PairProgramming_Module4_Bernoulli

HDS

2024-08-26

PairProgramming_Module4_Bernoulli, DSE5001

HD Sheets, Aug 26, 2024 checked 01/03/2025

Student Name: Ryan Waterman Partner Name: Date Completed: 2/5/25

Examples from Open Intro to Statistics by Diez et al

Question 4.13, Geometric series

4.13 Eye color, Part I. A husband and wife both have brown eyes but carry genes that make it possible fortheir children to have brown eyes (probability 0.75), blue eyes (0.125), or green eyes (0.125).

a. What is the probability the first blue-eyed child they have is their third child? Assume that the eye colors of the children are independent of each other.

Okay, this is the geometric series

we need two children without blue eyes in a row, then a blue eyed child

```
p(blue eyes on 3)  = P(not blue) *P(not blue)*P(blue)

p(blue) = 0.125
p(notblue)= 1-p(blue) =0.875
```

Here is the r calculation

```
pblue=0.125
pnotblue=1-pblue

p_blue_3= pnotblue*pnotblue*pblue

print(p_blue_3)
```

```
## [1] 0.09570312
```

About a 10% chance

Okay, I did this using “raw logic” and the product rule of unrelated events

We do also have dgeom()-pdf, pgeom()-cdf, qgeom-quantile and rgeom()- random value functions in r

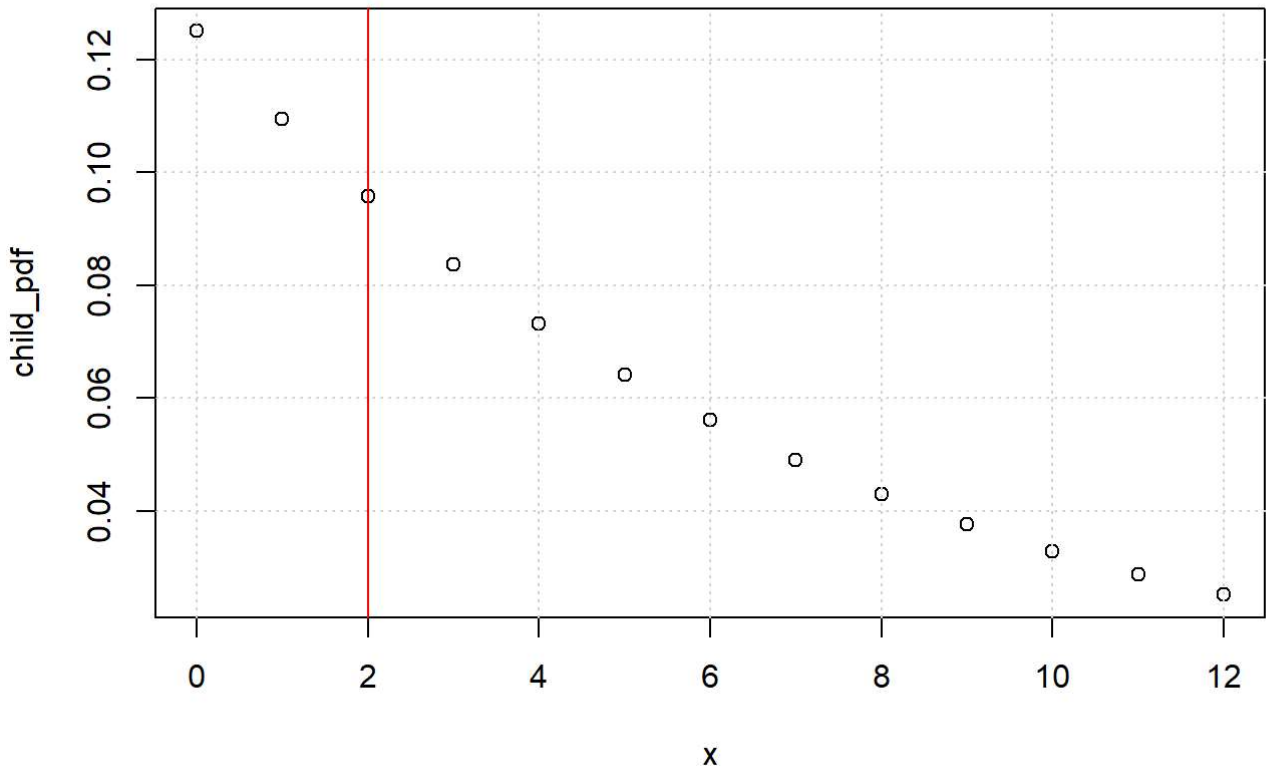
Note, this is a discrete distribution, we can have k=0,1,2,3,4….children, 2.3 children ain't gonna fly

