TRINITY COLLGE DUBLIN

School of Computer Science and Statistics

Extra Questions ST3009: Statistical Methods for Computer Science

Question 1. Suppose a continuous valued random variable X has probability density function:

$$f(x) = \begin{cases} 0 & x \leq 0 \\ 1 & 0 < x \leq 1 \\ 0 & x > 1 \end{cases}$$

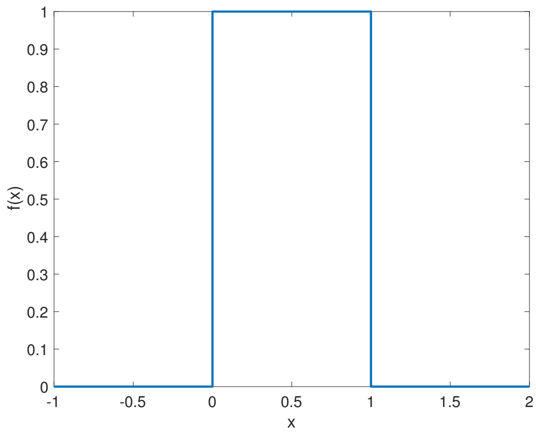

Figure 1: Plot of PDF $f(x)$

(a) Calculate $P(0 \leq X \leq 0.25)$
(b) Calculate $P(0 \leq X \leq 0.75)$
(c) Calculate $P(0.5 \leq X \leq 2)$
(d) Calculate the CDF for X

Solution

- The area under the PDF curve between 0 and 0.25, which is 0.25
- The area under the PDF curve between 0 and 0.75, which is 0.75
- The area under the PDF curve between 0.5 and 2. We split this into two parts, the area between 0.5 and 1 and the area between 1 and 2. The area between 0.5 and 1 is 0.5. The area between 1 and 2 is 0. So the total area is 0.5
- $$F(x) = \begin{cases} 0 & x \leq 0 \\ x & 0 < x \leq 1 \\ 1 & x > 1 \end{cases}$$

Question 2. Suppose a continuous valued random variable X has probability density function:

$$f(x) = \begin{cases} 0 & x \leq 0 \\ 2x & 0 < x \leq 1 \\ 0 & x > 1 \end{cases}$$

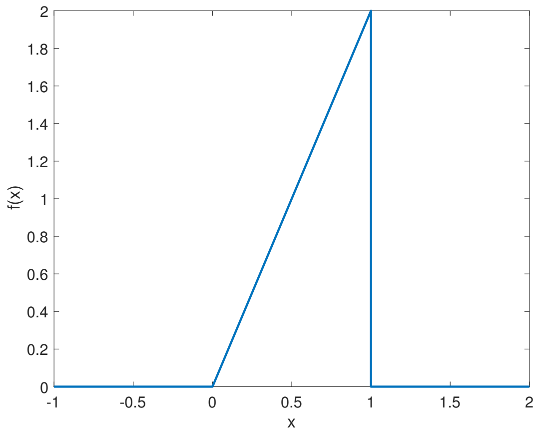

Figure 2: Plot of PDF $f(x)$

(a) Calculate $P(0 \leq X \leq 0.25)$
(b) Calculate $P(0 \leq X \leq 0.75)$
(c) Calculate $P(0.5 \leq X \leq 2)$
(d) Calculate the CDF for X
Solution

- The area under the PDF curve between 0 and 0.25, which is 0.0625 (the area of a triangle is given by half the base times the height).
- The area under the PDF curve between 0 and 0.75, which is 0.5625
- The area under the PDF curve betwee 0.5 and 2. We split this into two parts, the area between 0.5 and 1 and the area between 1 and 2. To calculate the area between 0.5 and 1 observer that it can be viewed as a rectangle of side 0.5 and height 1 with a triangle on top of base 0.5 and height 1. The area of the square is 0.5 and of the triangle is 0.25, so the overall area is 0.75. The area under the PDF between 1 and 2 is 0. So the total area under the PDF between 0.5 and 2 is 0.75
- $$F(x) = \begin{cases} 0 & x \leq 0 \\ x^2 & 0 < x \leq 1 \\ 1 & x > 1 \end{cases}$$

