Practice4

Qiang Liu

2022-09-27

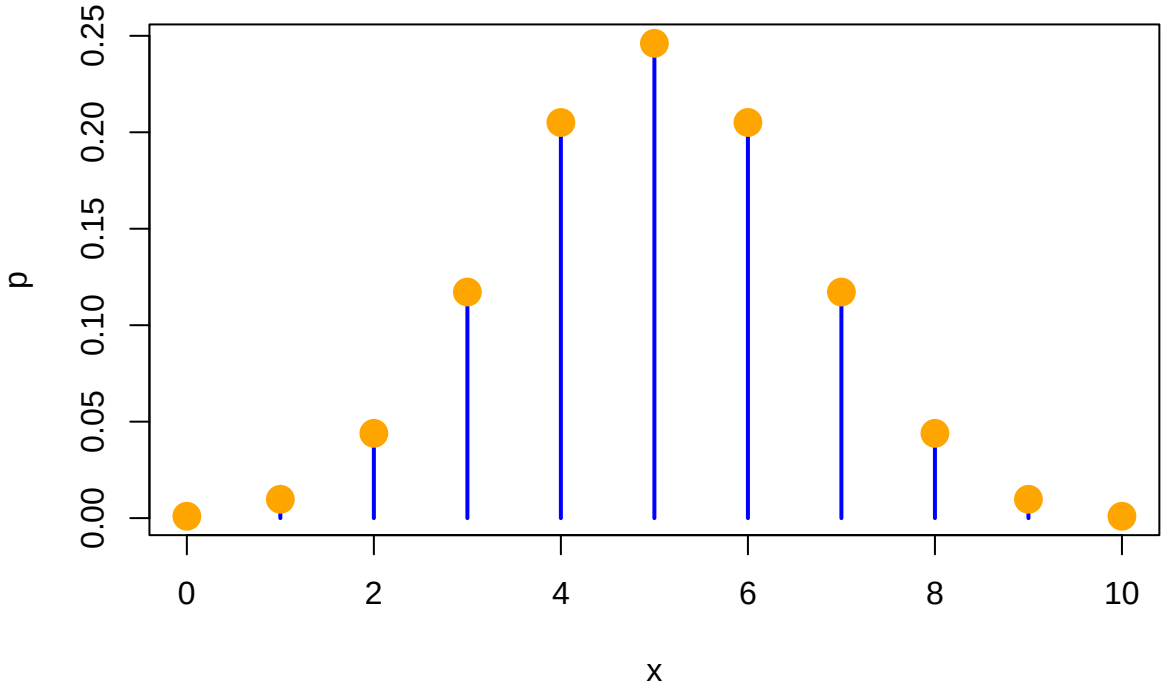
```
> p <- 0.5 # male
> q <- 1 # female
> n <- 10 # number of trial
> x <- 3 # number of male
>
> #using formula
> p3 <- choose(n,x)*p^x*q^(n-x)
> p3
```

```
## [1] 15
```

