MM 225: AI and Data Science

Continuous random variables

M P Gururajan and Hina A Gokhale

August 17, 2023

Lecture 4: Continuous random variables

Outline

Recall

Random variable

- Let P be the probability of a continuous random variable X
- Spinner: $P(E) = \int_E f(x)dx$
- $f(x)$: density function
- $f(x)dx$: Probability of outcome x

Density functions

Density function

Definition

Let X be a continuous real-valued random variable. A **density function** for X is a real-valued function f that satisfies

$$P(a \le X \le b) = \int_a^b f(x)dx \tag{1}$$

for all $a, b \in \mathbf{R}$

Remark

In this course, we only consider continuous random variables that possess a density function.

Remark

If $E \subset \mathbf{R}$, $P(X \in E) = \int_E f(x)dx$

Spinner example

- Spinner

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

A question

- Consider a circular target of unit radius
- Suppose darts are thrown
- What is the density function?

Methodology

- Recall the spinner case
- Consider, in this case, the dart falls in the upper half of the target, one quarter of the target and so on
- Let E be the subset of the target – cricle of radius x from the centre of the circular target, let us say!
- We need $P(E) = \int_E f(x)dx$
- What is $f(x)$?

Answer

What is $f(x)$?

- **Link:** https://www.menti.com/al35a47h2ezx
- **Code:** Go to menti.com and use code 1298 9480

Results

Results of the poll

Answer

- Dart game with circular target of unit radius

$$f(x, y) = \begin{cases} \frac{1}{\pi} & \text{if } x^2 + y^2 \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

Cumulative distribution function

Definition and theorem

Definition

Let X be a continuous real-valued random variable. Then the cumulative distribution function of X is defined as $F_X(x) = P(X \leq x)$

Theorem

Let X be a continuous real-valued random variable with density function f(x). Then, the function defined by

$$F(x) = \int_{-\infty}^{x} f(t)dt \tag{2}$$

is the cumulative distribution function of X. Further, $\frac{d}{dx}F(x) = f(x)$

Remark

All random variables have an associated CDF.

An example

Question

Let X be the random variable obtained by squaring a real number chosen at random from [0,1] with uniform probability. What is the cumulative distribution function and density of X?

Answer

Let U be the real number: $X = U^2$. If $0 \leq x \leq 1$ then,
$F_X(x) = P(X \leq x) = P(U^2 \leq x) = P(U \leq \sqrt{x}) = \sqrt{x}$.
Thus,

$$F_X(x) = \begin{cases} 0 & \text{if } x \leq 0 \\ \sqrt{x} & \text{if } 0 \leq x \leq 1 \\ 1 & \text{if } x \geq 1 \end{cases}$$

This implies

$$f_X(x) = \begin{cases} 0 & \text{if } x \leq 0 \\ \frac{1}{2\sqrt{x}} & \text{if } 0 \leq x \leq 1 \\ 0 & \text{if } x \geq 1 \end{cases}$$

Cumulative and density functions

Let us write a python script to plot the cumulative distribution and density functions:

$$F_X(x) = \begin{cases} 0 & \text{if } x \leq 0 \\ \sqrt{x} & \text{if } 0 \leq x \leq 1 \\ 1 & \text{if } x \geq 1 \end{cases}$$

$$f_X(x) = \begin{cases} 0 & \text{if } x \leq 0 \\ \frac{1}{2\sqrt{x}} & \text{if } 0 \leq x \leq 1 \\ 0 & \text{if } x \geq 1 \end{cases}$$

CDFandDF.py

```
import matplotlib.pyplot as plt
import numpy as np
import math

def F(x):
    if(x <=0): return 0
    elif(x>0 and x <=1): return math.sqrt(x)
    else: return 1

def f(x):
    if(x <=0): return 0
    elif(x>0 and x <=1): return 1/(2.*math.sqrt(x))
    else: return 0
```

CDFandDF.py

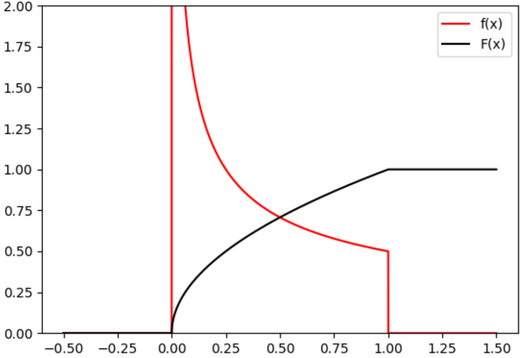
```
x = np.arange(-0.5,1.5,0.0001)

df=[]
for i in range(len(x)): df.append(f(x[i]))

cdf=[]
for i in range(len(x)): cdf.append(F(x[i]))

plt.plot(x,df,color='red')
plt.plot(x,cdf,color='black')
plt.ylim([0,2])
plt.legend(["f(x)","F(x)"],loc="upper right")
plt.show()
```

