STAT 8010 R Lab 7: Sampling Distribution and CLT

Whitney Huang

September 23, 2020

Contents

Normal approximation of Binomial Distribution

```
n = 400; p = 0.93
sum(dbinom(370:373, n, p))
```

```
## [1] 0.3009909
```

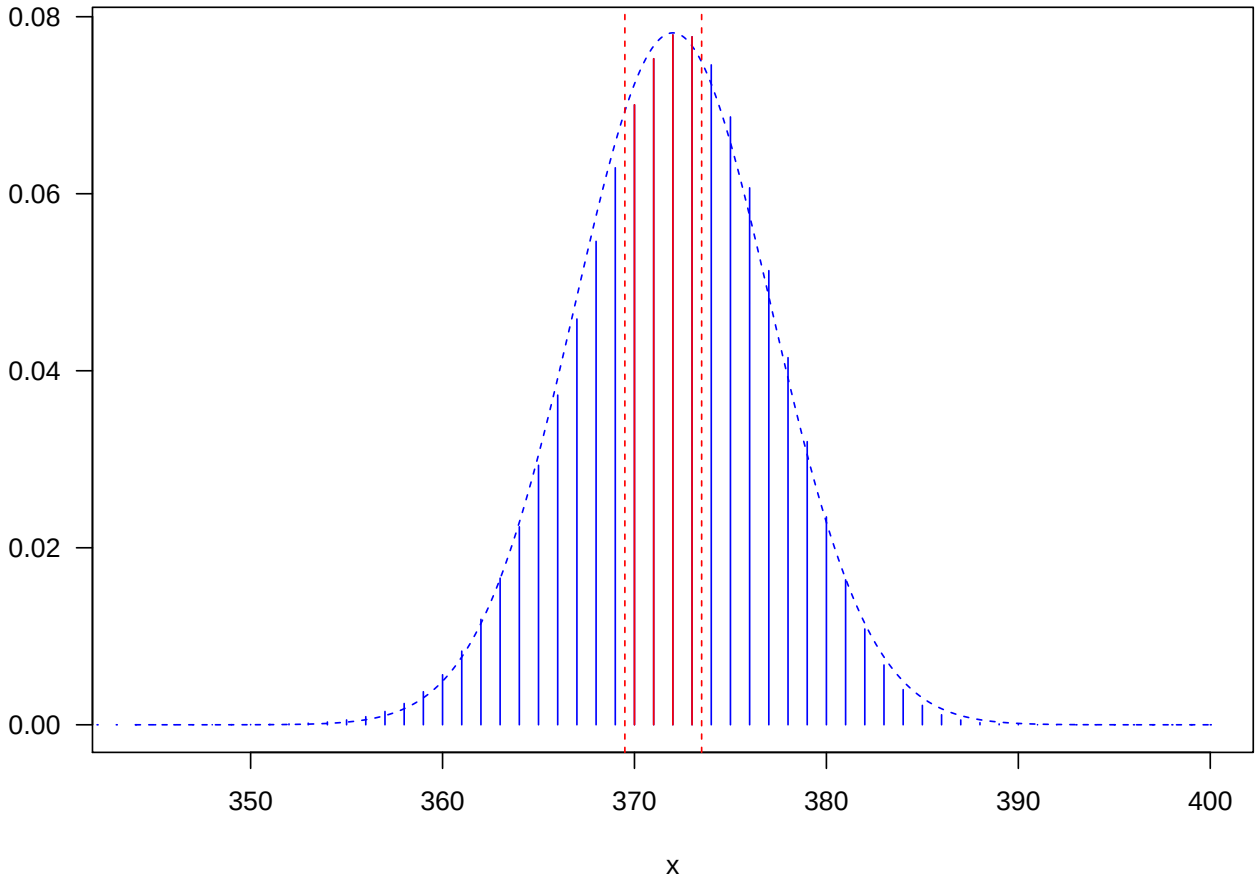
```
pbinom(373, n, p) - pbinom(369, n, p)
```

```
## [1] 0.3009909
```

```
mu = n * p; sigma = sqrt(n * p * (1 - p))
pnorm(373.5, mu, sigma) - pnorm(369.5, mu, sigma)
```

