Exercise 5: Sample Statistics (2 pts)

1. (1 pt) Suppose there is a dataset $x_1, \ldots, x_n$ sampled from a distribution with mean μ and variance σ^2. Compute the expected value of the sample mean: $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$. Describe any modifications that might be required to make the expected value μ (recall that μ and σ^2 are unknown).
   Ans: **Answer 5.1**
2. (1 pt) Suppose there is a dataset $x_1, \ldots, x_n$ sampled from a distribution with mean μ and variance σ^2. Compute the expected value of the sample variance: $\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$, where $\bar{x}$ is the sample mean from the previous part. Describe any modifications that might be required to make the expected value σ^2 (recall that μ and σ^2 are unknown).
   Ans: **Answer 5.2**

Answer 5.1

$$E[\bar{x}] = E\left[\frac{1}{n}\sum_{i=1}^{n} x_i\right] \tag{90}$$

$$= \frac{1}{n}\sum_{i=1}^{n} E[x_i] \tag{91}$$

$$\because \quad E[x_i] = \mu \forall i \tag{92}$$

$$= \frac{1}{n}\sum_{i=1}^{n} \mu \tag{93}$$

$$= \frac{1}{n} \cdot n\mu \tag{94}$$

$$= \mu \tag{95}$$

Hence:

$$\mu = E[\bar{x}] = E\left[\frac{1}{n}\sum_{i=1}^{n} x_i\right] \tag{96}$$

Hence, the expected value of the sample mean is μ. No additional modification is required.

Q.E.D.

Answer 5.2

$$E\left[\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2\right] = E\left[\frac{1}{n}\sum_{i=1}^{n}(x_i^2-2x_i\bar{x}+\bar{x}^2)\right] \tag{97}$$

$$= E\left[\frac{1}{n}(\sum_{i=1}^{n}x_i^2-2\sum_{i=1}^{n}x_i\bar{x}+\sum_{i=1}^{n}\bar{x}^2)\right] \tag{98}$$

$$\because \quad \sum_{i=1}^{n}x_i = n\bar{x} \leftarrow \text{ as stated in Answer 5.1} \tag{99}$$

$$= E\left[\frac{1}{n}(\sum_{i=1}^{n}x_i^2-2n\bar{x}\bar{x}+n\bar{x}^2)\right] \tag{100}$$

$$= E\left[\frac{1}{n}(\sum_{i=1}^{n}x_i^2-n\bar{x}^2)\right] \tag{101}$$

$$= \frac{1}{n}(\sum_{i=1}^{n}E\left[x_i^2\right]-nE\left[\bar{x}^2\right]) \tag{102}$$

$$\because \quad \text{Assume it is independent} \tag{103}$$

$$\therefore \quad E[x_i^2] = \sigma^2+\mu^2 \quad E[\bar{x}^2] = \frac{\sigma^2}{n}+\mu^2 \tag{104}$$

$$= \frac{1}{n}\left(\sum_{i=1}^{n}(\sigma^2+\mu^2)-n(\frac{\sigma^2}{n}+\mu^2)\right) \tag{105}$$

$$= \frac{1}{n}\left(n(\sigma^2+\mu^2)-n(\frac{\sigma^2}{n}+\mu^2)\right) \tag{106}$$

$$= \frac{1}{n}\left(n(\sigma^2+\mu^2-\frac{\sigma^2}{n}-\mu^2)\right) \tag{107}$$

$$= \frac{n-1}{n}\sigma^2 \tag{108}$$

$$\tag{109}$$

Hence, we need modification:

$$\sigma^2 = \frac{n}{n-1}E\left[\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2\right] \tag{110}$$

The expected variance σ^2 shall be multiplying $\frac{n}{n-1}$ with the sample variance.

In the special case of $n\to\infty$ (or simply large enough), we may assume $\sigma^2 = \lim_{n\to\infty}\frac{n}{n-1}E\left[\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2\right] = E\left[\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2\right]$.

Q.E.D.