Unbiased estimation for some standard problems.

**Problem 1** Suppose \( Y_1, Y_2, \ldots, Y_n \) is a random sample from a population with mean \( \mu \) and variance \( \sigma^2 \) (both unknown).

Find unbiased estimators of \( \mu \) and \( \sigma^2 \).

Again, note that \( Y_1, Y_2, \ldots, Y_n \) is a "random sample from a population" if means that \( Y_1, Y_2, \ldots, Y_n \) are independent random variables, and have a common distribution (unspecified here).

We know that the mean of this common distribution is \( \mu \) and the variance of this common distribution is \( \sigma^2 \), i.e.,

\[
E[Y_i] = \mu \quad \text{and} \quad V(Y_i) = \sigma^2
\]

for every \( i = 1, 2, \ldots, n \). The most obvious estimator for \( \mu \) is

\[
\hat{\mu} = \frac{1}{n} \sum_{i=1}^{n} Y_i = \bar{Y}.
\]

Note that

\[
E[\hat{\mu}] = E\left[\frac{1}{n} \sum_{i=1}^{n} Y_i\right] = \frac{1}{n} \sum_{i=1}^{n} E[Y_i] = \mu.
\]
Hence, \( \hat{\mu} \) is an unbiased estimator of \( \mu \).

**Definition:** The standard error of an estimator \( \hat{\theta} \) of a parameter \( \theta \) is defined as

\[
SE(\hat{\theta}) = \sqrt{E[(\hat{\theta} - \theta)^2]} = \sqrt{\text{MSE}(\hat{\theta})}.
\]

The standard error of \( \hat{\mu} \) is given by

\[
SE(\hat{\mu}) = \sqrt{\text{MSE}(\hat{\mu})} = \sqrt{\text{V}(\hat{\mu})} \quad (\text{because } \hat{\mu} \text{ is unbiased})
\]

\[
= \sqrt{\frac{1}{n} \sum_{i=1}^{n} \text{V}(Y_i)}
\]

\[
= \sqrt{\frac{1}{n^2} \sum_{i=1}^{n} \text{V}(Y_i)}
\]

\[
= \frac{\sigma}{\sqrt{n}}.
\]
Now let us find an unbiased estimator for $\sigma^2$.

Note that $\sigma^2$ is the population variance, i.e.,

$$\sigma^2 = \text{Average squared deviation from the mean in the population}$$

The only information we have is the sample data $Y_1, Y_2, \ldots, Y_n$. Hence, the most obvious strategy is the use sample quantities in place of population quantities for our procedure. Hence

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^{n} (Y_i - \overline{Y})^2 = \text{Average squared deviation from the sample mean $\overline{Y}$ in the sample.}$$

Let us calculate $E[\hat{\sigma}^2]$.

$$E[\hat{\sigma}^2] = E\left[\frac{1}{n} \sum_{i=1}^{n} (Y_i - \overline{Y})^2\right]$$

$$= E\left[\frac{1}{n} \sum_{i=1}^{n} (Y_i^2 - 2Y_i \overline{Y} + \overline{Y}^2)\right]$$

$$= E\left[\frac{1}{n} \sum_{i=1}^{n} Y_i^2 - 2\overline{Y} \sum_{i=1}^{n} Y_i + \overline{Y}^2\right]$$

$$= \frac{1}{n} \sum_{i=1}^{n} E[Y_i^2] - E[\overline{Y}^2]$$

$$= (\mu^2 + \sigma^2) - E[\overline{Y}^2] \quad (\text{Why?})$$

Let us concentrate on evaluating $E[\overline{Y}^2]$ and then substitute it back into the expression above.
\[
E[\bar{Y}^2] = E\left[\left(\frac{\sum_{i=1}^{n} Y_i}{n}\right)^2\right]
\]
\[
= \frac{1}{n^2} E\left[\sum_{i=1}^{n} Y_i^2 + \sum_{1 \leq i \neq j \leq n} Y_i Y_j\right]
\]
\[
= \frac{1}{n^2} \sum_{i=1}^{n} E[Y_i^2] + \frac{1}{n^2} \sum_{1 \leq i \neq j \leq n} E[Y_i Y_j]
\]
\[
= \frac{(\mu^2 + \sigma^2)}{n} + \frac{1}{n^2} \sum_{1 \leq i \neq j \leq n} \mu^2
\]
\[
= \frac{(\mu^2 + \sigma^2)}{n} + \frac{n(n-1)}{n^2} \mu^2
\]
\[
= \mu^2 + \frac{\sigma^2}{n}
\]
Here, \( E[\bar{Y}^2] = (\mu^2 + \sigma^2) - (\mu^2 + \frac{\sigma^2}{n}) = (n-1) \frac{\sigma^2}{n} \). This means \( \hat{\sigma}^2 \) is not unbiased for \( \sigma^2 \). However, a slight modification does the job.

Let \( s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (Y_i - \bar{Y})^2 = \frac{n}{n-1} \hat{\sigma}^2 \). Then,

\[
E[s^2] = \frac{n}{n-1} E[\hat{\sigma}^2] = \frac{n}{n-1} \times \frac{n-1}{n} \sigma^2 = \sigma^2.
\]