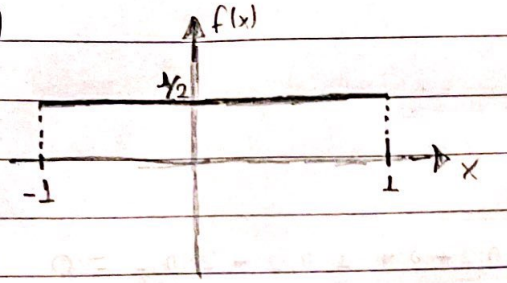
S T Q Q S S D

__/__/__

PPGEE2249 - Aprendizado de Máquina

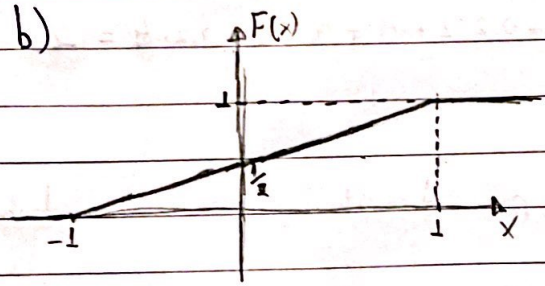
Pedro Alvim - 180108042

① a)



Como é uniforme - a probabilidade é constante em qualquer ponto. Assim, como a área de uma PDF precisa ser igual a 1, de -1 a 1 $f(x) = 1/2$.

b)



$$F(x) = \begin{cases} 0, & x < -1 \\ (x+1)/2, & -1 \le x \le 1 \\ 1, & x > 1 \end{cases}$$

c) $P(-0,2 < X < 0,2)$

$= F(0,2) - F(-0,2)$

$= \frac{0,2+1}{2} - \left(\frac{-0,2+1}{2}\right)$

$= 0,6 - 0,4$

$= 0,2 = 20\%$

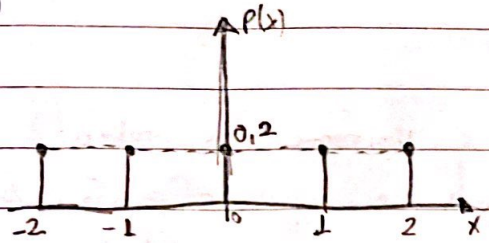
d) $E[X] = \int_{-\infty}^{\infty} x f(x)\,dx = \frac{x^2}{4}\Big|_{-1}^{1} = \frac{1}{4} - \frac{1}{4} = 0$

$E[X^2] = \int_{-\infty}^{\infty} x^2 f(x)\,dx = \frac{x^3}{6}\Big|_{-1}^{1} = \frac{1}{6} - \frac{(-1)}{6} = \frac{1}{3}$

$E[X^4] = \int_{-\infty}^{\infty} x^4 f(x)\,dx = \frac{x^5}{10}\Big|_{-1}^{1} = \frac{1}{10} - \frac{(-1)}{10} = \frac{1}{5}$

Var $E[(X - E(X))^2] = \int_{-\infty}^{\infty} (x-0)^2 f(x)\,dx = E[X^2] = \frac{1}{3}$

②



$p(x) = 0,2$

$$E[X] = \sum_{i=1}^{\infty} x_i p(x_i) = -2 \cdot 0,2 - 1 \cdot 0,2 + 0 + 1 \cdot 0,2 + 2 \cdot 0,2 = 0$$

$$Var[X] = \sum_{i=1}^{\infty} p_i (x_i - \mu)^2 = 0,2 \cdot 4 + 0,2 \cdot 1 + 0 + 0,2 \cdot 1 + 0,2 \cdot 4 = 2$$

③ PDF Conjunta, pode ser expressa pela distribuição Gaussiana bivariada

$$f_{X_1 X_2}(X_1, X_2) = \frac{1}{2\pi \sigma_{x_1} \sigma_{x_2} \sqrt{1-\rho^2}} \exp\left(\frac{-1}{2(1-\rho^2)}\left[\left(\frac{x_1-\mu_{x_1}}{\sigma_{x_1}}\right)^2 - 2\rho\left(\frac{x_1-\mu_{x_1}}{\sigma_{x_1}}\right)\left(\frac{x_2-\mu_{x_2}}{\sigma_{x_2}}\right) + \left(\frac{x_2-\mu_{x_2}}{\sigma_{x_2}}\right)^2\right]\right)$$

ρ = Pearson correlation coeficient = $\frac{Cov(X_1, X_2)}{\sigma_{x_1} \sigma_{x_2}}$ $\quad \mu_1 = -2 \quad \mu_2 = 1$ $\quad \sigma_1 = \sqrt{2} \quad \sigma_2 = 2$

$$\rho = \frac{-0,8}{\sqrt{2} \cdot 2} \qquad \rho^2 = 0,08$$

Voltando

$$= \frac{1}{2\pi \cdot \sqrt{2} \cdot 2\sqrt{1-0,08}} \exp\left(\frac{-1}{2(1-0,08)}\left[\frac{(x_1+2)^2}{2} + \frac{2 \cdot 0,8}{\sqrt{2} \cdot 2}\left(\frac{x_1+2}{\sqrt{2}}\right)\left(\frac{x_2-1}{2}\right) + \frac{(x_2-1)^2}{4}\right]\right)$$

$$f_{x_2}(X_1, X_2) = \frac{1}{4\pi\sqrt{1,84}} \exp\left(\frac{-1}{3,68}\left[(x_1+2)^2 + 0,4(x_1+2)(x_2-1) + \frac{(x_2-1)^2}{2}\right]\right)$$

In [38]:
```
import pandas as pd
import math

df = pd.read_csv('housing.csv')
df.head()
```

Out[38]:

	longitude	latitude	housing_median_age	total_rooms	total_bedrooms	population	households	median_income	median_house_value
0	-122.23	37.88	41.0	880.0	129.0	322.0	126.0	8.3252	452600.0
1	-122.22	37.86	21.0	7099.0	1106.0	2401.0	1138.0	8.3014	358500.0
2	-122.24	37.85	52.0	1467.0	190.0	496.0	177.0	7.2574	352100.0
3	-122.25	37.85	52.0	1274.0	235.0	558.0	219.0	5.6431	341300.0
4	-122.25	37.85	52.0	1627.0	280.0	565.0	259.0	3.8462	342200.0

Choosing median_income and median_house_value

Mean

In [39]:
```
m1 = df['median_income'].mean()
m2 = df['median_house_value'].mean()
```

Variance

In [40]:
```
v1 = df['median_income'].var()
v2 = df['median_house_value'].var()
```

Standard Deviation

In [41]:
```
s1 = math.sqrt(v1)
s2 = math.sqrt(v2)
```

Covariance

In [42]:
```
cov = ((df['median_income'] - m1)*(df['median_house_value'] - m2)).sum() / len(df)
```

Correlation

In [44]:
```
corr = cov / (s1*s2)
```

In [52]:
```
print(f'A correlação entre media de renda do bairro e a media de preço das casas do bairro é de: {corr:.4f}')
```

```
A correlação entre media de renda do bairro e a media de preço das casas do bairro é de: 0.6880
```

In [ ]:
```

```