Notice k is the number of failures before a success, so we have k=2 here, two non-blue, then 1 blue

The parameter we care about is p(blue eyes)= 0.125 per child and k=2

We can plot the geometric series probabilities from 1 to 12 children

```
x=0:12

child_pdf=dgeom(x, prob=0.125)

plot(x,child_pdf)
grid()

abline(v=2,col="red")
```



We can see the probability of the first two not having blue eyes and the third have blue looks to be about 0.095 or so

We can get the exact value

```
dgeom(2,prob=0.125)
```

```
## [1] 0.09570312
```

b. On average, how many children would such a pair of parents have before having a blue-eyed child? What is the standard deviation of the number of children they would expect to have until the first blue-eyed child?

The mean value of a geometric series, is 1/p, where p is the probability of success. Here p=0.125

```
p=0.125
1/p
```

```
## [1] 8
```

so on average, 8 kids before the first child comes up with blue eyes

the SD of a geometric series is $(1-p/p^{2)}0.5$- see the notes

```
((1-p)/p^2)^0.5
```

```
## [1] 7.483315
```

Question 4.17 Binomial

4.17 Underage drinking, Part I. Data collected by the Substance Abuse and Mental Health Services Administration (SAMSHA) suggests that 69.7% of 18-20 year olds consumed alcoholic beverages in any given year

a. Suppose a random sample of ten 18-20 year olds is taken. Is the use of the binomial distribution appropriate for calculating the probability that exactly six consumed alcoholic beverages? Explain.

