Department of Mathematics and Statistics

UNIVERSITY OF CANTERBURY
Te Whare Wānanga o Waitaha
CHRISTCHURCH NEW ZEALAND

CSE Exercises Wk 1

(1) The sensitivity and specificity of a medical diagnostic test for a disease are defined as

sensitivity = P(test is positive | patient has the disease),

specificity = P(test is negative | patient does not have the disease).

Suppose that a medical test has a sensitivity of 0·7 and a specificity of 0·95. If the prevalence of the disease in the general population is 1%, find

(a) the probability that a patient who tests positive actually has the disease,

(b) the probability that a patient who tests negative is free from the disease.

(2) The detection rate and false alarm rate of an intrusion sensor are defined as

detection rate = P(detection declared | intrusion),

false alarm rate = P(detection declared | no intrusion).

If the detection rate is 0·999 and the false alarm rate is 0·001, and the probability of an intrusion occurring is 0·01, find

(a) the probabity that there is an intrusion when a detection is declared,

(b) the probability that there is no intrusion when no detection is declared.

(3) Let A and B be events such that $P(A) \neq 0$ and $P(B) \neq 0$. When A and B are disjoint, are they also independent? Explain clearly why or why not.

(4) Let $a, b \in \mathbb{R}$, such that $a < b$. Let X be a Uniform(a, b) random variable whose PDF is given by

$$f(x) = c\, \mathbb{1}_{[a,b]}(x) = \begin{cases} c, & a \leq x \leq b, \\ 0, & \text{otherwise}, \end{cases}$$

where c is a constant.

(a) Find the following in terms of a and b:

(i) c,

(ii) $E(X)$,

(iii) $E(X^2)$,

(iv) $V(X)$.

(b) Plot the Uniform (a, b) PDF and CDF for $a = -1$ and $b = 1$.

(5) Starting from the definition of the variance of a random variable (Definition 20), show that

$$V(X) = E(X^2) - E(X)^2.$$

(6) Let X be a discrete random variable with PMF given by

$$f(x) = \begin{cases} \frac{x}{10}, & x \in \{1, 2, 3, 4\}, \\ 0, & \text{otherwise}. \end{cases}$$

(a) Find (i) $P(X = 0)$,

(ii) $P(2.5 < X < 5)$,

(iii) $E(X)$,

(iv) $V(X)$.

(b) Write down the CDF of X.

(b) Plot the PMF and CDF of X.

⑦ Let X be a discrete random variable whose CDF is given by

$$F(x) = \begin{cases} 0 , & x < 1.1 , \\ 0.5 , & 1.1 \leq x < 2.3 , \\ 0.7 , & 2.3 \leq x < 3.7 , \\ 1 , & x \geq 3.7 . \end{cases}$$

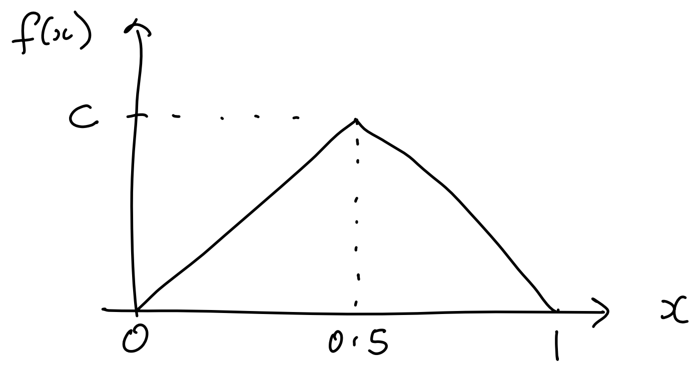
(a) Write down the PMF of X.

(b) Plot the PMF and CDF of X.

(c) Find (i) $E(X)$,

(ii) $V(X)$.

⑧ Let X be a random variable with PDF given in the figure below:



(a) What is the value of c ?