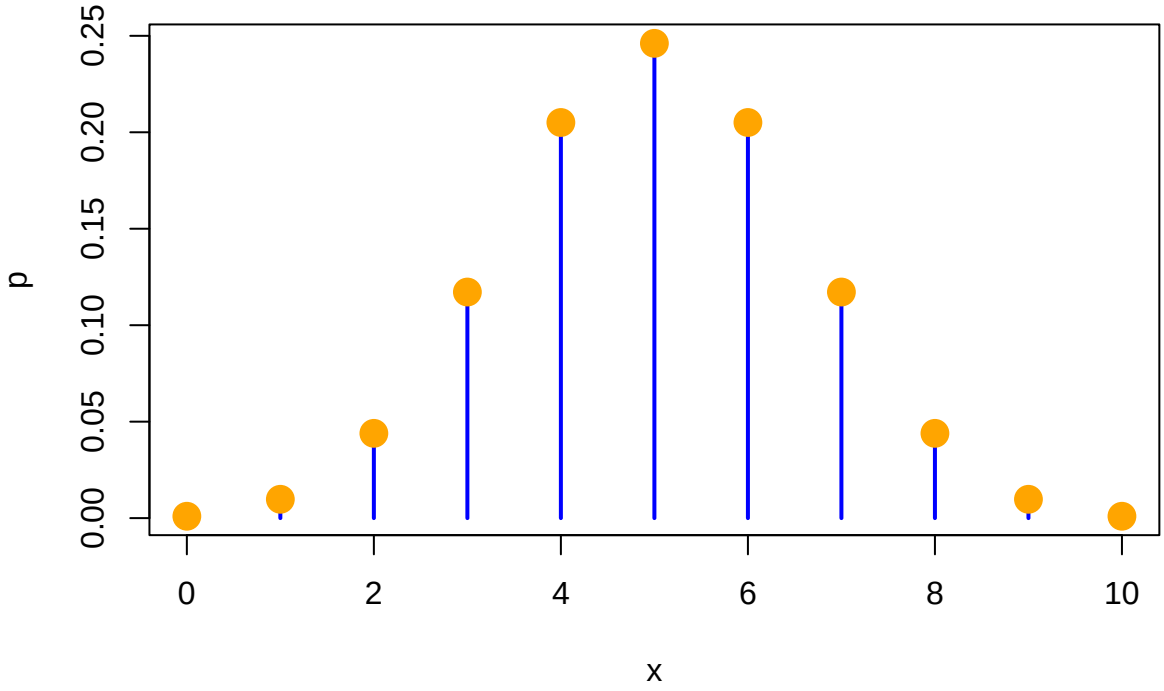
```
> dbinom(x,size = n,prob = p)
```

```
## [1] 0.1171875
```

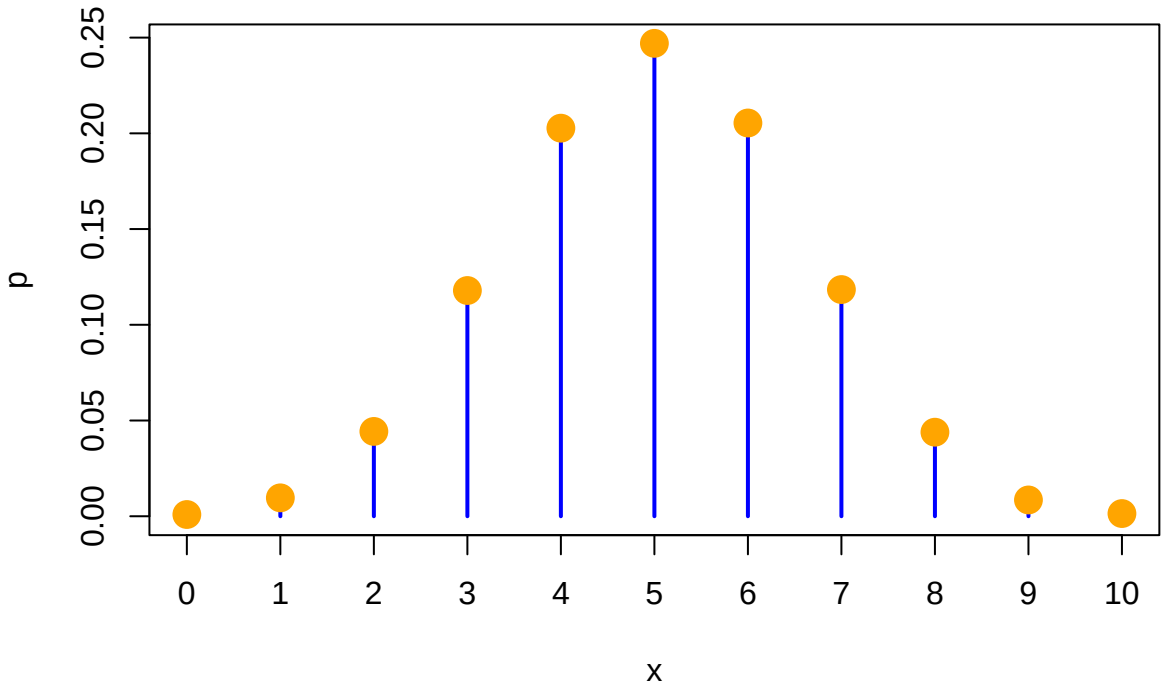
```
> # density plot
> x <- 0:n
> dens <- dbinom(x,size = n,prob = p) # density for the binomial distribution
> plot(x,dens,type = "h", lwd = 2, col = "blue",ylab = "p")
> points(x,dens,pch = 16,cex = 2,col = "orange")
```



```
> # distribution plot
> dist <- pbinom(x,size = n,prob = p) # distribution for the binomial distribution
> plot(x,dens,type = "h", lwd = 2, col = "blue",ylab = "p")
> points(x,dens,pch = 16,cex = 2,col = "orange")
```



```
> # calculate the population expectation, variance and moment
> mu <- n*p
> sigma2 <- n*p*(1-p)
>
> #sampling
> set.seed(123456)
> N <- 10000
> x <- rbinom(N,size = n,prob = p) # random generation for the binomial distribution
> y <- table(x)/length(x)
> plot(y,xlim = c(0,n),type = "h",lwd = 2,col = "blue",xlab = "x",ylab = "p")
> points(as.numeric(names(y)),y,pch = 16,cex = 2,col = "orange")
```



```
> # mean, sd and skewness/kurtosis
> mean(x)
```

```
## [1] 5.0006
```

```
> var(x)
```

```
## [1] 2.494449
```