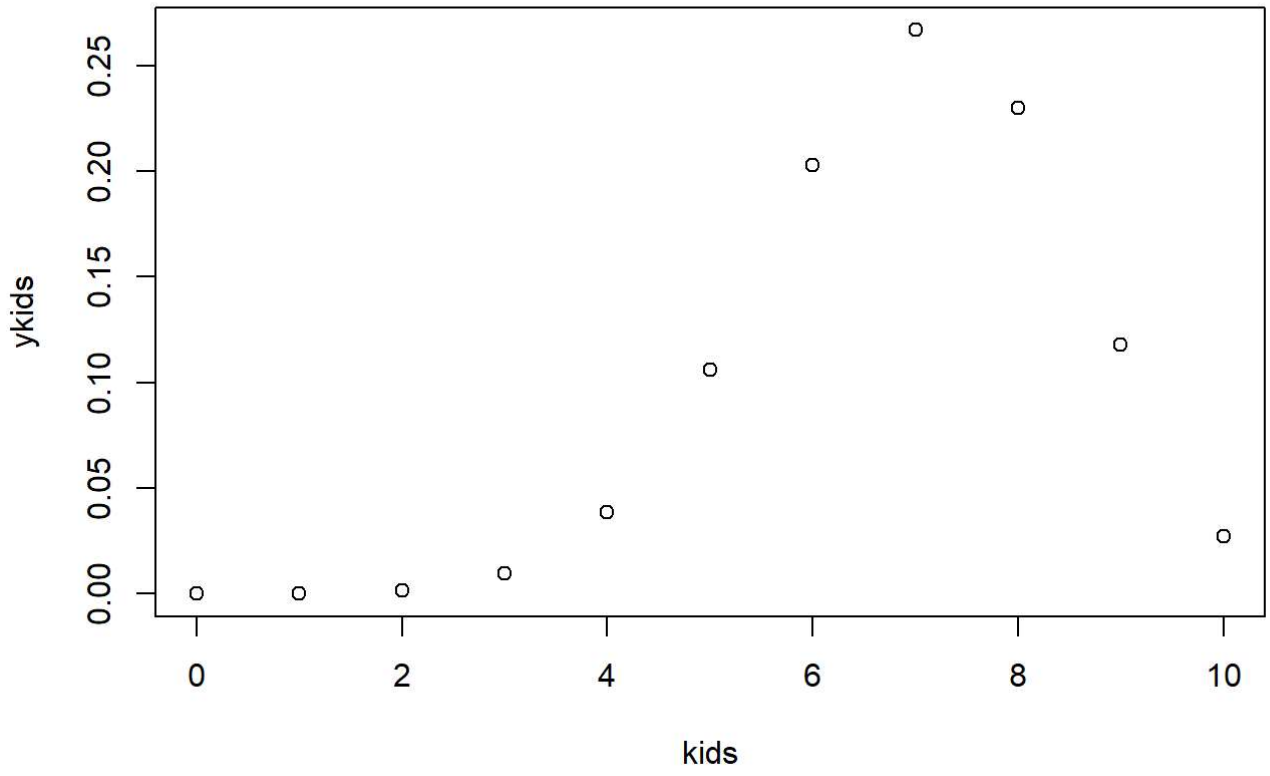
This is the probability of picking 6 out of 10 trials that are successes (ie drink), where p=0.697 that each one is a success (ie drinks)

Dang- I have just called underage drinking a success, a tad sub-optimal there…

This is a Binomial, with N=10, p=0.697 and k=6

Here is the pdf

```
kids=0:10

#r uses "size" for the value of N,  we want dbinom to get the pdf

ykids=dbinom(kids,size=10,prob=0.697)

plot(kids,ykids)
```



Not

surprisingly, a max around seven

here mean=p*N= 6.97

Variance =p(1-p)N

std dev= Variance^0.5

b. Calculate the probability that exactly 6 out of 10 randomly sampled 18- 20 year olds consumed analcoholic drink.

we have available the usual r family of probability functions

dbinom (pdf), pbinom (cdf), qbinom (quantile), rbinom (random value)

this is a discrete distribution, kids are integers

```
#we want dbinom() at N=0,k=6, p=0.697

dbinom(x=6,size=10,p=0.697)
```

```
## [1] 0.2029488
```

c. What is the probability that exactly four out of ten 18-20 year olds have not consumed an alcoholic beverage?

Probability that 4 out of 10 did not. This is the same as the probability that 10-4=6 did