(b) Write down the formulas for the PDF and CDF :

$$f(x) = ?$$

$$F(x) = ?$$

(c) Plot the CDF .

(d) Find (i) $E(X)$,

(ii) $V(X)$.

(9) The covariance of two random variables X and Y is defined as

$$\operatorname{Cov}(X,Y) := E((X-E(X))(Y-E(Y))).$$

(a) Show, starting from the definition, that

$$\operatorname{Cov}(X,Y) = E(XY) - E(X)E(Y).$$

(b) When $\operatorname{Cov}(X,Y) = 0$, X and Y are said to be "uncorrelated". Show that if X and Y are independent, then they are also uncorrelated, i.e. show that

$$X \text{ and } Y \text{ independent} \implies X \text{ and } Y \text{ uncorrelated}.$$

(Note, however, that the converse is generally not true.)

(10) Let $X_1, \ldots, X_n$ be random variables. Their joint CDF is defined as

$$F(x_1, \ldots, x_n) := P(X_1 \leq x_1, \ldots, X_n \leq x_n) .$$

By repeated application of the definition of conditional probability, show that the joint CDF admits the following "telescopic" representation:

$$F(x_1, \ldots, x_n) = F(x_n | x_1, \ldots, x_{n-1}) F(x_{n-1} | x_1, \ldots, x_{n-2}) \cdots$$

$$\cdots F(x_2 | x_1) F(x_1)$$

$$= F(x_1) \prod_{i=2}^{n} F(x_i | x_1, \ldots, x_{i-1}) ,$$

where $F(x_i | x_1, \ldots, x_{i-1})$ denotes the conditional probability, $P(X_i \leq x_i | X_1 \leq x_1, \ldots, X_{i-1} \leq x_{i-1})$.

Note that the same telescopic representation exists for a joint PMF (defined in the obvious way), and even for a joint PDF (which will be defined later on).

Solutions

① Let $+$ = test positive,
$-$ = test negative,
D = patient has disease,
ND = patient does not have disease.

Given: $P(D) = 0.01$, $P(ND) = 1 - P(D) = 0.99$,

sensitivity $= P(+|D) = 0.7$, $P(-|D) = 1 - P(+|D) = 0.3$,

specificity $= P(-|ND) = 0.85$, $P(+|ND) = 1 - P(-|ND) = 0.15$.

By Bayes' Theorem,

(a) $$P(D|+) = \frac{P(+|D)\,P(D)}{P(+)}$$

$$= \frac{P(+|D)\,P(D)}{P(+|D)\,P(D) + P(+|ND)\,P(ND)}$$

$$= \frac{(0.7)(0.01)}{(0.7)(0.01) + (0.15)(0.99)} = 0.045.$$

(b) $$P(ND|-) = \frac{P(-|ND)\,P(ND)}{P(-)}$$

$$= \frac{P(-|ND)\,P(ND)}{P(-|D)\,P(D) + P(-|ND)\,P(ND)}$$

$$= \frac{(0.85)(0.99)}{(0.3)(0.01) + (0.85)(0.99)} = 0.9964.$$

(2) Let D = detection,
ND = no detection,
I = intrusion,
NI = no intrusion,

Given $P(I) = 0.01$, $P(NI) = 1 - P(I) = 0.99$,

detection rate $= P(D|I) = 0.999$, $P(ND|I) = 1 - P(D|I) = 0.001$,

false alarm rate $= P(D|NI) = 0.001$, $P(ND|NI) = 1 - P(D|NI) = 0.999$.

By Bayes' Theorem,

(a) $$P(I|D) = \frac{P(D|I)\,P(I)}{P(D)} = \frac{P(D|I)\,P(I)}{P(D|I)\,P(I) + P(D|NI)\,P(NI)}$$

$$= \frac{(0.999)(0.01)}{(0.999)(0.01) + (0.001)(0.99)} = 0.9098.$$

(b) $$P(NI|ND) = \frac{P(ND|NI)\,P(NI)}{P(ND)} = \frac{P(ND|NI)\,P(NI)}{P(ND|I)\,P(I) + P(ND|NI)\,P(NI)}$$

$$= \frac{(0.999)(0.99)}{(0.001)(0.01) + (0.999)(0.99)} = 0.99999.$$

③ A, B events with $P(A) \neq 0$ and $P(B) \neq 0$.