Poisson distribution

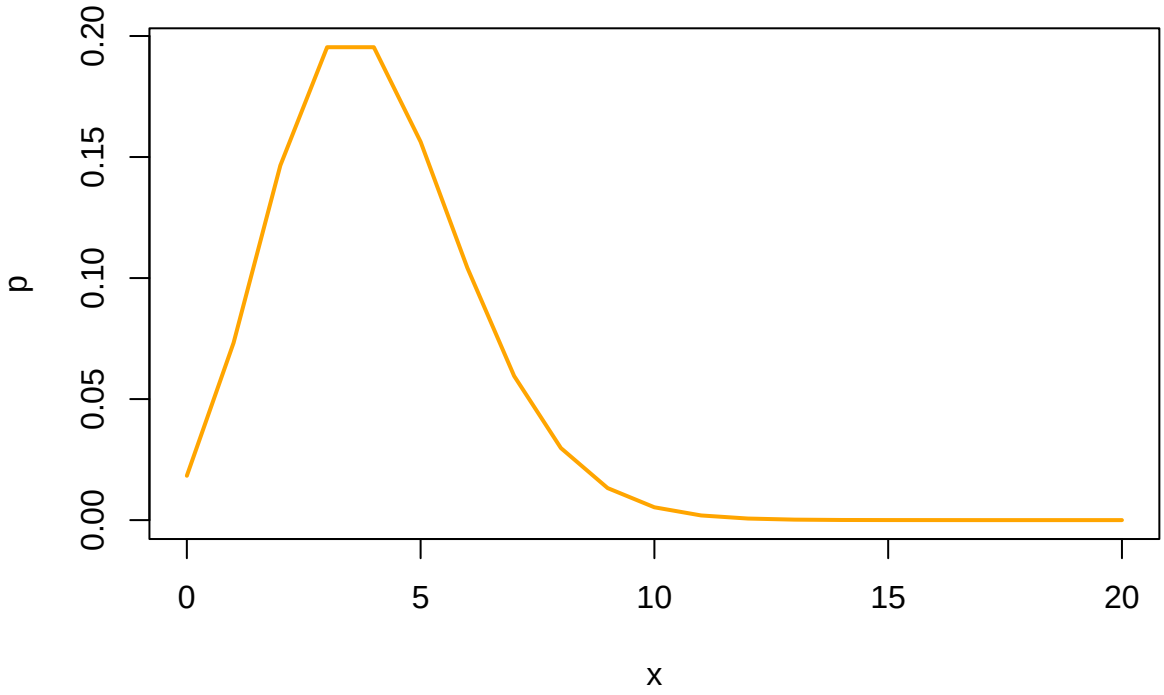
```
> p <- 0.01 # male
> q <- 1 - p # female
> n <- 400 # number of trial
> x <- 5 # number of male
>
> # using binomial distribution
> dbinom(x,size = n,prob = p)
```

```
## [1] 0.1570805
```

```
> # using Poisson distribution for approximation
> lambda <- n*p
> dpois(x,lambda = lambda)
```

```
## [1] 0.1562935
```

