

MPO 624-2018/ Applied Data Analysis

Exam (Module 1)

30 Minutes, 30 points

1) For a probability distribution function (PDF) $p(x)$, or a dataset $X = \{X_i\}$, fill in the following table:

	Name	Integral form	Bar notation
0 th moment			
1 st moment	mean		
2 nd moment			$\overline{X^2}$
3 rd moment		$\int_{-\infty}^{\infty} x^3 p(x) dx$	
4 th moment	(related to) kurtosis		$\overline{X^4}$

(10 marks)

2) Suppose an anomaly dataset $Y' = \{Y'_i\}$ is a perfect fit to (that is, is exactly equal to) $5X'^3$. Sketch the scatterplot.

(a) Describe the results a,b of an order-2 polynomial fit:

$$Y' = aX' + bX'^2 + \text{residual (whose RMS is minimized)?}$$

(b) Describe the results a,b,c,d of an order-4 polynomial fit,

$$Y' = aX' + bX'^2 + cX'^3 + dX'^4 + \text{residual (whose RMS is minimized)?}$$

(6 marks)

3) If $X = \{3, -3, 0\}$ and $Y = \{4, 4, -8\}$, are the **vectors** X and Y orthogonal? Are the **deviations** X' and Y' correlated?

(2 marks)

4) Suppose you wake up one day not feeling very well. You browse through the internet to find that your symptoms match with a disease Stasisytis, an extremely rare disease affecting only 1 in every 100000 people in the world. You also learn that if that you have a Stasisytis, the probability of exhibiting these symptoms is 0.95. Suppose that the symptoms are also 100% certain symptoms of flu, which is seen in 1% of the population during this season. What is the probability that you have Stasisytis?

(9 marks)

5) Suppose variable V is a **standardized** version of V_{raw} , a Gaussian (Normal) distributed variable.

a. How much less likely (or probable) is a value of $V=2$ than a value of $V=0$?

b. “ “ “ “ “ a value of $V=4$ than a value of $V=0$?

c. How many times will you see values of V very close to $V=2$, before you see a value very close to $V=6$? (Hint: The answer doesn't depend on the exact definition of “very close”, for small enough tolerances – use the same math as a. and b.).

(3 marks)