STAT3023 Statiscal Inference

Lab Week 2

Tutor: Wen Dai

SID: 470408326

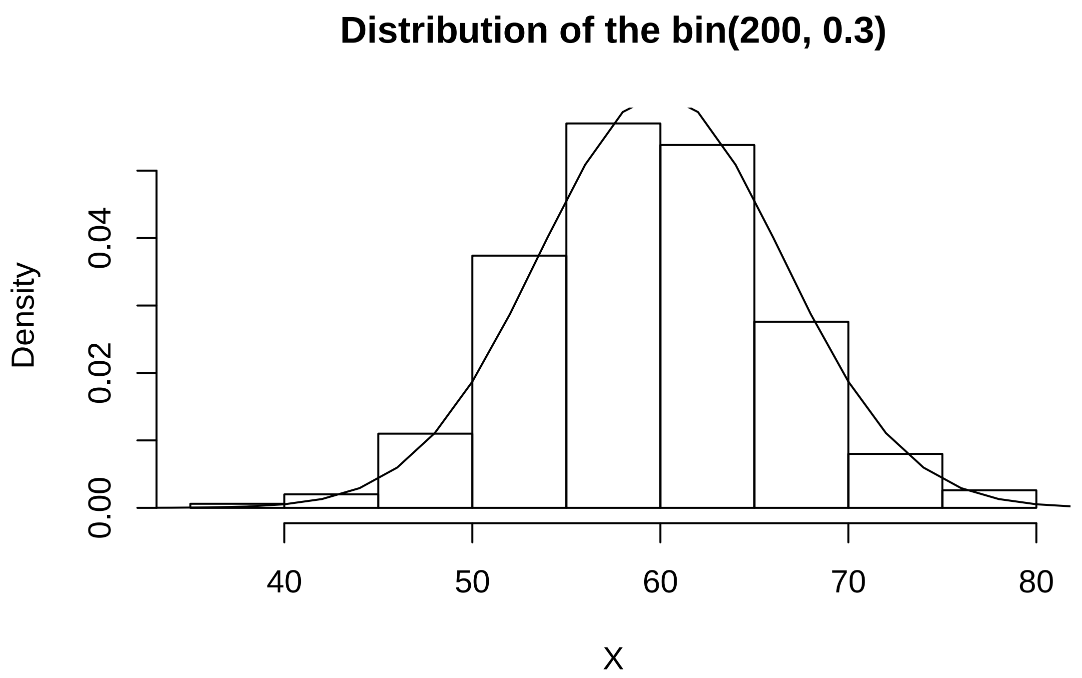
School of Mathematics and Statistics

The University of Sydney

Semester 1, 2021

1. (a) We generate 1000 samples, sampling from a binomial distribution with size being 200 and $p = 0.3$

```
samples = 1000
size = 200
p = 0.3
X = rbinom(samples, size, p)
hist(X, prob = TRUE, main = "Distribution of the bin(200, 0.3)")
X_mean = size * p
X_sd = sqrt(size * p * (1 - p))
curve(dnorm(x, mean = X_mean, sd = X_sd), xlim = c(0, 200), add = TRUE)
```



(b) We find the $P(45 \leq X \leq 55)$

```
pbinom(55, size, p) - pbinom(45, size, p)

[1] 0.2343248
```

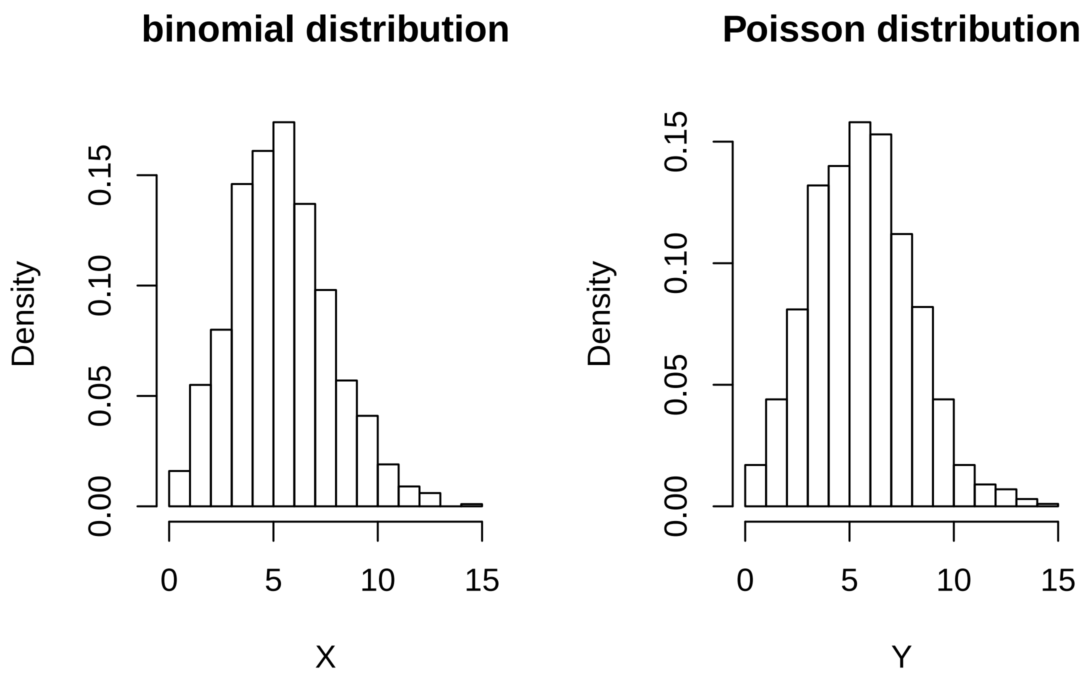
(c) We use the normal approximation to the binomial with mean np and standard deviation $np(1-p)$ to approximate the same probability. With the continuity correction, our probability becomes

$$P(45 \leq X \leq 55) = P(44.5 \leq X \leq 55.5)$$

```
pnorm(55.5, mean = X_mean, sd = X_sd) - pnorm(45.5, mean = X_mean, sd = X_sd)

[1] 0.2310965
```

