Part IA Engineering, Sample Statistics Questions

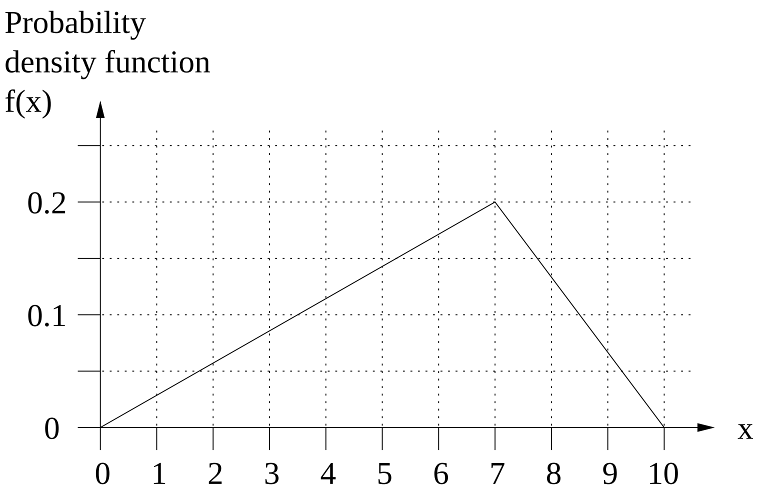
1 **(long)**

(a) Explain the difference between the *population standard deviation* and the *sample standard deviation*. State the formula for each of them and provide an example illustrating how each can be correctly used. [10]

(b) An experimental reading is taken 50 times. It is believed that the experiment is producing unbiased results that do not vary as a function of time. The sum of the readings is 5.0 units and the sum of the squares of the readings is 1.0 units^2. Estimate the range in which you are 95% confident that the true experimental results lies. [14]

(c) Suppose that all the individual readings were available together with the time at which they were each taken. Suggest a way of checking that the experimental results do not vary as a function of time. (You do not have to perform the check that you suggest.) [6]

2 **(short)** The graph below shows the probability density function of a population.



(a) What is the probability that a sample from this population has a value greater than 7? [2]

(b) Calculate the mean and variance of this population. [8]

3 **(short)** For each of the cases below, either find a set of five numbers to satisfy the condition or explain why it cannot be achieved.

(a) All the numbers are less in magnitude than their mean. [2]

(b) All the numbers are greater in magnitude than their mean. [2]

(c) All the numbers are less than their standard deviation. [2]

(d) All the numbers are less in magnitude than their standard deviation. [2]

(e) All the numbers are greater in magnitude than their standard deviation. [2]

Solutions to Sample Statistics Questions

1. (a) The population standard deviation is simply the root mean square deviation of a set of values from their mean.

$$\text{Standard Deviation: } \sigma = \sqrt{\frac{\sum_{i=1}^{N}(x_i - \mu)^2}{N}}$$

where μ is the mean of the population and N is the number of elements in the population being considered. This formula is appropriate when you have access to the values of all the members of the population. You would could use it to calculate the standard deviation of the widths of all the spaces in a particular car park.

The sample standard deviation is useful when you don't have access to the underlying population but you wish to estimate its standard deviation based on the properties of a sample.

$$\text{Estimate of Standard Deviation (based on sample): } s = \sqrt{\frac{\sum_{i=1}^{n}(x_i - m)^2}{n-1}}$$

Here m is the mean of the values in the sample and n is the number of values in the sample. This can be used if you have a sample of values from an underlying distribution. For example you might pick 10 parts from a production line during a day, and wish to estimate the standard deviation of (say) the lengths of the whole population of parts produces by the production line that day.

(b) We estimate the standard deviation of the experimental values as 0.1010 therefore our estimate of the standard deviation of $\overline{x}$ is $0.1010/\sqrt{50} = 0.0143$. We would expect the true experimental results to have a 95% chance of lying within two standard deviations from the mean, i.e. in the range 0.0714 to 0.1286.

(c) Plot a graph of the values against the time they were acquired and see if there is a trend or systematic variation as a function of time. If it looks as if the variation is linear, fit a line to the values and on the graph and see if it has non-zero gradient.

2. (a) $\mathrm{P}(x > 7) = 3 \times 0.2/2 = 0.3$

(b)

$$\mu = \int_{-\infty}^{+\infty} x f(x)\, dx = \int_0^7 x^2 \frac{0.2}{7}\, dx + \int_7^{10} x \frac{10 - x}{15}\, dx$$
$$= \frac{17}{3}$$

$$\sigma = \sqrt{\int_{-\infty}^{+\infty} (x-\mu)^2 f(x)\, dx} = \sqrt{\int_0^7 \left(x - \frac{17}{3}\right)^2 x \frac{0.2}{7}\, dx + \int_7^{10} \left(x - \frac{17}{3}\right)^2 \frac{10 - x}{15}\, dx}$$
$$= 2.095$$

3. (a) Impossible: at least one number must be greater than the mean.

(b) -5, -6, 4, 3, 4: mean is zero, all numbers are greater in magnitude than zero.

(c) -5, -5, -5, -5, -5: standard deviation is zero, all number are less than zero.

(d) Impossible: cannot make standard deviation bigger than spread of the numbers.

(e) -5, -5, -5, -5, -5: standard deviation is zero, all numbers are greater in magnitude than zero.