EXERCISE 1: Answers

This an exercise in the theory and application of the normal approximation when the underlying distribution is not normal. It also gives you an opportunity to refresh your memory on how to conduct a hypothesis test and construct a confidence interval.

1. Consider a Bernoulli random variable. Specifically \( X \in \{0, 1\} \) and \( P_r(x = 1) = \pi \) and \( P_r(x = 0) = 1 - \pi \).

   (a) Obtain the population mean and variance of distribution. \( E[X] = 1 \times P_r(x = 1) + 0 \times P_r(x = 0) = \pi \). \( E[(X - \pi)^2] = (1 - \pi)^2 \times P_r(x = 1) + (0 - \pi)^2 \times P_r(x = 0) = (1 - \pi)^2 \times \pi + \pi^2 \times (1 - \pi) = \pi - 2\pi^2 + \pi^3 + \pi^2 - \pi^3 = \pi - \pi^2 = \pi(1 - \pi) \).

   (b) Obtain the mean and variance of the sample mean \( \hat{\pi} = \overline{X}_n \) based on a sample size of \( n \). Assume an i.i.d. sample. \( E[\hat{\pi}] = E[\overline{X}_n] = E[\frac{1}{n}\sum_{i=1}^{n}X_i] = \frac{1}{n}\sum_{i=1}^{n}E[X_i] = \pi \). \( E[(\hat{\pi} - \pi)^2] = E[(\frac{1}{n}\sum_{i=1}^{n}(X_i - \pi))^2] = \frac{1}{n^2}\sum_{i=1}^{n}(X_i - \pi))^2 \) = \( \frac{\sigma^2}{n} \) where \( \sigma^2 \) is the variance of the underlying distribution.

   (c) Using the law of large numbers show that \( \lim_{n \to \infty} \hat{\pi} = \pi \). Khintchine’s theorem states that if an i.i.d. variable has a first moment, then its sample average converges in probability to the expectation as the sample size goes to infinity. Since \( X_i \) is i.i.d. and \( E[X_i] = \pi \) exists then \( \lim_{n \to \infty} \hat{\pi} = \pi \).

   (d) Using the central limit theorem obtain the limiting distribution of \( (\hat{\pi} - \pi)/(\pi(1 - \pi)/n)^{1/2} \). The Lindberg-Levy form of the central limit theorem states that if an i.i.d. variable has a first moment and second moment, then its \( \sqrt{n} \) times the sample average less the mean converges in distribution to a normal distribution with mean zero and the second moment as its variance. Since \( X_i \) is i.i.d., \( E[X_i] = \pi \), and \( \text{var}[X] = \pi(1 - \pi) \) then \( \sqrt{n}(\hat{\pi} - \pi) \to_d N(0, \pi(1 - \pi)) \).

   (e) What should be the limiting distribution of \( (\hat{\pi} - \pi)/(\pi(1 - \pi)/n)^{1/2} \). Rewrite \( (\hat{\pi} - \pi)/(\pi(1 - \pi)/n)^{1/2} = n^{1/2}(\hat{\pi} - \pi)/(\pi(1 - \pi))^{1/2} = \frac{n^{1/2}(\hat{\pi} - \pi)}{(\pi(1 - \pi))^{1/2}} = \frac{n^{1/2}(\hat{\pi} - \pi)}{(\pi(1 - \pi))^{1/2}} = \frac{n^{1/2}(\hat{\pi} - \pi)}{(\pi(1 - \pi))^{1/2}} \). From (d) the term in the first brackets converges to a standard normal, and by (c) and continuity of the function in the second bracket w.r.t. \( \hat{\pi} \) the second terms converges in probability to unity.

2. In a survey of 400 likely voters, 215 responded that they would vote for the incumbent and 185 responded that they would vote for the challenger. Let \( \pi \) denote the fraction of all likely voters that preferred the incumbent at the time of the survey, and let \( \hat{\pi} \) be the fraction of the survey respondents that preferred the incumbent.

   (a) Use the survey results to obtain a value for the estimator \( \hat{\pi} \). \( \hat{\pi} = \overline{X}_n = 215/400 = .5375 \).
(b) Based on your findings in (1), obtain an appropriate estimate of the variance of \( \hat{\pi} \). From above \( \text{var}[X]/n = \pi(1 - \pi)/n \) and an estimator is \( \hat{\pi}(1 - \hat{\pi})/n = .5375(1 - .5375)/400 = .0006215 \).

(c) For \( \alpha = .05 \), test the null \( H_0 : \pi = 0.5 \) vs. \( H_1 : \pi \neq 0.5 \). From (1.d) \( \sqrt{n}(\hat{\pi} - \pi) \rightarrow_d N(0, \pi(1 - \pi)) \) and \( \sqrt{n}(\hat{\pi} - \pi)/\sqrt{\pi(1 - \pi)} \rightarrow_d N(0,1) \) so \( 20 \times (\hat{\pi} - .5)/\sqrt{.5 \times .5} = 40 \times .0375 = 1.5 \) is a draw from \( N(0,1) \). A critical value for a .05 two-sided test is 1.96. Since our realization does not exceed this in absolute value we fail to reject. Alternatively, we could use \( \hat{\pi} \) in estimating the denominator of our ratio and obtain similar and asymptotically equivalent results.

(d) For \( \alpha = .05 \), test the null \( H_0 : \pi = 0.5 \) vs. \( H_1 : \pi > 0.5 \). The critical value for a .05 one-sided test is 1.645. Since our realization does not exceed this we fail to reject, although we are closer than with a two-sided test.

(e) Do the results from (2.c) and (2.d) agree? Why or why not?

3. Using the data in (2):

(a) Construct a 95% confidence interval for \( \pi \). Let \( z = \sqrt{n}(\hat{\pi} - \pi)/\sqrt{\pi(1 - \pi)} \) then \( \Pr[-1.96 \leq z \leq 1.96] = .95 \) asymptotically. Algebra yields \( \Pr[\hat{\pi} - 1.96\sqrt{\pi(1 - \pi)}/n \leq \pi \leq \hat{\pi} + 1.96\sqrt{\pi(1 - \pi)}/n] = .95 \). Thus \( .5375 \pm 1.96\sqrt{.25/400} = .5375 \pm .049 \) or \( (.4885,.5865) \) is a 95% interval.

(b) What does the interval tell us? For \( \alpha = .05 \) this yields the set of null hypotheses that will not reject.

(c) Construct a 99% confidence interval for \( \pi \). Using the same approach as in (3a), we obtain \( .5375 \pm 2.57\sqrt{.25/400} = .5375 \pm .06425 \) or \( (.4733,.6018) \) is a 99% interval.

(d) Why is the interval in (3.b) wider? The tail content is smaller at 1% rather than 5% so the interval it covers is wider.

(e) Using only the calculation from this problem, test the hypothesis \( H_0 : \pi = 0.5 \) vs. \( H_1 : \pi \neq 0.5 \) at the 5% significance level. From (3.a) a 95% interval is \( (.4885,.5865) \). Since 0.5 is included in this interval we do not reject \( H_0 : \pi = 0.5 \).