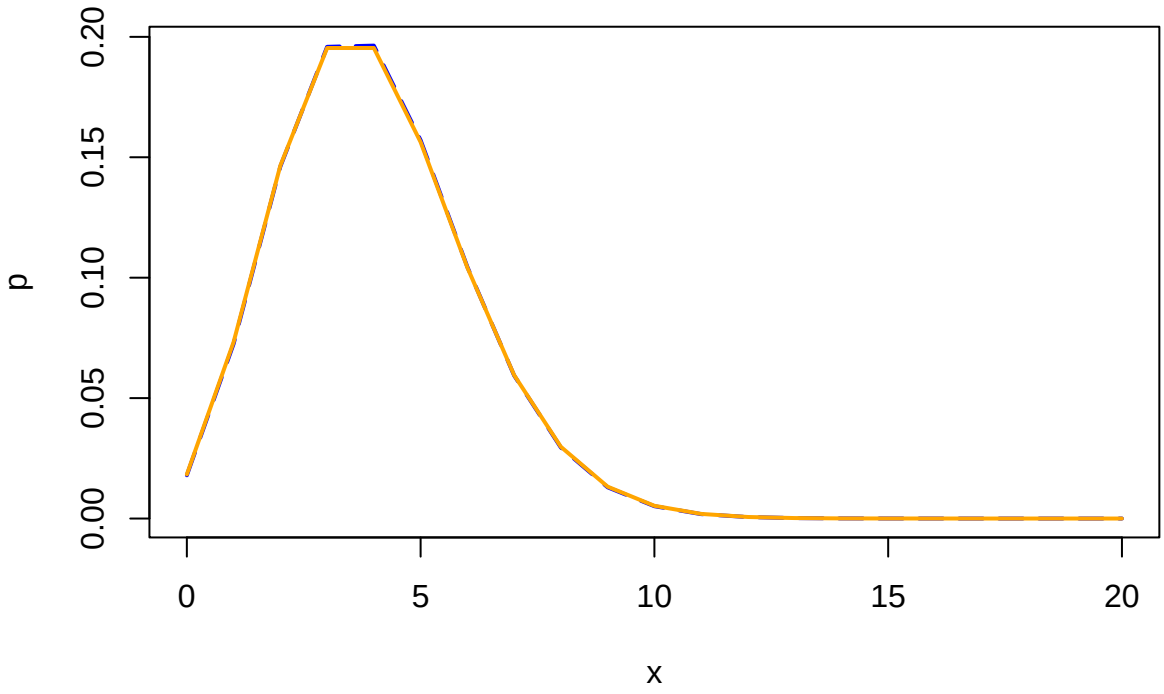
```
> # density plot
> x <- 0:20
> dens <- dpois(x,lambda = lambda) # density for the Poisson distribution
> plot(x,dens,type = "l",lwd = 2,col = "orange",ylab = "p")
```