When A and B are disjoint, $A \cap B = \phi$, and so

$$P(A \cap B) = P(\phi) = 0. \quad \text{———} \quad (1)$$

Suppose A and B are also independent. Then

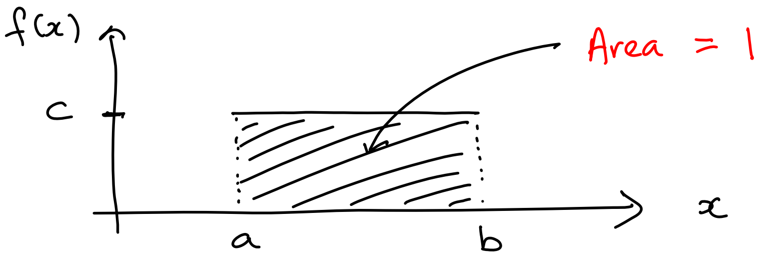
$$P(A \cap B) = P(A)\,P(B) \neq 0,$$

which contradicts (1).

Hence, when A and B are disjoint, they cannot also be independent.

④ Given $a, b \in \mathbb{R}$, such that $a < b$, and $X \sim \text{Uniform}(a, b)$ with PDF, $f(x) = c\,\mathbb{1}_{[a,b]}(x)$.

(a) (i) Sketch of PDF:



Since the area under the density between a and b must integrate to 1, and in this case corresponds to a rectangle with length $b-a$, the width c must be equal to $1/(b-a)$. Mathematically,

$$\int_{\mathbb{R}} f(x)\, dx = 1$$

$$\iff \int_{\mathbb{R}} c\, \mathbb{1}_{[a,b]}(x)\, dx = 1$$

$$\iff c \int_a^b 1\, dx = 1$$

$$\iff c \cdot x \Big|_a^b = 1$$

$$\iff c\,(b-a) = 1$$

$$\iff c = \frac{1}{(b-a)} .$$

(ii)
$$E(X) = \int_{\mathbb{R}} x f(x)\, dx$$

$$= \frac{1}{(b-a)} \int_a^b x\, dx$$

$$= \frac{1}{(b-a)} \cdot \frac{x^2}{2}\Big|_a^b$$

$$= \frac{1}{(b-a)} \frac{(b^2-a^2)}{2}$$

$$= \frac{1}{(b-a)} \frac{(b-a)(b+a)}{2} = \frac{a+b}{2} .$$

(iii)
$$E(X^2) = \int_{\mathbb{R}} x^2 f(x)\, dx$$

$$= \frac{1}{(b-a)} \int_a^b x^2\, dx$$

$$= \frac{1}{(b-a)} \cdot \frac{x^3}{3}\Big|_a^b$$

$$= \frac{1}{(b-a)} \frac{(b^3-a^3)}{3}$$

$$= \frac{1}{(b-a)} \frac{(b-a)(a^2+ab+b^2)}{3}$$

$$= \frac{a^2+ab+b^2}{3} .$$

(iv) $V(X) = E(X^2) - E(X)^2$

$$= \frac{a^2+ab+b^2}{3} - \left(\frac{a+b}{2}\right)^2$$

$$= \frac{1}{12}\left[4(a^2+ab+b^2) - 3(a^2+2ab+b^2)\right]$$

$$= \frac{1}{12}(a^2 - 2ab + b^2)$$

$$= \frac{(b-a)^2}{12}.$$

(b) Modify unif01pdf.m, unif01cdf.m and plotunif.m to get the required plots.

