Statistical Distributions: continous, discrete, properties and simulations

Pietro Colaguori 1936709

November 2023

1 Continous Distributions

Continuous statistical distributions are mathematical functions that describe the probabilities of various outcomes in a continuous random variable. Unlike discrete distributions, which model random variables with distinct, separate values, continuous distributions deal with variables that can take any real value within a specified range.

Some key concepts related to continous distributions are:

1. **Probability Density Function (PDF)**: the PDF denoted as $f(x)$ represents the likelihood of a random variables to assume a specific value or falling in a certain interval, in particular the area under the PDF curve represents the probability of a random variable falling in the corresponding interval.

2. **Cumulative Distributon Function (CDF)**: the CDF, denoted as $F(x)$ represents the probability of a random variable assuming a value that is $\leq x$.

3. **Expected value and variance**: computing the expected value $E(X)$ and the variance $Var(X)$ are computed using integrals, the expected value is also known as μ and represents the average value of a random variable while instead the variance is also known as σ^2 and represents the spread or dispersion of the distribution.

4. **Common continous distributions**: here are listed some famous continous distributions:

   - **Normal Distribution**: The bell-shaped curve is characterized by its mean (μ) and standard deviation (σ). The standard normal distribution has a mean of 0 and a standard deviation of 1.
   - **Uniform Distribution**: All values within a specified range are equally likely. The PDF is a horizontal line over the range.

- **Exponential Distribution**: Models the time until an event occurs in a Poisson process. Commonly used in reliability analysis and queuing theory.
- **Gamma Distribution**: Generalizes the exponential distribution and includes an additional shape parameter. Used in fields such as physics, finance, and reliability engineering.

2 Discrete Distributions

Discrete probability distributions are mathematical functions that describe the probabilities associated with a set of distinct and separate outcomes in a random variable. Unlike continuous distributions, which deal with variables that can take any value within a range, discrete distributions model variables with specific, isolated values. Here are some key concepts related to discrete distributions:

1. **Probability Mass Function (PMF)**: In discrete distributions, probabilities are described by a probability mass function (PMF), denoted as $P(X = x)$, where X is the random variable and x is a specific value. The PMF gives the probability of the random variable taking a particular value. It satisfies two conditions:
   - $P(X = x) \geq 0, \forall x$ (non-negativity)
   - $\sum_x P(X = x) = 1$ (normalization)
2. **Cumulative Distribution Function (CDF)**: The cumulative distribution function (CDF) for discrete distributions, denoted as $F(x)$, gives the probability that the random variable is less than or equal to a specific value x.The CDF is obtained by summing the probabilities from the PMF up to a given value.
3. **Expected value and variance**: the expected value μ is computed as $E(X) = \sum x \cdot P(X = x)$. The variance σ^2 measures the spread or dispersion of the distribution and is calculated as $Var(X) = E((X - \mu)^2) = \sum (x - \mu)^2 \cdot P(X = x)$.
4. **Common discrete distributions**: here are listed some famous discrete distribution functions:
   - **Bernoulli Distribution**: Models a binary outcome, such as success/failure or yes/no. Characterized by a single parameter p, the probability of success.
   - **Binomial Distribution**: Describes the number of successes in a fixed number of independent Bernoulli trials. Characterized by parameters n (number of trials) and p (probability of success).

- **Poisson Distribution**: Models the number of events occurring in a fixed interval of time or space. Characterized by a single parameter λ, the average rate of events.
- **Geometric Distribution**: Models the number of trials needed for the first success in a sequence of independent Bernoulli trials. Characterized by a single parameter p, the probability of success on each trial.

3 Properties

Let's first go through the properties of continous distributions:

1. **Infinite possible values**: Continuous random variables can take an infinite number of values within a range.
2. **Probability at specific points**: The probability of a continuous random variable taking a specific value is technically zero.
3. **Integration for probability**: probabilities are obtained by integrating the PDF over an interval.
4. **Smooth curves**: PDFs for continuous distributions are smooth curves without jumps.

