinom(21,30,misprint)
errorpage40 <- dbinom(31,40,misprint)
errorpage50 <- dbinom(41,50,misprint)

# combine five 10-pages ranges
errorpage <- cbind(errorpage10, errorpage20, errorpage30, errorpage40, errorpage50)

# set the table names
```

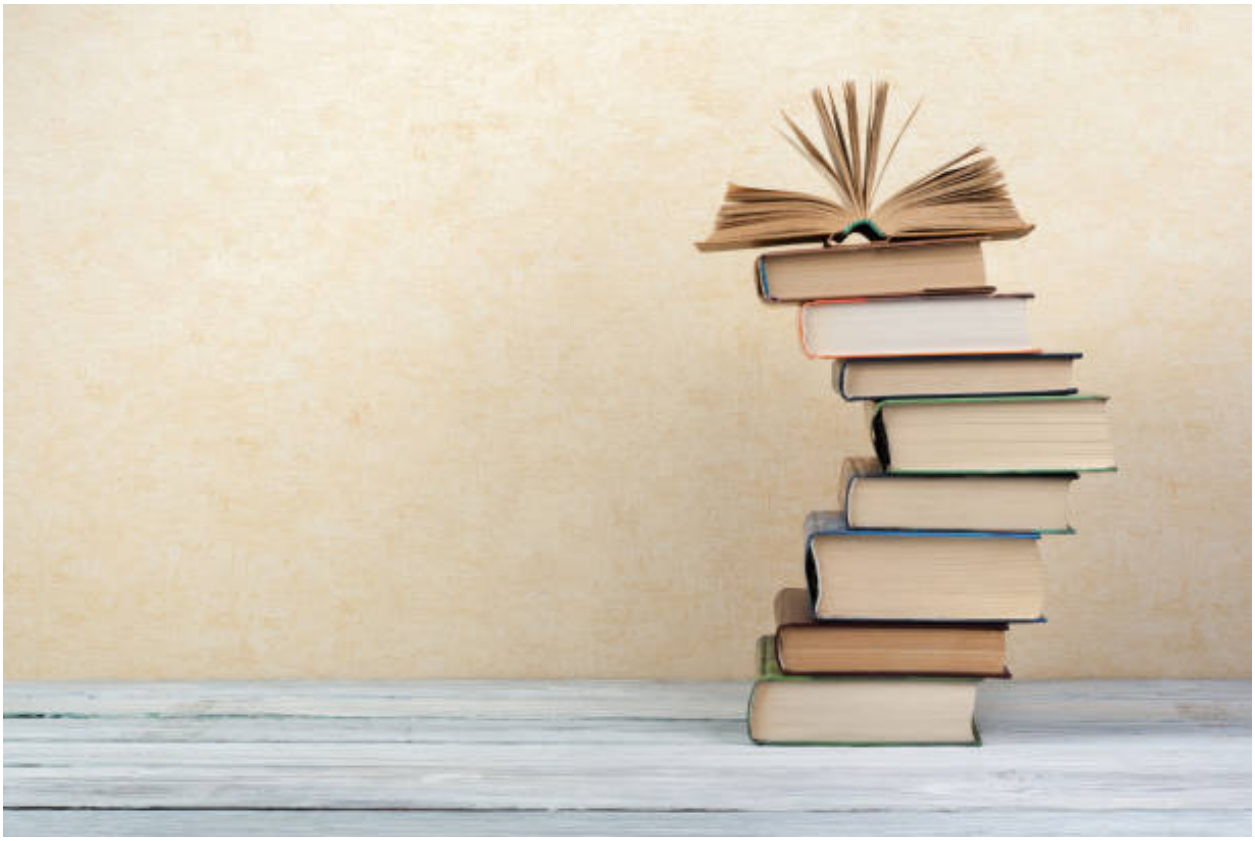
```r
colnames(errorpage) <- c("page 1-10","page 11-20","page 21-30","page 31-40","page 41-50")
rownames(errorpage) <- c("0 error","1 error","2 errors","3 errors","4 errors","5errors","6 errors")

# construct table
Errorpage <- kable(errorpage, "latex")

# table styling
kable_styling(Errorpage, bootstrap_options = "striped", font_size=16 )
```

	page 1-10	page 11-20	page 21-30	page 31-40	page 41-50
0 error	0.000	0.001	0.010	0.046	0.098
1 error	0.002	0.164	0.076	0.008	0.000
2 errors	0.096	0.020	0.000	0.000	0.000
3 errors	0.357	0.000	0.000	0.000	0.000
4 errors	0.324	0.000	0.000	0.000	0.000
5errors	0.143	0.000	0.000	0.000	0.000
6 errors	0.044	0.000	0.000	0.000	0.000

```r
# insert pic
# pic URL:https://github.com/sichen55/MA615-homework.git
knitr::include_graphics('books.JPG')
```



```
#finished
```

Note: For every page, use possion distribution to calculate the possibility of making certain number of mistakes.

Use page range by every 10 pages, otherwise if we list every page, we may need to use loop.

Use binomial function to find the error number in each page, where the probability found in step 1 is the p value in binomial equation.

set and style the table to expected structure.