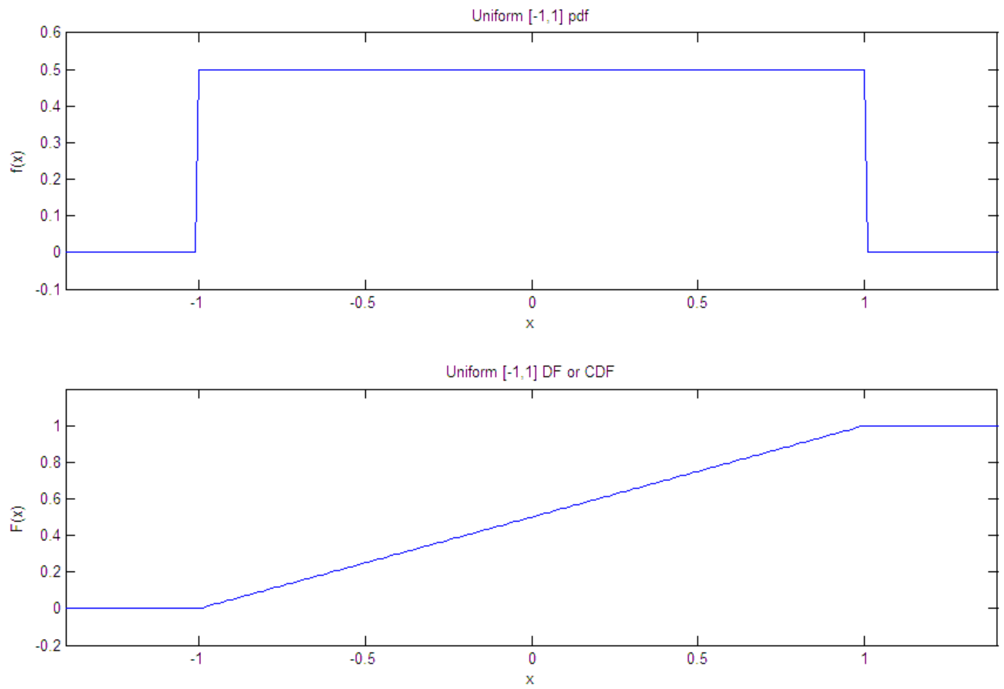
Note that for $x \in [a,b]$, the Uniform (a,b) CDF is

$$F(x) = \int_a^x f(z)\,dz$$ (here z is a dummy variable of integration)

$$= \frac{1}{(b-a)}\int_a^x 1 \, dz$$

$$= \frac{1}{(b-a)} \cdot z\Big|_a^x$$

$$= \frac{x-a}{b-a}.$$



(5) $$V(X) := E((X - E(X))^2)$$

$$= E(X^2 - 2X\,E(X) + E(X)^2)$$

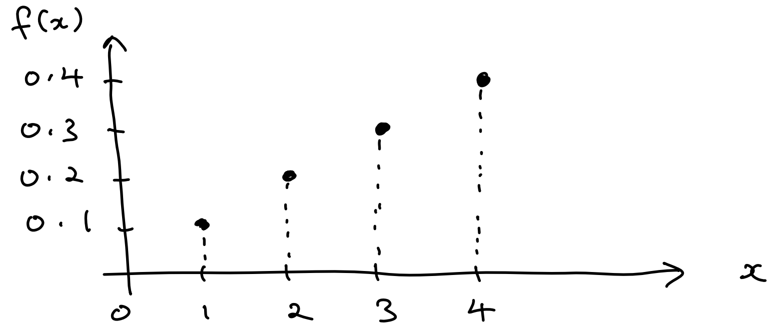
$$= E(X^2) - 2E(X)E(X) + E(X)^2$$

$$= E(X^2) - E(X)^2 .$$

(6) Given that X is discrete with PMF

$$f(x) = \begin{cases} \frac{x}{10}, & x \in \{1, 2, 3, 4\}, \\ 0, & \text{otherwise}. \end{cases}$$

Sketch of PMF :



(a) (i) $P(X = 0) = 0$.