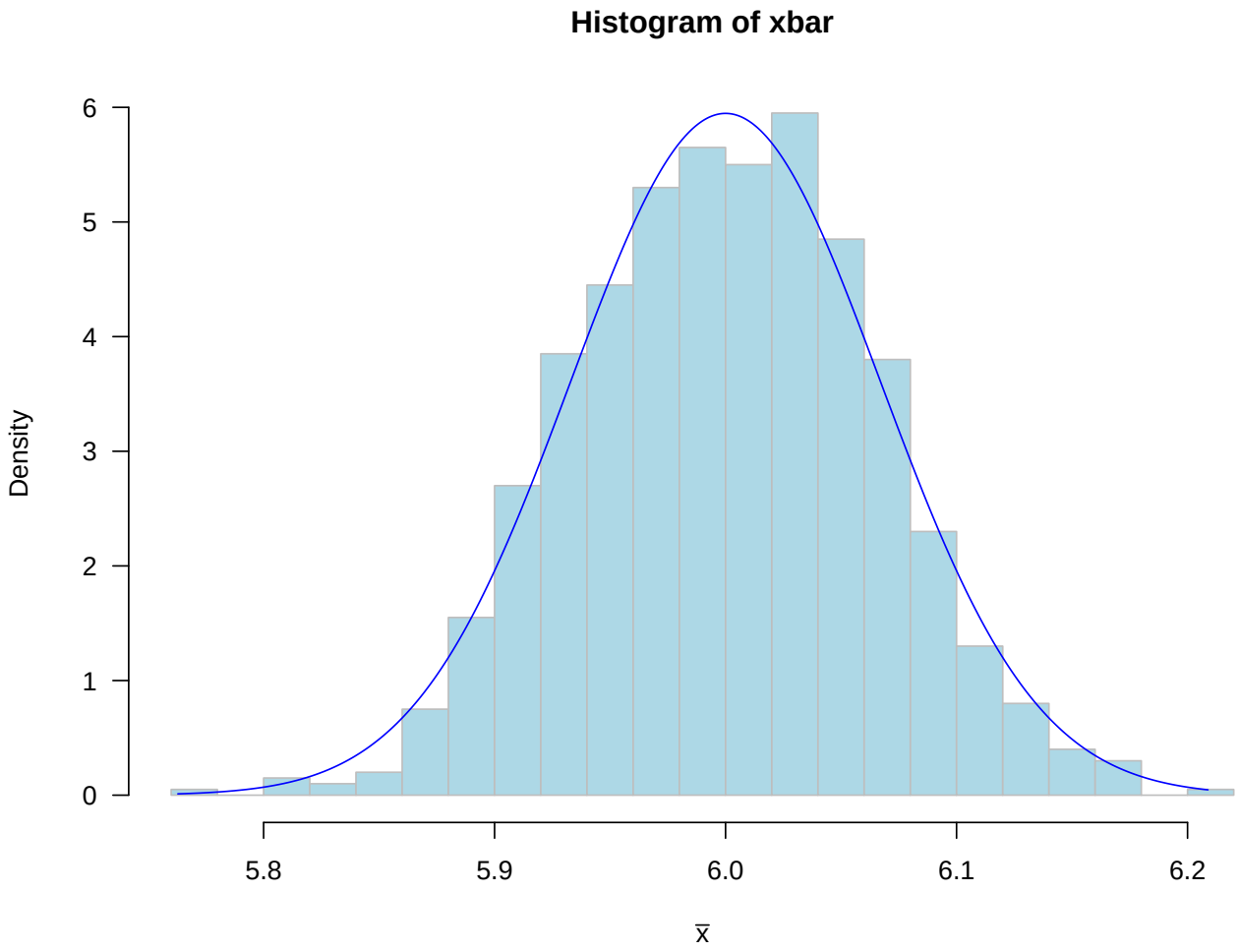
```
## [1] 0.3035037
```

```
plot(0:400, dbinom(0:400, n, p), type = "h", las = 1,
     xlim = c(344, 400), xlab = "x", ylab = "", col = "blue")
xg <- seq(344, 400, 0.1); yg <- dnorm(xg, mu, sigma)
lines(xg, yg, col = "blue", lty = 2)
abline(v = c(369.5, 373.5), lty = 2, col = "red")
lines(370:373, dbinom(370:373, n, p), type = "h", col = "red")
```

