Exercise: Parametric bootstrapping for binomial data

Let x be the number of ”positives” from some experiment (e.g. classification in supervised learning), where the total number of responses was 100. We will assume that responses are i.i.d, and thus the appropriate statistical model is

$$x \sim Binom(100, p), \quad p \in (0, 1)$$

In the experiment, the observed x had the value 19.

- Estimate p from data. Write down the expression for the estimator.

For binomial data, one may construct an approximate $(1 - \alpha)$ confidence interval by[1]:

$$\hat{p} \pm \sqrt{\frac{p(1-p)}{n}} \cdot z_{1-\alpha/2} \tag{0.1}$$

where n is the number of data points and $z_{1-\alpha/2}$ is the $z_{1-\alpha/2}$ quantile in a normal distribution.

- Find an approximate 95% CI for p when $x = 19$ and $n = 100$ using Eq. (0.1).

Parametric bootstrap:

- Simulate a large number (at least 1000) of data where we use our estimate of p as input. For each replica, estimate p in the statistical model.
- Plot a histogram of $\hat{p}$ (ie. the estimates of p).
- Using our simulations, find a 95% param. bootstrap CI for p.
- Compare with the approximate CI that you found using formula (0.1).
- Repeat the exercise where $x = 25, 35, 50, 70, 80, 95$. Do you see a pattern?

Inspiration:

```
# Quantiles in a normal distribution:
qnorm( 0.99 ) # 99% quantile

# Simluate a large number from Binom(n,p) in one go:
rbinom(10, 100, 0.5)
# 10 observations from Binom(n,p) where n = 100, p = 0.5

# Histograms: use hist([...])
```

[1] This is the statndard textbook way of doing CI for binomial data