(ii) $P(2.5 < X < 5) = P(X = 3) + P(X = 4)$
$= 0.3 + 0.4$
$= 0.7$.

(iii) $E(X) = \sum_x x f(x)$

$= 1(0.1) + 2(0.2) + 3(0.3) + 4(0.4)$

$= 0.1 + 0.4 + 0.9 + 1.6$

$= 3$

(iv) $E(X^2) = \sum_{x} x^2 f(x)$

$= 1^2(0.1) + 2^2(0.2) + 3^2(0.3) + 4^2(0.4)$

$= 1(0.1) + 4(0.2) + 9(0.3) + 16(0.4)$

$= 0.1 + 0.8 + 2.7 + 6.4$

$= 10$

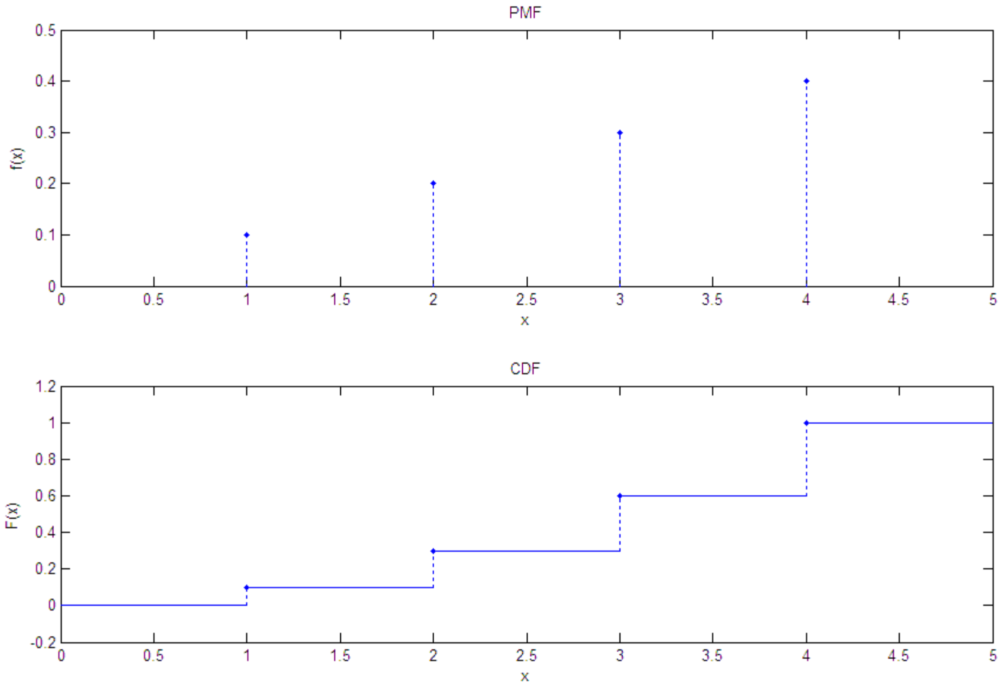
$V(X) = E(X^2) - E(X)^2$

$= 10 - 3^2 = 1$.

(b) CDF of X is

$$F(x) = \begin{cases} 0, & x < 1, \\ 0.1, & 1 \leq x < 2, \\ 0.3, & 2 \leq x < 3, \\ 0.6, & 3 \leq x < 4, \\ 1, & x \geq 4. \end{cases}$$

(c)



(7) Given that X is discrete with CDF

$$F(x) = \begin{cases} 0, & x < 1.1, \\ 0.5, & 1.1 \leq x < 2.3, \\ 0.7, & 2.3 \leq x < 3.7, \\ 1, & x \geq 3.7. \end{cases}$$

(a) PMF of X is

$$f(x) = \begin{cases} 0.5, & x = 1.1, \\ 0.2, & x = 2.3, \\ 0.3, & x = 3.7, \\ 0, & \text{otherwise}. \end{cases}$$

(c) (i) $E(X) = \sum_x x f(x)$

$= (1.1)(0.5) + (2.3)(0.2) + (3.7)(0.3) = 2.12$.