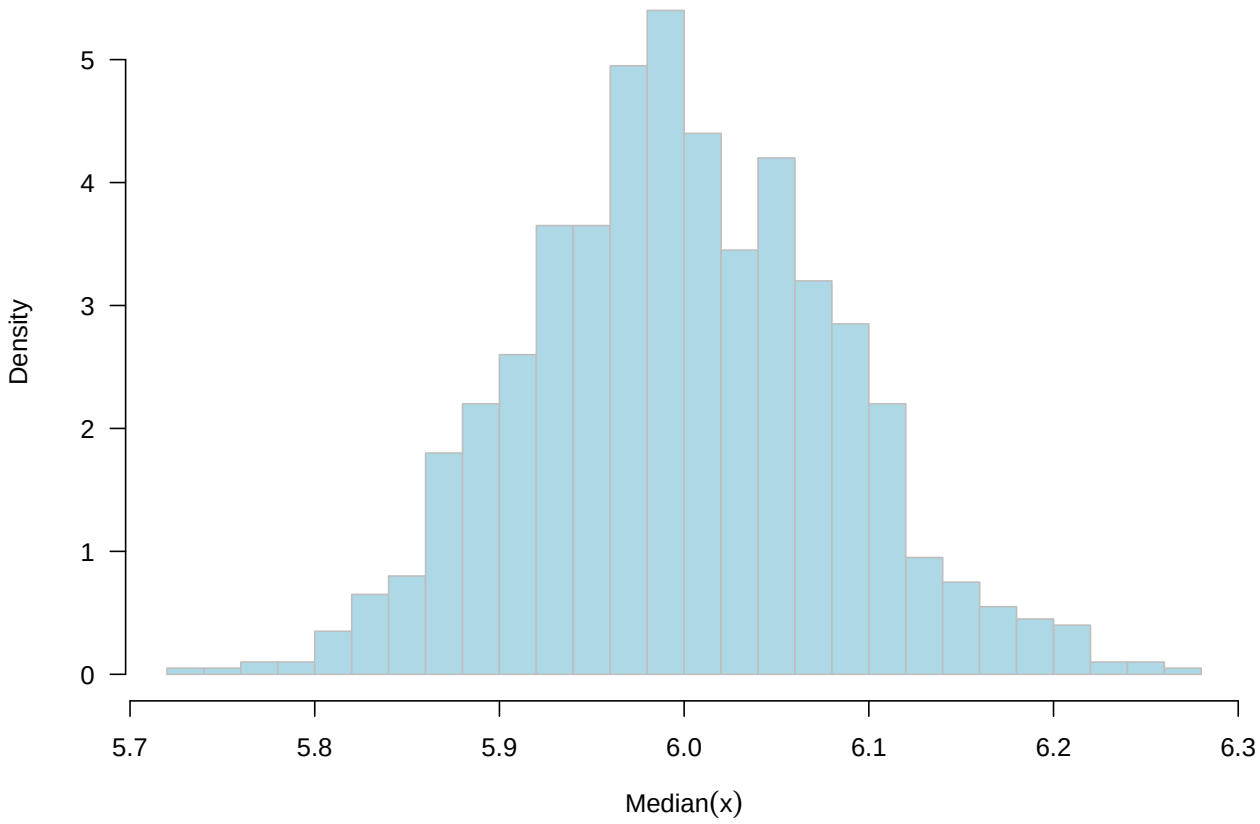


Sampling Distribution

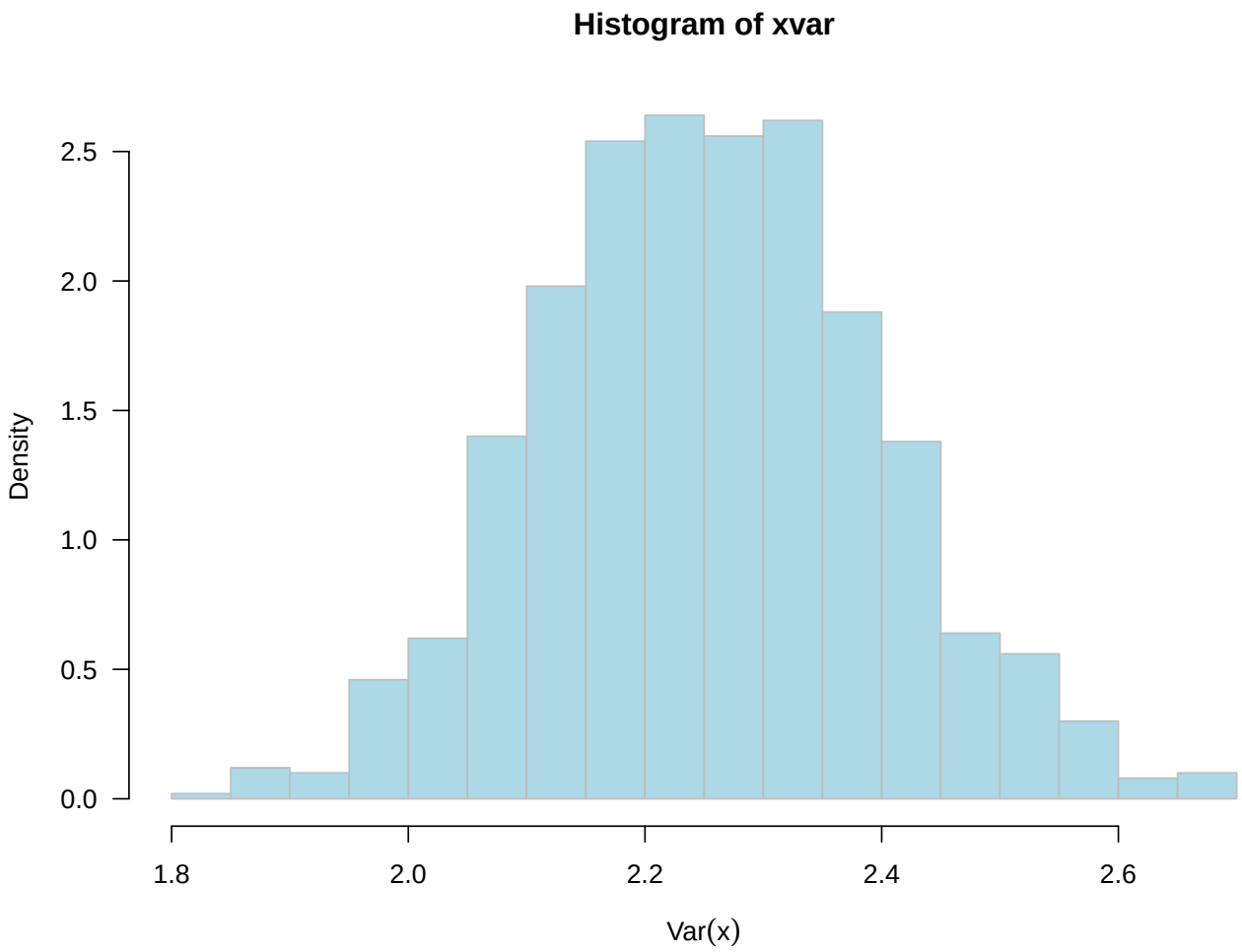
```
set.seed(123)
x <- replicate(1000, rnorm(500, 6, 1.5))
xbar <- apply(x, 2, mean)
hist(xbar, 30, col = "lightblue", border = "gray", prob = T,
     las = 1, xlab = expression(bar(x)))
xg <- seq(min(xbar), max(xbar), len = 1001); yg = dnorm(xg, 6, 1.5 / sqrt(500))
lines(xg, yg, col = "blue")
```




```
xmed <- apply(x, 2, median)
hist(xmed, 30, col = "lightblue", border = "gray", prob = T,
     las = 1, xlab = expression("Median"(x)))
```




```
xvar <- apply(x, 2, var)
hist(xvar, 30, col = "lightblue", border = "gray", prob = T,
     las = 1, xlab = expression("Var"(x)))
```



CLT

Animation

```
set.seed(123)
sim <- replicate(120, rexp(n = 100, 0.2))
library(animation)
saveLatex({
  for (i in 1:120){
    par(mfrow = c(1, 2), mar = c(5, 4, 1, 1), las = 1)
brk <- seq(0, max(sim), length.out = 25)
hist(sim[, i], breaks = brk, prob = T, border = "gray", main = " ",
xlab = "x", xlim = range(sim), ylim = c(0, 0.25), cex.lab = 1.25,
cex.axis = 1.25)
rug(sim[, i])

xg <- seq(0, range(sim)[2], length.out = 1000)
lines(xg, dexp(xg, 0.2), lwd = 1.5)
abline(v = mean(sim[, i]), col = "blue", lwd = 1.5)