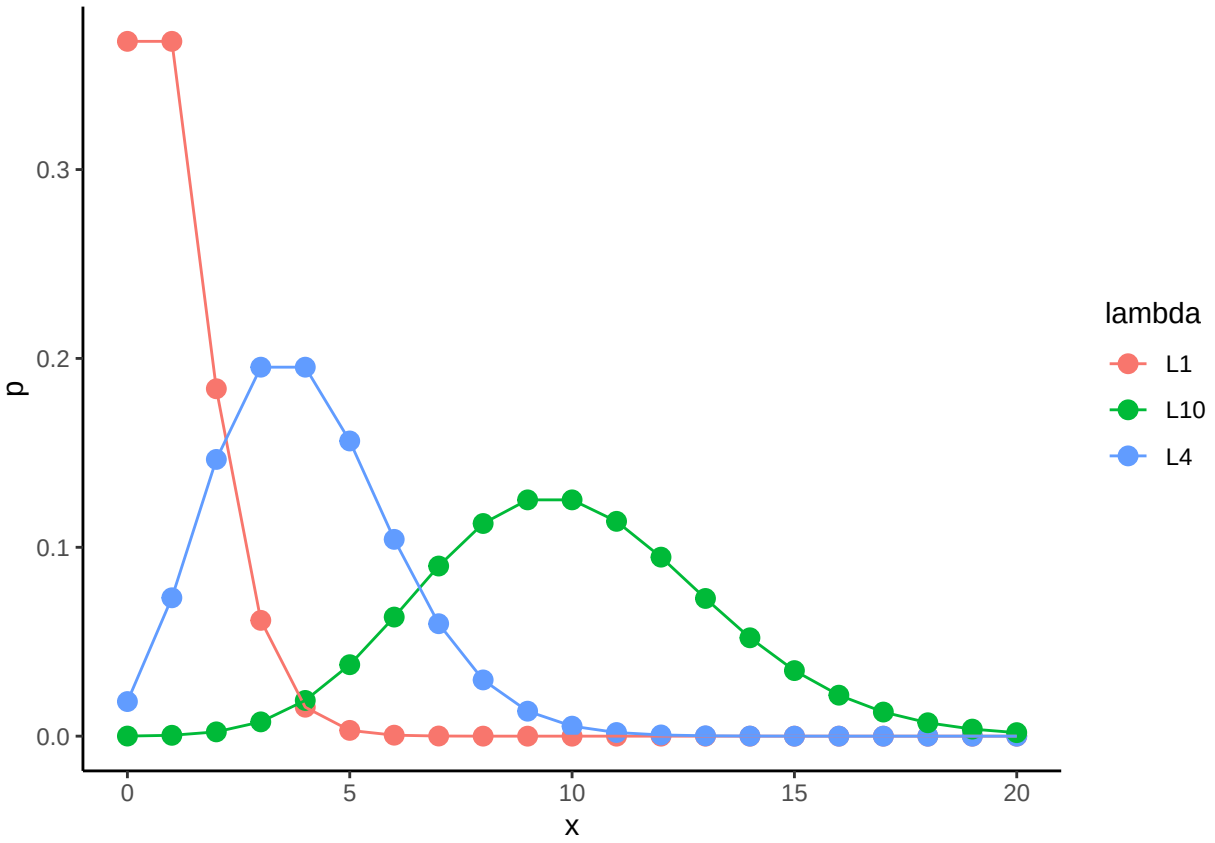

```
> # Compare the Poisson and bionomial distributions
> p <- 0.01
> n <- 400
> x <- 0:20
> bd <- dbinom(x,size = n,prob = p)
> pd <- dpois(x,lambda = n*p)
>
> plot(x,bd,type = "l",ylab = "p",col = "blue",lwd = 2,lty = 5)
> lines(x,pd,col = "orange",lwd = 2)
>
> # with different lambda
```

```
> lambdas <- c(1,4,10)
> x <- 0:20
> ps <- lapply(lambdas,function(lambda){
+   dpois(x,lambda = lambda)
+ })
>
> # list --> data.frame
> ps <- as.data.frame(ps,col.names = sprintf("L%s",lambdas))
>
> # reformat: wide --> long
> library(tidyr)
> df <- data.frame(x = x,ps)%>%gather(lambda,p,L1:L10)
> library(ggplot2)
```

```
## Warning:   'ggplot2' R 4.2.1
```




```
> df%>%ggplot(aes(x,p,color = lambda)) +
+   geom_point(size = 3) + geom_line() +
+   theme_classic()
```



Hypergeometric distribution

```
> # dhtper(x,m,n,k,log = FALSE)
> m <- 10;n <- 7;k <- 8
> x <- 0:(k+1)
> rbind(phyper(x,m,n,k),dhyper(x,m,n,k))
```

```
##      [,1]         [,2]       [,3]     [,4]      [,5]      [,6]      [,7]
## [1,]    0 0.0004113534 0.01336898 0.117030 0.4193747 0.7821884 0.9635952
## [2,]    0 0.0004113534 0.01295763 0.103661 0.3023447 0.3628137 0.1814068
##            [,8]       [,9] [,10]
## [1,] 0.99814891 1.00000000     1
## [2,] 0.03455368 0.00185109     0
```

```
> all(phyper(x,m,n,k) == cumsum(dhyper(x,m,n,k)))
```

```
## [1] FALSE
```