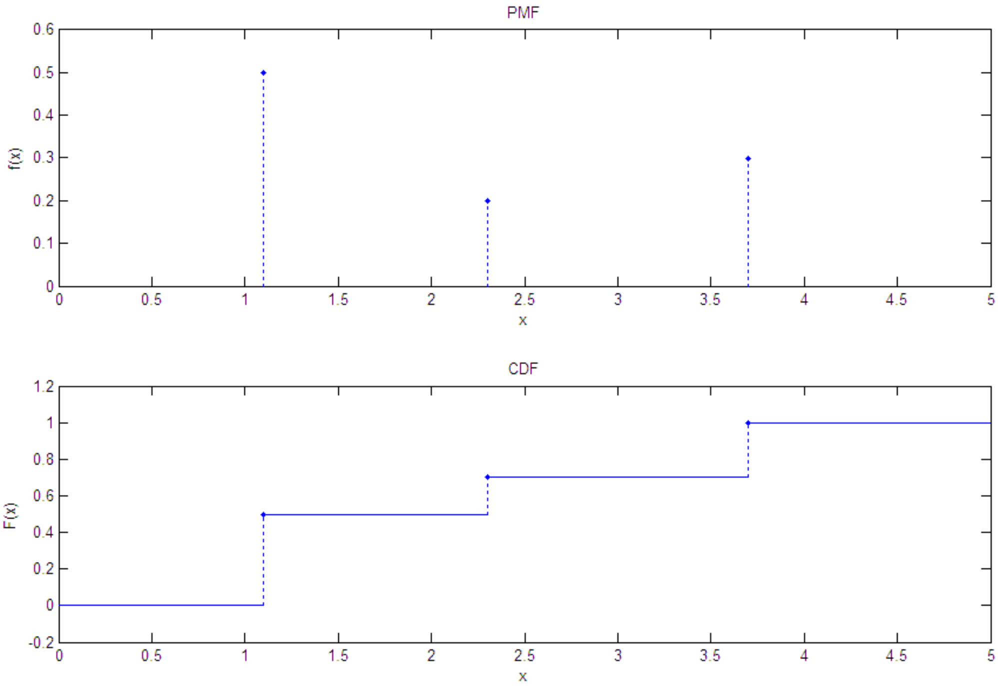
(ii) $E(X^2) = \sum_x x^2 f(x)$

$= (1.1^2)(0.5) + (2.3^2)(0.2)(3.7^2)(0.3) = 5.77$

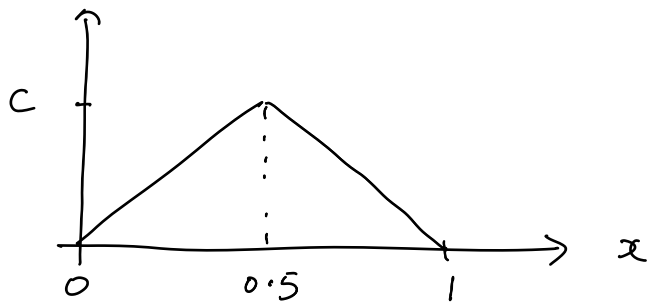
$V(X) = E(X^2) - E(X)^2$

$= 5.77 - 2.12^2 = 1.2756$.

(b)


PMF
f(x)
x
0.6
0.5
0.4
0.3
0.2
0.1
0
0.5
1
1.5
2
2.5
3
3.5
4
4.5
5
CDF
F(x)
1.2
1
0.8
0.6
0.4
0.2
-0.2

(8) Given that X has PDF as shown:



(a) Since the area under the density must be 1 and, in this case, the area is given by a triangle with base 1, c must be equal to 2.