2. (a) We simulate drawing 1000 times from the binomial and the poisson distribution here:

```
simulations = 1000
size = 200
p = 0.03
lambda = 6

# 1000 draws of a binomial distribution
X = rbinom(simulations, size, p)
# 1000 draws of a poisson distribution
Y = rpois(simulations, lambda)
# plot with 1 row and 2 columns
par(mfrow=c(1,2))
# Historgrams of X and Y
hist(X, probability = TRUE, main = "binomial distribution", breaks = 15)
hist(Y, probability = TRUE, main = "Poisson distribution", breaks = 15)
```



(b) We find $P(X \leq 5)$ and $P(Y \leq 5)$

```
pbinom(5, size = 200, prob = 0.03)

[1] 0.4432292

ppois(5, lambda = 6)

[1] 0.4456796
```

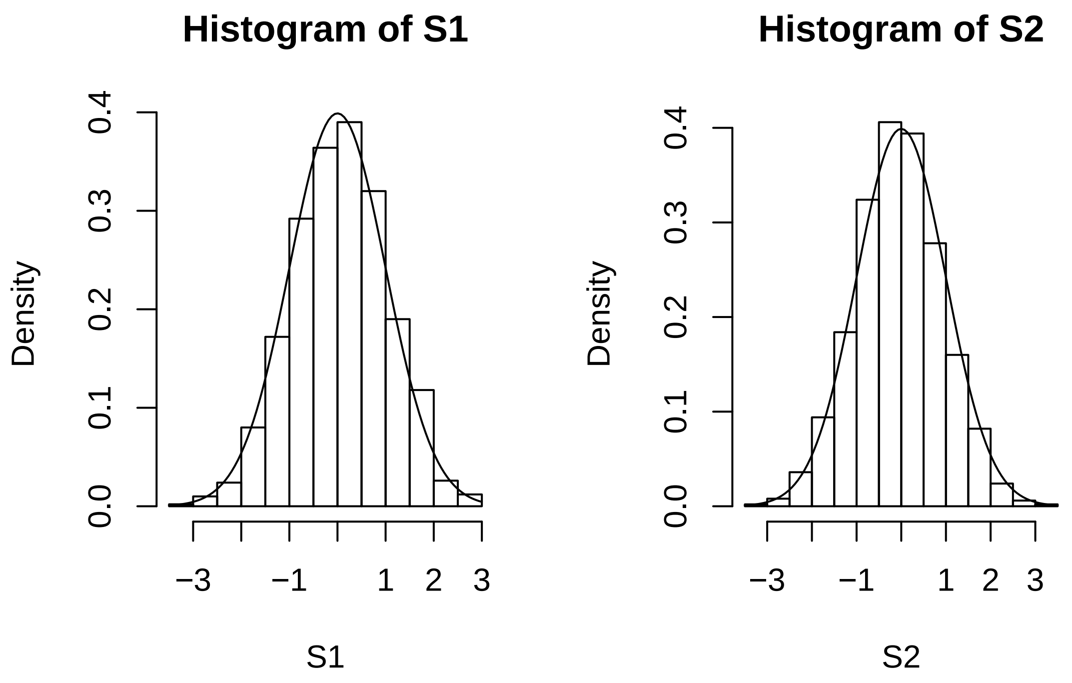
3. Recall that for n iid random variables $X_1, ..., X_n$ the standardised sum is given by:

$$S = \frac{\sum_{i=1}^{n} (X_i - \mathbb{E}[x_i])}{\sqrt{n\text{Var}[X_1]}}$$

(a) Generate 1000 realisations of the sum of n (where $n = 5$) from the uniform distirbution $unif(0, 1)$. Note that here, $\mathbb{E}[X_i] = \frac{1}{2}$ and that $\text{Var}[X_i] = \frac{1}{12}$. Then do the same for $n = 100$ and caclulate the standardised residuals for both.

```
set.seed(3023)
sim_number = 1000
n1 = 5
n2 = 100
mu = 1/2
sigma_squared = 1/12

# sample from the uniform distribution n1 times and repeat the experiment
# sim_number of times
temp1 = (runif(sim_number * n1) - mu)/sqrt(n1 * sigma_squared)
mat_temp1 = matrix(temp1, ncol = n1)
dim(mat_temp1)

[1] 1000    5

S1 = apply(mat_temp1, 1, sum)

temp2 = (runif(sim_number * n2) - mu)/sqrt(n2 * sigma_squared)
mat_temp2 = matrix(temp2, ncol = n2)
dim(mat_temp2)

[1] 1000  100

S2 = apply(mat_temp2, 1, sum)
```

```
par(mfrow = c(1, 2))
hist(S1, probability = TRUE)
curve(dnorm(x), add = TRUE)
hist(S2, probability = TRUE)
curve(dnorm(x), add = TRUE)
```


Histogram of S1
Density
0.0
0.1
0.2
0.3
0.4
−3
−1
1
2
3
S1
Histogram of S2
Density
0.0
0.1
0.2
0.3
0.4
−3
−1
1
2
3
S2