4 Confidence Interval Introduction

(a) $\frac{\sigma^2}{\epsilon^2}$

Since $\sigma = \sqrt{\text{var}(X)}$, so $\text{var}(X) = \sigma^2$. Thus, using Chebyshev's Inequality, we have that the upper bound would be:

$$\mathbb{P}[|X - \mu| \geq \epsilon] \ \leq \ \frac{\text{var}(X)}{\epsilon^2} = \frac{\sigma^2}{\epsilon^2}$$

(b) Direct Proof

Since the event $|X - \mu| < \epsilon$ is equivalent to the event $-\epsilon < X - \mu < \epsilon$, which is then equivalent to the event $X - \epsilon < \mu < X + \epsilon$ by properties of inequalities, which is then equivalent to the event $\mu \in (X - \epsilon, X + \epsilon)$ by definition, so the event space of $|X - \mu| < \epsilon$ is the same as the event space of $\mu \in (X - \epsilon, X + \epsilon)$, and they have the same sample space.

Thus, $\mathbb{P}[|X - \mu| < \epsilon] \ = \ \mathbb{P}\big\{\mu \in (X - \epsilon, X + \epsilon)\big\}$, as desired.

Q.E.D.

(c) $\epsilon = 2\sqrt{5}\sigma$

We wish to have $\mathbb{P}\big\{\mu \in (X - \epsilon, X + \epsilon)\big\} \geq 95\%$. Using our results from parts (a) and (b), so we have that $\mathbb{P}\big\{\mu \in (X - \epsilon, X + \epsilon)\big\} = \mathbb{P}[|X - \mu| < \epsilon] = \mathbb{P}[\overline{|X - \mu| \geq \epsilon}] = 1 - \mathbb{P}[|X - \mu| \geq \epsilon] \geq 1 - \frac{\sigma^2}{\epsilon^2}$. This means that if we can choose ϵ such that $1 - \frac{\sigma^2}{\epsilon^2} \geq 95\%$, then we can guarantee that

$$\mathbb{P}\big\{\mu \in (X - \epsilon, X + \epsilon)\big\} \ \geq \ 1 - \frac{\sigma^2}{\epsilon^2} \ \geq \ 95\%$$

Thus, we can calculate that for $1 - \frac{\sigma^2}{\epsilon^2} \geq 95\%$, so we need:

$$\epsilon^2 \geq 20\sigma^2$$

$$\implies \epsilon \geq 2\sqrt{5}\sigma$$

(d) $\mathbb{E}[\overline{X}] = \mu$, $\text{var}(\overline{X}) = \frac{\sigma^2}{n}$

Since we're given that $n \in \mathbb{Z}^+$ is a constant, and that μ is the mean for X (i.e. $\mu = \mathbb{E}[X]$), and that $X_1, \ldots, X_n$ are i.i.d. samples, as well as that $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$, so we can utilize Theorem 15.1 to get that:

$$\mathbb{E}[\overline{X}] = \mathbb{E}[\frac{1}{n}\sum_{i=1}^{n} X_i] = \frac{1}{n} \cdot \mathbb{E}[X_1 + \cdots + X_n] = \frac{1}{n} \cdot (\mathbb{E}[X_1] + \cdots + \mathbb{E}[X_n]) = \frac{1}{n} \cdot (\mu + \cdots + \mu)$$

$$\implies \mathbb{E}[\overline{X}] = \frac{1}{n} \cdot (n\mu) = \mu$$

And also, using Theorem 16.3 and a result from Note 16 we have:

$$\text{var}(\overline{X}) = \text{var}(\frac{1}{n}\sum_{i=1}^{n} X_i) = (\frac{1}{n})^2 \cdot \text{var}(X_1 + \cdots + X_n) = \frac{1}{n^2} \cdot \big(\text{var}(X_1) + \cdots + \text{var}(X_n)\big)$$

Then, since $\text{var}(X) = \sigma^2$ by definition, so:

$$\text{var}(\overline{X}) = \frac{1}{n^2} \cdot (\sigma^2 + \cdots + \sigma^2) = \frac{1}{n^2} \cdot (n\sigma^2) = \frac{\sigma^2}{n}$$

(e) $\epsilon = \sqrt{\frac{20\sigma^2}{n}}$

We can repeat the process of parts (a) to (c) to choose a proper width ϵ of the confidence interval.