Now, let's consider the properties of discrete distributions:

1. **Countable outcomes**: Discrete random variables have a countable number of possible outcomes.
2. **Probability at specific points**: The PMF gives the probability of the random variable taking specific values.
3. **Sum of probabilities**: The sum of all probabilities in the PMF equals 1.
4. **Non-negative probabilities**: Probabilities in the PMF are non-negative.
5. **Integer values**: The discrete random variables takes on integer values.

4 Simulations

Thanks to various Python libraries, I've been able to compute and display both the PDF and CDF of a Gamma distribution (continous case) and the PMF and CDF of a Geometric distribution (discrete case). Below I'll show the code as well as the graphs produced.

```python
import numpy as np
import streamlit as st
import matplotlib.pyplot as plt
from scipy.stats import gamma

def geometric_pmf(p, k):
    return (1 - p) ** (k - 1) * p

def geometric_cdf(p, k):
    return 1 - (1 - p) ** k

p = 0.5  # probability of success
k_values = np.arange(1, 11)  # range of values for k

pmf_values = [geometric_pmf(p, k) for k in k_values]
cdf_values = [geometric_cdf(p, k) for k in k_values]

plt.subplot(2, 1, 1)
plt.bar(k_values, pmf_values)
plt.xlabel('k')
plt.ylabel('PMF')
plt.title('PMF of Geometric Distribution')

plt.subplot(2, 1, 2)
plt.bar(k_values, cdf_values)
plt.xlabel('k')
plt.ylabel('CDF')
plt.title('CDF of Geometric Distribution')

plt.tight_layout()  # Adjust spacing between subplots

st.pyplot()
plt.show()

# Continous distributions: Gamma distribution
def plot_gamma_distribution(alpha, beta):
    x = np.linspace(0, 10, 100)
    pdf = gamma.pdf(x, alpha, scale=1/beta)
    cdf = gamma.cdf(x, alpha, scale=1/beta)

    plt.subplot(2, 1, 1)
    plt.plot(x, pdf)
    plt.xlabel('x')
    plt.ylabel('PDF')
    plt.title('PDF of Gamma Distribution')
```

```python

    plt.subplot(2, 1, 2)
    plt.plot(x, cdf)
    plt.xlabel('x')
    plt.ylabel('CDF')
    plt.title('CDF of Gamma Distribution')

    plt.tight_layout()  # Adjust spacing between subplots

    st.pyplot()
    plt.show()

alpha = st.slider('Alpha', 1, 10, 2)
beta = st.slider('Beta', 1, 10, 2)

plot_gamma_distribution(alpha, beta)
```

The output produced is the following, note that for the Gamma distribution I used parameters $\alpha = \beta = 3$.

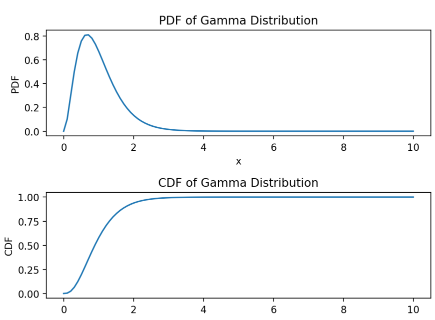

Figure 1: Gamma PDF and CDF.

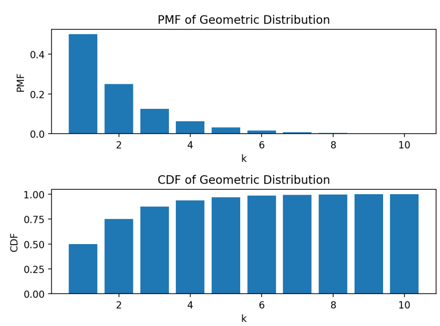

Figure 2: Geometric PMF and CDF.