**Key points**

* We can run Monte Carlo simulations to compare with theoretical results assuming a value of p.
* In practice, p is unknown. We can corroborate theoretical results by running Monte Carlo simulations with one or several values of p.
* One practical choice for p when modeling is X¯, the observed value of X^ in a sample.

**Code: Monte Carlo simulation using a set value of p**

p <- 0.45 # unknown p to estimate

N <- 1000

# simulate one poll of size N and determine x\_hat

x <- sample(c(0,1), size = N, replace = TRUE, prob = c(1-p, p))

x\_hat <- mean(x)

# simulate B polls of size N and determine average x\_hat

B <- 10000 # number of replicates

N <- 1000 # sample size per replicate

x\_hat <- replicate(B, {

x <- sample(c(0,1), size = N, replace = TRUE, prob = c(1-p, p))

mean(x)

})

**Code: Histogram and QQ-plot of Monte Carlo results**

library(tidyverse)

library(gridExtra)

p1 <- data.frame(x\_hat = x\_hat) %>%

ggplot(aes(x\_hat)) +

geom\_histogram(binwidth = 0.005, color = "black")

p2 <- data.frame(x\_hat = x\_hat) %>%

ggplot(aes(sample = x\_hat)) +

stat\_qq(dparams = list(mean = mean(x\_hat), sd = sd(x\_hat))) +

geom\_abline() +

ylab("X\_hat") +

xlab("Theoretical normal")

grid.arrange(p1, p2, nrow=1)