First, denoting the mean of $\overline{X}$ as $\mathbb{E}[\overline{X}] = \nu$, and we calculate an upper bound on $\mathbb{P}\big[|\overline{X} - \nu| \geq \epsilon\big]$, which, using Chebyshev's Inequality, is:

$$\mathbb{P}\big[|\overline{X} - \nu| \geq \epsilon\big] \ \leq \ \frac{\text{var}(\overline{X})}{\epsilon^2} = \frac{\frac{\sigma^2}{n}}{\epsilon^2} = \frac{\sigma^2}{n \cdot \epsilon^2}$$

Now, since the event $|\overline{X} - \nu| < \epsilon$ is equivalent to the event $-\epsilon < \overline{X} - \nu < \epsilon$, which is then equivalent to the event $\overline{X} - \epsilon < \nu < \overline{X} + \epsilon$ by properties of inequalities, which is then equivalent to the event of $\nu \in (\overline{X} - \epsilon, \overline{X} + \epsilon)$ by definition, so the event space of $|\overline{X} - \nu| < \epsilon$ is the same as the event space of $\nu \in (\overline{X} - \epsilon, \overline{X} + \epsilon)$. Thus, $\mathbb{P}[|\overline{X} - \nu| < \epsilon] \ = \ \mathbb{P}\big\{\nu \in (\overline{X} - \epsilon, \overline{X} + \epsilon)\big\}$.

Then, $\mathbb{P}\big\{\nu \in (\overline{X} - \epsilon, \overline{X} + \epsilon)\big\} \ = \mathbb{P}[|\overline{X} - \nu| < \epsilon] = \mathbb{P}[\overline{|\overline{X} - \nu| \geq \epsilon}] = 1 - \mathbb{P}[|\overline{X} - \nu| \geq \epsilon]$. Since $\mathbb{P}\big[|\overline{X} - \nu| \geq \epsilon\big] \leq \frac{\sigma^2}{n \cdot \epsilon^2}$, so $\mathbb{P}\big\{\nu \in (\overline{X} - \epsilon, \overline{X} + \epsilon)\big\} = 1 - \mathbb{P}[|\overline{X} - \nu| \geq \epsilon] \ \geq \ 1 - \frac{\sigma^2}{n \cdot \epsilon^2}$.

Since we wish to have $\mathbb{P}\big\{\nu \in (\overline{X} - \epsilon, \overline{X} + \epsilon)\big\} \geq 95\%$, so if $1 - \frac{\sigma^2}{n \cdot \epsilon^2} \geq 95\%$, then we can guarantee our desired result. With σ being known, so we can calculate:

$$1 - \frac{\sigma^2}{n \cdot \epsilon^2} \geq 95\%$$

$$\implies \frac{\sigma^2}{n \cdot \epsilon^2} \leq 0.05$$

$$\implies 0.05\epsilon^2 \geq \frac{\sigma^2}{n}$$

$$\implies \epsilon^2 \geq \frac{20\sigma^2}{n}$$

$$\implies \epsilon \geq \sqrt{\frac{20\sigma^2}{n}}$$

Thus, $\epsilon = \sqrt{\frac{20\sigma^2}{n}}$ is an appropriate width of the confidence interval for the desired result, i.e. guaranteeing $\mathbb{P}\big\{\nu \in (\overline{X} - \epsilon, \overline{X} + \epsilon)\big\} \geq 95\%$.

(Confirmed: as n increases, ϵ decreases.)