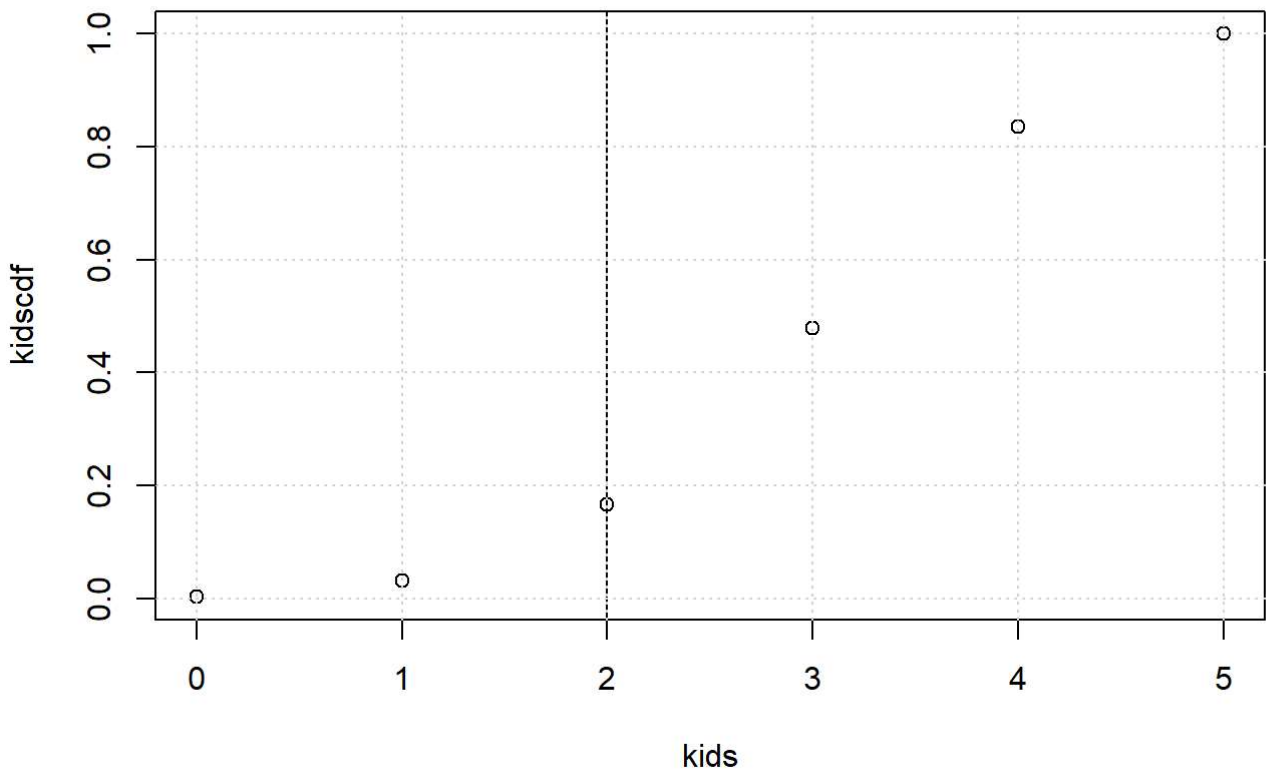
which was 20.2% as above

d. What is the probability that at most 2 out of 5 randomly sampled 18-20 year olds have consumed alcoholic beverages?

The phrase at most here should be a clue that we want the cummulative cdf

this is the cdf P(x,<=2)

```
#here's the plot,

kids=0:5

kidscdf=pbinom(kids,size=5, p=0.697)

plot(kids,kidscdf)

abline(v=2)

grid()
```



Looks like about 18% or 0.18

Here's the exact answer

```
pbinom(2,size=5,p=0.697)
```

```
## [1] 0.1670716
```

e. What is the probability that at least 1 out of 5 randomly sampled 18-20 year olds have consumed alcoholic beverages?

at least 1 of 5

this is 1-P(x<1) = 1-p(x==0)

```
1-pbinom(0,size=5,p=0.697)
```

```
## [1] 0.997446
```

Question 4.21

4.21 Game of dreidel. A dreidel is a four-sided spinning top with the Hebrew letters nun, gimel, hei, and shin, one on each side. Each side is equally likely to come up in a single spin of the dreidel. Suppose you spin a dreidel three times. Calculate the probability of getting

a. at least one nun?

