Homework 1

ESE 402/542

Due Wednesday, Sept. 22, 2021 at 11:59pm

Type or scan your answers as a single PDF file and submit on Canvas.

Tips:

- Don't worry too much about precision of numerical answers (3-4 decimal places should be enough).
- When asked to compute confidence intervals, Central Limit Theorem is your friend. Computing exact distributions of the sample mean can often be hard.
- Linearity of expectation is your friend.

Problem 1. Counting review.

(a) The game of Mastermind starts in the following way: One player selects 4 pegs, each peg having 6 possible colors, and places them in a line. The second player then tries to guess the sequence of colors. What is the probability of guessing correctly?

(b) A child has 6 blocks, 3 of which are red and 3 of which are green. How many patterns can she make by placing them all in a line? If she is given three white blocks, how many total patterns can she make by placing all nine blocks in a line?

Problem 2. A random variable X is distributed according to $Gamma(\alpha, \beta)$. Recall that the pdf of a gamma random variable is given to be $f(x|\alpha, \beta) = \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-x/\beta}$. Also recall that $\Gamma(\alpha + 1) = \alpha\Gamma(\alpha)$ for $\alpha > 0$.

(a) Compute $\mathbf{E}[X]$

(b) Compute $\mathbf{Var}[X]$

(c) Find α and β such that X is a chi-squared random variable with p degrees of freedom. Recall that if $X = \sum_{i=1}^{p} z_i^2, z_i \sim N(0, 1)$, then $X \sim \chi_p^2$.

Problem 3. Find the moment-generating function of a random variable $X \sim Bernoulli(p)$, and use the MGF you computed to find its mean, variance, and third moment $\mathbf{E}[X^3]$. Recall that pmf of Bernoulli distribution is $f(x) = p^x(1 - p)^{1-x}$.

Problem 4. Suppose that a company ships packages that are variable in weight, with an average weight of 30 lb and a standard deviation of 15. Assuming that the packages come from a large number of different customers so that it is reasonable to model their weights as independent random variables, find (approximately) the probability that 100 packages will have a total weight exceeding 3500 lb.

Problem 5. A particular area contains 8,000 condominium units. In a survey of the occupants with sample size 100, 12% of the respondents said they planned to sell their condos within the next year;

(a) Compute the 95% confidence interval for the estimated probability of people planning to sell. (Hint: The variance of the Bernoulli distribution with parameter p is given by $p(1-p)$).

(b) Suppose that another survey is done of another condominium project of 12,000 units. The sample size is 200, and the proportion planning to sell in this sample is .18. What is the standard error of this estimate? Give a 90% confidence interval.

(c) Suppose we use the notation $\hat{p}_1 = .12$ and $\hat{p}_2 = .18$ to refer to the proportions in the two samples. Let $\hat{d} = \hat{p}_1 - \hat{p}_2$ be an estimate of the difference, d, of the two population proportions p_1 and p_2. Using the fact that $\hat{p}_1$ and $\hat{p}_2$ are independent random variables, calculate the variance and standard error of $\hat{d}$.

(d) Because $\hat{p}_1$ and $\hat{p}_2$ are approximately normally distributed, so is $\hat{d}$. Use this fact to construct 99%, 95%, and 90% confidence intervals for d. Is there clear evidence that p_1 is really different from p_2?

Problem 6. Let $\hat{\theta}$ be an estimator for a scalar parameter θ of a parametric model. The mean squared error of θ is define as

$$\mathrm{MSE}(\hat{\theta}, \theta) = \mathbb{E}[|\hat{\theta} - \theta|^2]. \tag{1}$$

Show that

$$\mathrm{MSE}(\hat{\theta}, \theta) = (\mathrm{bias}(\hat{\theta}, \theta))^2 + \mathrm{Var}(\hat{\theta}) \tag{2}$$

where $\mathrm{bias}(\hat{\theta}, \theta)$ is the the difference between the average of the estimate and the true parameter; i.e., $\mathbb{E}[\hat{\theta}] - \theta$.

Problem 7. There are $p = 25\%$ of total population that will vote for candidate A in a coming election. In a survey of sample size to be $n = 100$, $\hat{p}$ of interviewees said they would vote for A. Assume that the votes of interviewees are i.i.d.

(a) Find δ such that $P(|\hat{p} - p| \geq \delta) \approx 0.025$.

(b) If, in the sample, $\hat{p} = 0.25$, will the 95% confidence interval for p contain the true value of p?