(b)

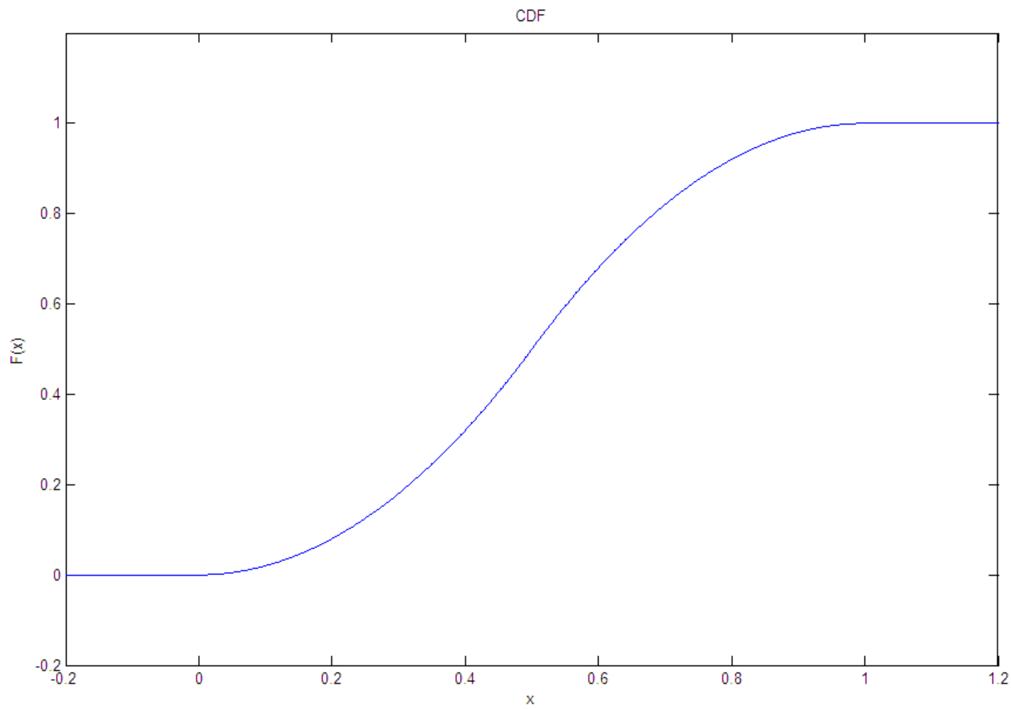
$$f(x) = \begin{cases} 4x, & 0 \le x < 0.5, \\ 4(1-x), & 0.5 \le x < 1, \\ 0, & \text{otherwise}. \end{cases}$$

$$F(x) = \int_{-\infty}^{x} f(z)\,dz$$ (here z is a dummy variable of integration)

$$= \begin{cases} 0, & x < 0, \\ 4\int_0^x z\,dz, & 0 \le x < 0.5, \\ 0.5 + 4\int_{0.5}^{x} (1-z)\,dz, & 0.5 \le z < 1, \\ 1, & x \ge 1, \end{cases}$$

$$= \begin{cases} 0, & x < 0, \\ 2x^2, & 0 \leq x < 0.5, \\ -2x^2+4x-1, & 0.5 \leq x < 1 \\ 1, & x \geq 1. \end{cases}$$

(c)



(d) (i) $E(X) = \int_{\mathbb{R}} x\, f(x)\, dx$

$$= 4\int_0^{1/2} x^2\, dx + 4\int_{1/2}^{1} x(1-x)\, dx$$

$$= 4\left[\left.\frac{x^3}{3}\right|_0^{1/2} + \left.\left(\frac{x^2}{2} - \frac{x^3}{3}\right)\right|_{1/2}^{1} \right]$$

$$= 4\left[\frac{1}{24} + \left(\frac{1}{2} - \frac{1}{3}\right) - \left(\frac{1}{8} - \frac{1}{24}\right) \right]$$