Question 3. Suppose a continuous valued random variable X has probability density function:

$$f(x) = \begin{cases} 0 & x \leq 0 \\ 4x & 0 < x \leq 0.5 \\ 4 - 4x & 0.5 < x \leq 1 \\ 0 & x > 1 \end{cases}$$

(a) Calculate $P(0 \leq X \leq 0.25)$
(b) Calculate $P(0 \leq X \leq 0.75)$
(c) Calculate $P(0.5 \leq X \leq 2)$
Solution

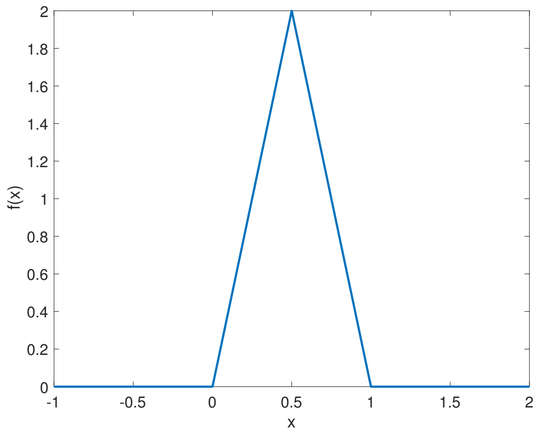

Figure 3: Plot of PDF $f(x)$

- The area under the PDF curve between 0 and 0.25, which is $0.25 \times 1/2 = 0.125$ (the area of a triangle is given by half the base times the height).
- The area under the PDF curve between 0 and 0.75. We split this into the area between 0 and 0.5 and the area between 0.5 and 0.75. The area of the triangle between 0 and 0.5 is $0.5 \times 2/2 = 0.5$. The area between 0.5 and 0.75 can be viewed as a rectangle of base 0.25 and height 1 with a triangle of height of base 0.25 and height 1 on top of it. So the overall area is $1 \times 0.25 + 0.25 \times 1/2 = 0.375$. Adding this to the area 0 and 0.5 gives a total of 0.875. Alternatively, we know that the total area under the PDF is 1. The area under the triangle between 0.75 and 1 is $0.25 \times 1/2 = 0.125$ and so the area of the PDF between 0 and 0.75 must be 1-0.125=0.875.
- The area between 0.5 and 1 is 0.5. The area between 1 and 2 is 0. So the area between 0.5 and 2 is 0.5.

Question 4. Suppose a continuous valued random variable X has probability density function:

$$f(x) = \begin{cases} 0 & x \leq 0 \\ 1 & 0 < x \leq 1 \\ 0 & x > 1 \end{cases}$$

(a) Calculate $\int_0^1 x dx$. Hint: recall that this is the area under a triangle with base 1 and height 1.

(b) What is the expected value of X ?

Solution

- The area under a triangle with base 1 and height 1, so area is 0.5.
- The expected value is $E[X] = \int_{-\infty}^{\infty} x f(x) dx = \int_0^1 x dx$ since $f(x) = 0$ for $x < 0$ and $x > 1$. Now $\int_0^1 x dx$ is the area under a triangle with base 1 and height 1, $E[X] = 0.5$

Question 5. The CDF of a continuous valued random variable X is

$$F(x) = \begin{cases} 0 & x \leq 0 \\ x & 0 < x \leq 1 \\ 1 & x > 1 \end{cases}$$

(a) Calculate $P(X \leq 0.25)$
(b) Calculate $P(0 \leq X \leq 0.25)$
(c) Calculate $P(0.5 \leq X \leq 0.75)$
(d) Calculate $P(2 \leq X \leq 3)$
(e) Sketch a graph of the CDF.
Solution