x_bar_1 <- apply(sim, 2, mean)

library(MASS)
norm_mle1 <- fitdistr(x_bar_1[1:120], "normal")$estimate
```

```
hist(x_bar_1[1:120], 12, prob = T, border = "white",  main = " ",
     xlab = expression(bar(x)[100]), xlim = c(3.2, 6.8), ylim = c(0, 1),
     cex.lab = 1.25, cex.axis = 1.25)
rug(x_bar_1[1:i], col = "blue", lwd = 1.25)
xg1 <- seq(3.2, 6.8, length.out = 1000)

}
hist(x_bar_1[1:120], 12, prob = T, border = "gray",  main = " ",
     xlab = " ", xlim = c(3.2, 6.8), ylim = c(0, 1), add = T)
lines(xg1, dnorm(xg1, mean = norm_mle1[1], sd = norm_mle1[2]),
                 col = "blue", lty = 2, lwd = 1.8)
lines(xg1, 100 * dgamma(xg1 * 100, 100, 0.2), lwd = 1.5)
abline(v = mean(x_bar_1[1:120]), col = "blue")
abline(v = 5, col = "red")

}, img.name = "CLT", ani.opts = "controls,width=0.95\\textwidth",
latex.filename = ifelse(interactive(), "CTL_demos.tex", ""),
nmax = 120, ani.dev = "pdf", ani.type = "pdf", ani.width = 10,
interval = 0.3,
ani.height = 5,documentclass = paste("\\documentclass{article}",
                                     "\\usepackage[papersize={10in,5in},margin=0.1in]{geometry}",
                                     sep = "\n"))
```