$$= 4\left(\frac{1}{8}\right) = \frac{1}{2} \quad \text{(as expected)}.$$

(ii) $E(X^2) = \int_{\mathbb{R}} x^2 f(x)\, dx$

$$= 4\int_0^{1/2} x^3\, dx + 4\int_{1/2}^{1} x^2(1-x)\, dx$$

$$= 4\left[\left.\frac{x^4}{4}\right|_0^{1/2} + \left.\left(\frac{x^3}{3} - \frac{x^4}{4}\right)\right|_{1/2}^{1} \right]$$

$$= 4\left[\frac{1}{64} + \left(\frac{1}{3} - \frac{1}{4}\right) - \left(\frac{1}{24} - \frac{1}{64}\right) \right]$$

$$= 4\left(\frac{1}{32} + \frac{1}{24}\right) = \frac{7}{24}$$

$$V(X) = E(X^2) - E(X)^2$$

$$= \frac{7}{24} - \left(\frac{1}{2}\right)^2$$

$$= \frac{7}{24} - \frac{1}{4}$$

$$= \frac{1}{24}.$$

(9) (a) $\mathrm{Cov}(X,Y) := E\big((X-E(X))(Y-E(Y))\big)$

$$= E\big(XY - XE(Y) - E(X)Y + E(X)E(Y)\big)$$

$$= E(XY) - E(X)E(Y) - E(X)E(Y) + E(X)E(Y)$$

$$= E(XY) - E(X)E(Y) .$$

(b) We will show this for discrete random variables and revisit it for continuous random variables later on after the concept of a joint PDF is introduced. So suppose that X and Y are discrete. Their joint PMF is

$$f(x,y) := P(X=x, Y=y)$$

Recall from Section 3.5 that when X and Y are independent,

$$P(X=x, Y=y) = P(X=x)\,P(Y=y).$$

Hence, when X and Y are independent,

$$f(x,y) = f(x)\,f(y)$$

i.e. their joint PMF is a product of their marginal PMF's.

Now, $E(XY) := \sum_x \sum_y xy\, f(x,y)$

and if X and Y are independent, then

$$E(XY) = \sum_x \sum_y x\, y\, f(x)\, f(y)$$

$$= \left[\sum_x x f(x)\right] \left[\sum_y y f(y)\right]$$

$$= E(X)\ E(Y)$$

i.e. $E(XY) - E(X)E(Y) = \mathrm{cov}(X, Y) = 0$

i.e. X and Y are uncorrelated.

(10) $F(x_1, \ldots, x_n)$

$= P(X_1 \leq x_1, \ldots, X_n \leq x_n)$

$= P(X_n \leq x_n \mid X_1 \leq x_1, \ldots, X_{n-1} \leq x_{n-1})\, P(X_1 \leq x_1, \ldots, X_{n-1} \leq x_{n-1})$

$= P(X_n \leq x_n \mid X_1 \leq x_1, \ldots, X_{n-1} \leq x_{n-1})$

$\quad \cdot P(X_{n-1} \leq x_{n-1} \mid X_1 \leq x_1, \ldots, X_{n-2} \leq x_{n-2})\, P(X_1 \leq x_1, \ldots, X_{n-2} \leq x_{n-2})$

$\vdots$ and so on

$= P(X_n \leq x_n \mid X_1 \leq x_1, \ldots, X_{n-1} \leq x_{n-1})$

$\quad \cdot P(X_{n-1} \leq x_{n-1} \mid X_1 \leq x_1, \ldots, X_{n-2} \leq x_{n-2}) \cdots$

$\quad \cdots P(X_2 \leq x_2 \mid X_1 \leq x_1)\, P(X_1 \leq x_1)$

$= F(x_n \mid x_1, \ldots, x_{n-1})\, F(x_{n-1} \mid x_1, \ldots, x_{n-2}) \cdots$

$\quad \cdots F(x_2 \mid x_1)\, F(x_1)$.