Problem Section 5

KEY

Monday Feb 5 2024

Learning Outcomes

The problems are designed to build conceptual understanding and problem-solving skills. The emphasis is on learning to find, evaluate and build confidence. The specific tasks include:

- Calculate probabilities using the Central Limit Theorem.
- Construct a confidence interval using Chebychev's inequality
- Calculate the sample size for a desired margin of error
- Back up and support work with relevant explanations

Exercises

1. To utilize the Central Limit Theorem for the sum of independent random variables, we must find two quantities: the expectation and variance.

   We have $E[X] = 0 \times .3 + 1 \times .5 + 2 \times .2 = .9$.

   To find the variance, let us first find the second moment, $E[X^2]$.

   $E[X^2] = 0^2 \times .3 + 1^2 \times .5 + 2^2 \times .2 = 1.3$

   So $Var[X] = 1.3 - (.9)^2 = .49$

   By the CLT for sums we know for i.i.d. RV's $X_1, \ldots, X_{400}$:

$$S = X_1 + X_2 + \cdots + X_{400} \sim N(\mu = n \times E[X], \sigma = \sqrt{n \times Var[X]})) = N(360, 14)$$

To approximate the probability that the grocer bought enough half-gallons of milk we may find:

$$P(S \leq 390) = 0.9839377$$

Thus the probability the grocer ordered enough using the CLT is 0.9839377.

2. We want to find n so that the half length of a 90% (large sample) confidence interval for the mean width is less than or equal to 0.1 mm. The critical value for a 90% confidence interval is 1.645. Therefore we want to solve the equation

$$1.645 \times \frac{\sigma_0}{\sqrt{n}} \leq 0.1,$$

$$\begin{aligned} n &\geq 1.645^2 \times \frac{0.4^2}{0.1^2}, \\ &\geq 43.296 \end{aligned}$$

Therefore, we need to measure $n = 44$ toothpicks.

To verify that the half length of the 90% confidence interval for the mean width is no larger than 0.1mm for this value of n, we can do the following calculation.

```r
1.645 * 0.4/sqrt(44)
```

```
## [1] 0.09919723
```

3. To find a confidence interval for μ_0 that has at least 95% confidence we must find L, U s.t.

$$P(L \leq \mu_0 \leq U) \geq .95$$

Per the hint, let us proceed with Chebyshev's inequality.

Define the RV $\bar{X} = \frac{1}{n} \sum_i^n X_i$.

We have by Thm 18.1, $E[\bar{X}] = E[X] = \mu_0$ and $Var(\bar{X}) = \frac{Var(X)}{n} = \frac{\sigma_0^2}{n}$.

Thus by Chebyshev's Inequatility:

$$P(|\bar{X} - \mu_0| \geq k\frac{\sigma_0}{\sqrt{n}}) \leq \frac{1}{k^2}$$

We see that this is almost in the form of the CI, but we wish for a $\geq$ on the outside of the probability. Let us thus reverse this inequality by multiplying by -1 and subtracting by 1 on each side.

$$1 - P(|\bar{X} - \mu_0| \geq k\frac{\sigma_0}{\sqrt{n}}) = P(|\bar{X} - \mu_0| \leq k\frac{\sigma_0}{\sqrt{n}}) \geq 1 - \frac{1}{k^2}$$

Now we have the term $1 - \frac{1}{k^2}$ that will lower bound the probability. We can adjust this lower bound, by adjusting k. For now, let us leave it in terms of k and we can simplify once we have a better idea of the term inside the probability.

Remember, we wish to find a probability of the form: $P(L \leq \mu_0 \leq U)$. We have:

$$P(|\bar{X} - \mu_0| \leq k\frac{\sigma_0}{\sqrt{n}}) = P(-k\frac{\sigma_0}{\sqrt{n}} \leq \bar{X} - \mu_0 \leq k\frac{\sigma_0}{\sqrt{n}})$$

This is very close to the desired form, save for an $\bar{X} - \mu_0$ term in the middle, rather than just μ_0.

Multiplying the inequality by negative 1, and then subtracting the $\bar{X}$ on both sides yields:

$$-k\frac{\sigma_0}{\sqrt{n}} \leq \bar{X} - \mu_0 \leq k\frac{\sigma_0}{\sqrt{n}} \Rightarrow \bar{X} - k\frac{\sigma_0}{\sqrt{n}} \leq \mu_0 \leq \bar{X} + k\frac{\sigma_0}{\sqrt{n}}$$

Thus we have that:

$$P(\bar{X} - k\frac{\sigma_0}{\sqrt{n}} \leq \mu_0 \leq \bar{X} + k\frac{\sigma_0}{\sqrt{n}}) \geq 1 - \frac{1}{k^2}$$

This exactly in the form we want! We have $L = \bar{X} - k\frac{\sigma_0}{\sqrt{n}}$, $U = \bar{X} + k\frac{\sigma_0}{\sqrt{n}}$, and our lower bound confidence level as $1 - \frac{1}{k^2}$. We assume n and σ_0^2 are known, so we must simply find the value of k such that $1 - \frac{1}{k^2} = .95$.

$$
\begin{aligned}
1 - \frac{1}{k^2} &= .95 \\
\frac{1}{k^2} &= .05 \\
k^2 &= \frac{1}{.05} \\
k &= 4.472136
\end{aligned}
$$

Thus for k = 4.472136,

$$P(\bar{X} - k\frac{\sigma_0}{\sqrt{n}} \leq \mu_0 \leq \bar{X} + k\frac{\sigma_0}{\sqrt{n}}) \geq .95$$

That is, Chebychev's inequality says we need to add about 4.5 standard errors around $\bar{X}$ in order to get at least 95% confidence. This is much larger than 1.96 which is what the Central Limit Theorem (CLT) tells us to do. The difference is that Chebychev is a result that applies to *any* random variable regardless of distribution. The CLT is focused on the case when the distribution is normal.