P(at least one nun)=1-P(0 nun)

```
P_one_nun=1-((3/4)^3)
P_one_nun
```

```
## [1] 0.578125
```

```
P_one_nun_binom=1-pbinom(0,size=3,prob=0.25)
P_one_nun_binom
```

```
## [1] 0.578125
```

About 58%

b. exactly 2 nuns?

```
P_two_nun=dbinom(2,size=3,prob=0.25)
P_two_nun
```

```
## [1] 0.140625
```

About 14%

c. exactly 1 hei?

```
P_one_hei=dbinom(1,size=3,prob=0.25)
P_one_hei
```

```
## [1] 0.421875
```

About 42%

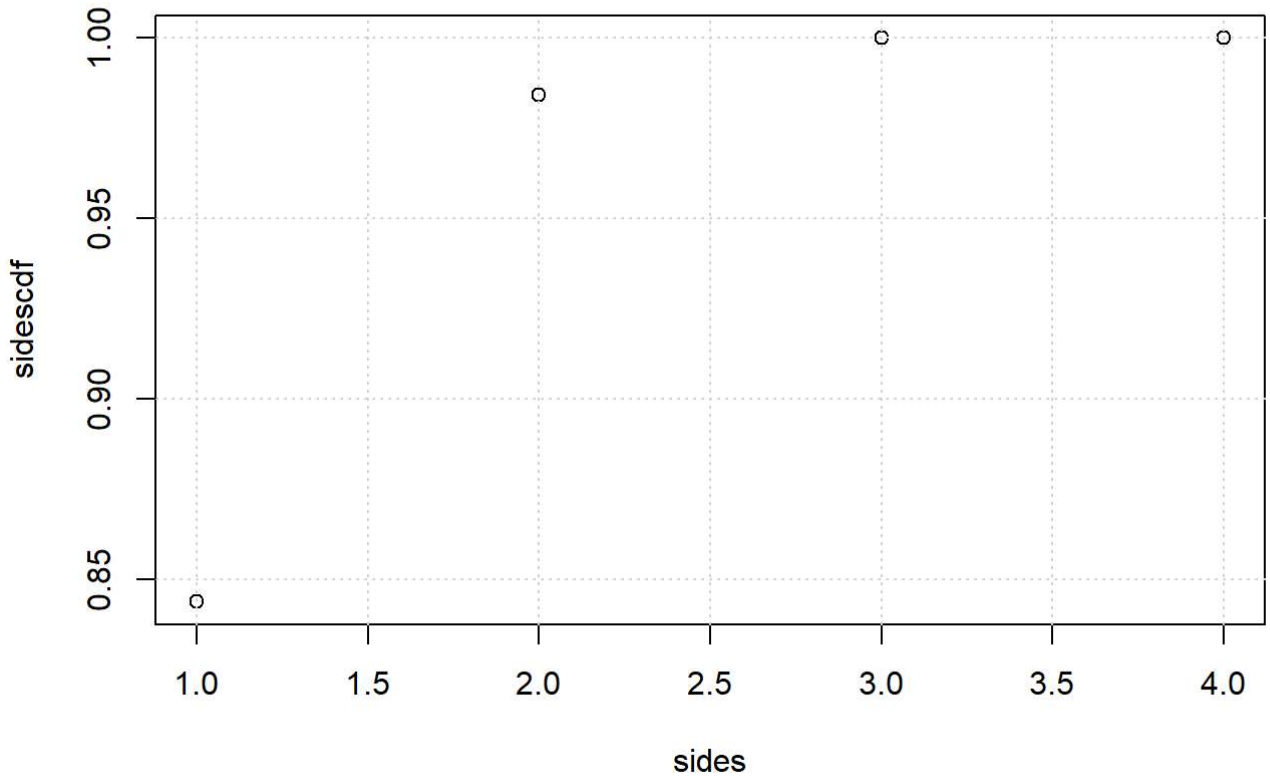
d. at most 2 gimels?

At most 2 indicates that we are interested in two or less gimels out of 3 spins, which points to the cdf

```
P_two_gimels=pbinom(2,size=3,prob=0.25)
P_two_gimels
```

```
## [1] 0.984375
```

```
sides=1:4
sidescdf=pbinom(sides,size=3, p=0.25)
plot(sides,sidescdf)
grid()
```



About 98%

Question/Action

What is p for each letter (nun,gimel,hei, shin)?