Sample Size and the Normal Approximation

```
set.seed(123)
simExp <- replicate(5000, rexp(n = 5000, 0.2))

xbarEX5 <- apply(simExp[1:5,], 2, mean)
xbarEX30 <- apply(simExp[1:30,], 2, mean)
xbarEX100 <- apply(simExp[1:100,], 2, mean)
xbarEX500 <- apply(simExp[1:500,], 2, mean)
xbarEX5000 <- apply(simExp, 2, mean)


par(mfrow = c(2, 3), las = 1)
hist(xbarEX5, 24, prob = T, main = "", xlab = expression(bar(x)[5]),
     ylab = "Density", border = "gray", col = "lightblue", cex.lab = 1.25)
abline(v = 5, lwd = 1.5)
xgEX5 <- seq(0, 20, length.out = 1000)
lines(xgEX5, dnorm(xgEX5, mean = 5, sd = 5 / sqrt(5)), lty = 2, lwd = 1.3)
hist(xbarEX30, 24, prob = T, main = "Exponential", xlab = expression(bar(x)[30]),
     ylab = "Density", border = "gray", col = "lightblue", cex.lab = 1.25)
abline(v = 5, lwd = 1.5)
xgEX30 <- seq(2, 10, length.out = 1000)
lines(xgEX30, dnorm(xgEX30, mean = 5, sd = 5 / sqrt(30)), lty = 2, lwd = 1.3)
hist(xbarEX500, 24, prob = T, main = "", xlab = expression(bar(x)[500]),
     ylab = "Density", border = "gray", col = "lightblue", cex.lab = 1.25)
abline(v = 5, lwd = 1.5)
xgEX500 <- seq(4, 6, length.out = 1000)
lines(xgEX30, dnorm(xgEX30, mean = 5, sd = 5 / sqrt(500)), lty = 2, lwd = 1.3)

set.seed(123)
```

```r
simN <- replicate(5000, rnorm(n = 500))

xbarN5 <- apply(simN[1:5,], 2, mean)
xbarN30 <- apply(simN[1:30,], 2, mean)
xbarN500 <- apply(simN[1:500,], 2, mean)


hist(xbarN5, 24, prob = T, main = "", xlab = expression(bar(x)[5]),
     ylab = "Density", border = "gray", col = "lightblue", cex.lab = 1.25)
abline(v = 0, lwd = 1.5)
xgN5 <- seq(-2, 2, length.out = 1000)
lines(xgN5, dnorm(xgN5, mean = 0, sd = 1 / sqrt(5)), lty = 2, lwd = 1.3)
hist(xbarN30, 24, prob = T, main = "Normal", xlab = expression(bar(x)[30]),
     ylab = "Density", border = "gray", col = "lightblue", cex.lab = 1.25)
abline(v = 0, lwd = 1.5)
xgN30 <- seq(-0.7, 0.7, length.out = 1000)
lines(xgN30, dnorm(xgN30, mean = 0, sd = 1 / sqrt(30)), lty = 2, lwd = 1.3)
hist(xbarN500, 24, prob = T, main = "", xlab = expression(bar(x)[500]),
     ylab = "Density", border = "gray", col = "lightblue", cex.lab = 1.25)
abline(v = 0, lwd = 1.5)
xgN500 <- seq(-0.2, 0.2, length.out = 1000)
lines(xgN500, dnorm(xgN500, mean = 0, sd = 1 / sqrt(500)), lty = 2, lwd = 1.3)
```