EXERCISE 3: Answers

This an exercise in the theory and application of the normal approximation when the underlying distribution is not normal.

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

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

   (b) Obtain the mean and variance of the sample mean $\bar{X} = \frac{\sum_{i=1}^{n} X_i}{n}$ based on a sample size of $n$. Assume an i.i.d. sample. $E[\bar{X}] = E[\frac{\sum_{i=1}^{n} X_i}{n}] = \frac{1}{n} \sum_{i=1}^{n} E[X_i] = \pi$. $E[(\bar{X} - \pi)^2] = E[(\frac{1}{n} \sum_{i=1}^{n} X_i - \pi)^2] = \frac{1}{n} \sum_{i=1}^{n} (X_i - \pi)^2 = \frac{\text{var}[X]}{n} = \pi(1 - \pi)/n$.

   (c) Using the law of large numbers show that $\text{plim}_{n \to \infty} \bar{X} = \pi$. 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 $\text{plim}_{n \to \infty} \bar{X} = \pi$.

   (d) Using the central limit theorem obtain the limiting distribution of $(\bar{X} - \pi)/(\pi(1 - \pi)/n)^{1/2}$. 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 $\sqrt{n}(\bar{X} - \pi) \to_{d} N(0, \pi(1 - \pi))$.

   (e) What should be the limiting distribution of $(\bar{X} - \pi)/(\pi(1 - \pi)/n)^{1/2}$. Rewrite $(\bar{X} - \pi)/(\pi(1 - \pi)/n)^{1/2} = n^{1/2}((\pi - \pi)/(\pi(1 - \pi)))^{1/2} = n^{1/2}((\pi - \pi)/(\pi(1 - \pi)))^{1/2}/((\pi - \pi)/(\pi(1 - \pi)))^{1/2}$.

   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. $\bar{X}$ the second terms converges in probability to unity.

2. We now consider maximum likelihood estimation of this problem.

   (a) Noting that some of the observations are zeros and others unity, write the joint probability of the $n$ observations. Use this probability to write the likelihood function for this problem. (Hint: the variable $x_i$ can be used as an indicator variable). The joint distribution is given by $f(y_1, y_2, ..., y_n; \pi) = L(\pi; y_1, y_2, ..., y_n) = \prod_{i=1}^{n} \pi^{y_i}(1 - \pi)^{(1 - y_i)}$ is the likelihood function and $\mathcal{L}(\pi; y_1, y_2, ..., y_n) = \sum_{i=1}^{n} [y_i \ln(\pi) + (1 - y_i) \ln(1 - \pi)]$ is the log-likelihood function.
(b) Obtain the maximum likelihood estimator of \( \pi \) for this problem and compare it to the estimator obtained above. The first-order condition is \( 0 = \partial \mathcal{L}(\pi; y_1, y_2, \ldots, y_n)/\partial \pi = \sum_{i=1}^{n} \frac{y_i(1/\pi) + (1 - y_i)(1/(1 - \pi))}{(1 - \pi)} \). Multiplying and dividing the first term by \( 1 - \pi \) and the second by \( \pi \) to obtain common denominators and then solving for \( \pi \) yields \( \hat{\pi} = \frac{1}{n} \sum_{i=1}^{n} y_i \) or the sample proportion as the maximum likelihood estimator of \( \pi \). This is just the sample average.

(c) Establish the large sample asymptotic properties of the maximum likelihood estimator such as asymptotic bias, consistency, and asymptotic normality. From before we showed that the sample average has mean \( \pi \) so is unbiased while the variance is \( \pi(1 - \pi)/n \), so we have asymptotic unbiasedness and convergence in quadratic mean and hence consistency. And the central limit theorem was used to show that \( \sqrt{n}(\hat{\pi} - \pi) \xrightarrow{d} N(0, \pi(1 - \pi)) \).

(d) Develop the Cramer-Rao bound for this problem. What can we then say about the efficiency properties of our estimator from (b)? The Cramer-Rao bound will be the inverse of \(-E[\partial^2 \ln f_i/\partial \pi^2]\) = \(E[-y_i \frac{1}{\pi^2} - (1 - y_i) \frac{1}{(1 - \pi)^2}]\). This is the covariance of the limiting distribution of \( \sqrt{n}(\hat{\pi} - \pi) \xrightarrow{d} N(0, \pi(1 - \pi)) \) so our estimator is efficient with the class of CUAN estimators and, since it is unbiased, it is efficient within the class of unbiased estimators.

(e) Suppose we are interested in testing the null hypothesis \( H_0 : \pi = \pi_0 \). How would we obtain the likelihood ratio statistic for this problem? What is its’ limiting distribution? From the text we have \( LR = 2[\mathcal{L}(\hat{\pi}) - \mathcal{L}(\pi_0)] = 2[\sum_{i=1}^{n} y_i \ln(\hat{\pi}/\pi_0) + (1 - y_i) \ln(\frac{1 - \hat{\pi}}{1 - \pi_0})] \) as the likelihood ratio for testing \( \pi = \pi_0 \). Asymptotically, this will have a limiting \( \chi^2 \) distribution.

3. 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 \( \hat{\pi} \). \( \hat{\pi} = \bar{X}_n = 215/400 = .5375 \).

(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) \( \sqrt{n}(\hat{\pi} - \pi) \xrightarrow{d} N(0, \pi(1 - \pi)) \) and \( \sqrt{n}(\hat{\pi} - \pi)/\sqrt{\pi(1 - \pi)} \xrightarrow{d} N(0, 1) \) so \( 20 \times (.5375 - .5)/\sqrt{.5 \times .5} = 40 \times .0375 = 1.5 \) 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?

4. 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 $0.5375 \pm 1.96\sqrt{0.25/400} = 0.5375 \pm 0.049$ or $(0.4885, 0.5865)$ 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 $0.5375 \pm 2.57\sqrt{0.25/400} = 0.5375 \pm 0.06425$ or $(0.4733, 0.6018)$ 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 $(0.4885, 0.5865)$. Since 0.5 is included in this interval we do not reject $H_0 : \pi = 0.5$. 

3