Density and distribution: $X = U^2$



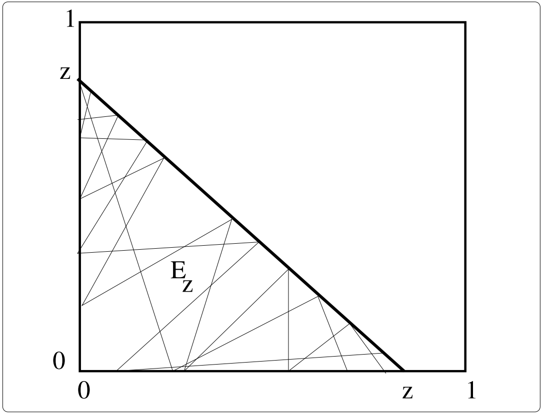
$F_X(x)$ is continuous; $f_X(x)$ is not (while X is continuous)!

Another example

Question

Let Z be the random variable obtained by adding two real numbers X and Y chosen at random from [0,1] with uniform probability. Derive the cumulative distribution and density functions of Z.

Sample space



Derivation

- Sample space Ω: unit square in $\mathbf{R}^2$ with uniform density
- Point $\omega \in \Omega$: (x,y) where x and y are chosen randomly
- $0 < Z < 2$
- E_z: Event $Z \leq z$
- Shaded area in the figure: E_z for any $0 \leq z \leq 1$
- How does the E_z for $1 \leq z \leq 2$ look like?

$$F_Z(z) = P(Z \leq z) = \text{Area of } E_z$$

$$F_Z(z) = \begin{cases} 0 & \text{if } z < 0 \\ \frac{z^2}{2} & \text{if } 0 \leq x \leq 1 \\ 1 - \frac{1}{2}(2-z)^2 & \text{if } 1 \leq z \leq 2 \\ 1 & \text{if } z > 2 \end{cases}$$

This implies

$$f_Z(z) = \begin{cases} 0 & \text{if } z < 0 \\ z & \text{if } 0 \leq x \leq 1 \\ 2 - z & \text{if } 1 \leq z \leq 2 \\ 0 & \text{if } z > 2 \end{cases}$$

Cumulative and density functions

Can you write a python script to plot the cumulative distribution and density functions for Z?

Some important distributions and densities

Continuous uniform density

Random variable U whose value represents the outcome of the experiment of choosing a real number at random from the interval [a,b]

$$f(\omega) = \begin{cases} \frac{1}{b-a} & \text{if } a \leq \omega \leq b \\ 0 & \text{otherwise} \end{cases}$$

Exponential density

Random variable describing the time lapse until something happens; for example, time between emission of particles from a radioactive source

$$f(\omega) = \begin{cases} \lambda \exp\left(-\lambda x\right) & \text{if } 0 \leq x \leq \infty \\ 0 & \text{otherwise} \end{cases}$$

λ: any positive constant

Normal density

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\left[\frac{(x-\mu)^2}{2\sigma^2}\right]} \tag{3}$$

μ and σ: parameters representing the centre and spread of the densities

Normal density

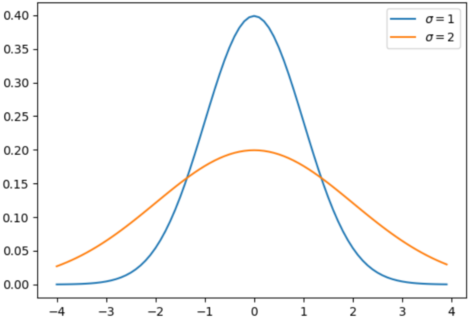


Figure: Normal density with $\mu = 0$.

Cumulative distribution function for normal

- Area under the curve for the normal distribution: leads to error function
- Why error?
- Error function: tabulated

Binomial distribution

Let an experiment be repeated n times; every time, there are two possible outcomes, say, either success or failure, with probabilities p and q respectively; q = 1-p. Let X be the random variable that gives the number of successes in these n trials. It is distributed as binomial distribution with the distribution function $b(n, p, k) = \binom{n}{k} p^k q^{n-k}$

Poisson distribution

Let X be the random variable of an experiment which is known to be distributed as Poisson distribution with parameter λ. Then, the distribution is given by $P(X = k) = \frac{\lambda^k}{k!} e^{-\lambda}$

Some remarks

- More distributions: introduced as we go along
- How to work with these distributions and densities in python: tutorials
- Before we proceed further, we will go back to some probability calculations – combinatorics, conditional
- Let us go back to problem of points in the next lecture

THANK YOU!!!