P for each is 25%

What is the N or size of this problem?

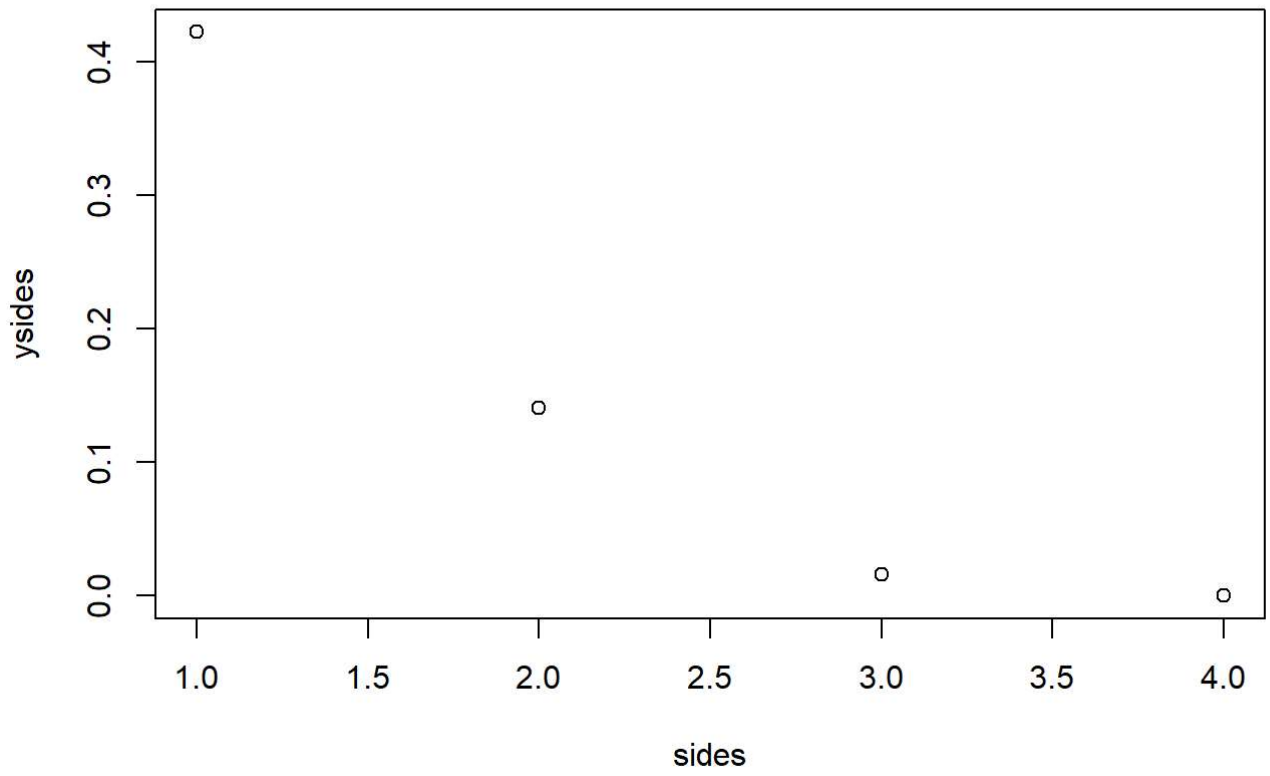
3

What distribution should you use?

Binomial

Plot the distribution of outcomes for N,k and p. You can cut and paste one of the earlier examples and just alter the N,k and p

```
N=3
k=1:4
p=0.25

#r uses "size" for the value of N,  we want dbinom to get the pdf

ysides=dbinom(k,size=N,prob=p)

plot(sides,ysides)
```



4.21a,b,c,d

We know the distribution and the parameters now.

For each question a,b,c,d

Is this the distribution (pdf) or cummulative distribution (cdf)?

Find the probability for each question, a,b,c,d