- $P(X \leq 0.25) = F(0.25) = 0.25$
- $P(0 \leq X \leq 0.25) = F(0.25) - F(0) = 0.25$
- $P(0.75 \leq X \leq 0.25) = F(0.75) - F(0.25) = 0.75 - 0.25 = 0.5$
- $P(2 \leq X \leq 3) = F(3) - F(2) = 1 - 1 = 0$

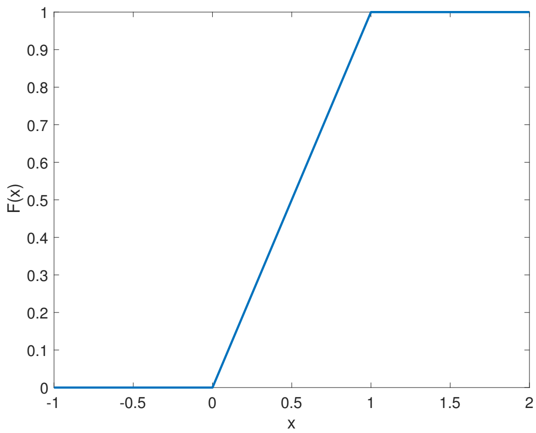

Figure 4: Plot of CDF $F(x)$

Question 6. The CDF of a continuous valued random variable X is

$$F(x) = \begin{cases} 0 & x \leq 0 \\ 1 - e^{-x} & x \geq 0 \end{cases}$$

(a) Calculate $P(X \leq 1)$
(b) Calculate $P(0 \leq X \leq 1)$
(c) Calculate $P(1 \leq X \leq 2)$
(d) Sketch a graph of the CDF.
Solution

- $P(X \leq 1) = F(1) = 1 - e^{-1}$
- $P(0 \leq X \leq 1) = F(1) - F(0) = 1 - e^{-1} - 0 = 1 - e^{-1}$
- $P(1 \leq X \leq 2) = F(2) - F(1) = 1 - e^{-2} - (1 - e^{-1} = e^{-1} - e^{-2}$

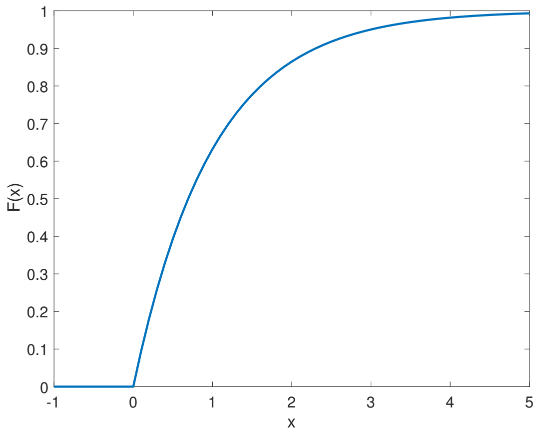

Figure 5: Plot of CDF $F(x)$

Question 7. You carry out a poll asking n eskimos selected independently and uniformly at random from the population whether they like warm weather. Let X_i be 1 if eskimo i likes warm weather and 0 otherwise. You calculate the sample average $X = \frac{1}{n}\sum_{i=1}^{n} X_i$ and use this as an estimate of the probability that an eskimo likes warm weather.

(a) State the central limit theorem

(b) How might the central limit theorem be used to obtain a confidence interval for X as an estimate of the probability that an eskimo likes warm weather?

(c) Discuss the assumptions made.

Solution

•

Question 8. In a study on cholestrol levels a sample of 12 men and women was chosen. The plasma cholestrol levels (mmol/L) of the subjects were as follows:

6.0,6.4,7.0,5.8,6.0,5.8,5.9,6.7,6.1,6.5,6.3,5.8

(a) Explain how you estimate the mean of the plasma cholestrol levels, and its 95% confidence interval, using bootstrapping

(b) Write a short matlab program to carry out these estimates. Compare with the confidence interval obtained in last weeks test question 7.

(c) Discuss the assumptions made when using bootstrapping.

(d) How might the central limit theorem be used to obtain a confidence interval ? How do the assumptions differ from when bootstrapping is used.