Scientific Computing Assignment 3

Munirat Idris

15th March 2024

1 Question1.1

Here's a pure Python code to compute the median of an arbitrary dataset:

def median(data):
sorted$_d ata = sorted(data)$
$n = len(sorted_d ata)$
$ifnreturn(sorted_d ata[n//2-1] + sorted_d ata[n//2])/2$
$else:$
$returnsorted_d ata[n//2]$

Example usage:
dataset = [7, 2, 5, 3, 9, 8, 6, 4, 3, 2]
print("Dataset:", dataset)
print("Median:", median(dataset))

2 Question1.2

To show that the alternative way to compute the variance σ^2 given by:

$$\sigma^2 = \frac{1}{N^2}\sum_{i<j}(x_i - x_j)^2$$

is accurate, we can start by expanding the expression:

$$\sigma^2 = \frac{1}{N^2}\sum_{i<j}(x_i-x_j)^2 = \frac{1}{N^2}\left((x_1-x_2)^2 + (x_1-x_3)^2 + \ldots + (x_1-x_N)^2 + (x_2-x_3)^2 + \ldots + (x_{N-1}-x_N)^2\right)$$

Now, notice that $\sum_{i=1}^{N} x_i^2 = Nx_1^2 + Nx_2^2 + \ldots + Nx_N^2$ and $\sum_{i=1}^{N} x_i = Nx_1 + Nx_2 + \ldots + Nx_N$, so:

$$\sigma^2 = \frac{1}{N^2}\left(\sum_{i=1}^{N} x_i^2 - \frac{1}{2}\left(\sum_{i=1}^{N} x_i\right)^2\right)$$

And we know that:

$$\frac{1}{N^2}\left(\sum_{i=1}^{N} x_i^2 - \frac{1}{2}\left(\sum_{i=1}^{N} x_i\right)^2\right) = \frac{1}{N^2}\sum_{i=1}^{N}(x_i - \bar{x})^2$$

So, the alternative way to compute the variance is indeed accurate.

3 Question1.3

Here's a Python code to compute the population variance using the provided equation:

def population$_v ariance(data):$
$N = len(data)$
$sum_x i_s quared = sum([x**2 forxindata])$
$sum_x i = sum(data)$

$variance = (sum_x i_s quared - (1/N)*(sum_x i**2)/N**2)/N$

return variance

Example usage:
data = [10, 15, 20, 25, 30, 31, 89, 76, 56, 23, 87, 45, 32, 98] Sample data

variance = population$_v ariance(data)$

$print("PopulationVariance:", variance)$

This code defines the function population variance

4 Question1.4

The difference between computing the sample and population variance arises from the denominator used in the variance formula. We divide the sample variance by (n - 1) instead of n to correct for bias and provide an unbiased estimate of the population variance. This correction is known as Bessel's correction. It comes from the fact that when we compute the sample mean (x), we use the data itself, which introduces some estimation error. Dividing by (n - 1) instead of n helps to adjust for this error and provides a more accurate estimate of the population variance when working with a sample of data.

Here's a Python code to compute and plot the difference between the sample variance and the population variance as a function of the sample size:

import numpy as np
import matplotlib.pyplot as plt

def population$_variance(data):$
$N = len(data)$
$mean = np.mean(data)$
$variance = np.sum((data - mean) * *2)/N$
$returnvariance$

def sample$_variance(data):$
$n = len(data)$
$mean = np.mean(data)$
$variance = np.sum((data - mean) * *2)/(n - 1)$
$returnvariance$

Generate sample sizes from 2 to 100
sample$_sizes = np.arange(5, 1001)$
$differences = []$

Compute the difference between sample variance and population variance for each sample size
n in sample$_sizes:$
$data = np.random.randn(n)Generaterandomdata$
$population_var = population_variance(data)$
$sample_var = sample_variance(data)$
$differences.append(sample_var - population_var)$

Plot the difference as a function of sample size
plt.plot(sample$_sizes, differences)$
$plt.xlabel('SampleSize')$
$ylabel('Difference(SampleVariance - PopulationVariance)')$
$plt.title('DifferencebetweenSampleVarianceandPopulationVariance')$
$plt.grid(True)$
$plt.show()$

This code generates random data for various sample sizes, computes the sample and population variances, and then calculates the difference between the two.

5 Question1.5

To prove that the Gaussian probability distribution function (PDF) integrates to 1, we must integrate the PDF over its entire range and show that the result equals 1.

The Gaussian PDF is given by:

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

Given that = 170 and = 7, the Gaussian PDF becomes:

$$f(x) = \frac{1}{\sqrt{2\pi(7)^2}} \exp\left(-\frac{(x-170)^2}{2(7)^2}\right)$$

Now, integrating this PDF from negative infinity to positive infinity:

$$\int_{-\infty}^{\infty} f(x)\, dx$$

Substitute the expression for the Gaussian PDF:

$$\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi(7)^2}} \exp\left(-\frac{(x-170)^2}{2(7)^2}\right) dx$$

Let's denote the integrand as $g(x)$:

$$g(x) = \frac{1}{\sqrt{2\pi(7)^2}} \exp\left(-\frac{(x-170)^2}{2(7)^2}\right)$$

Using the properties of the Gaussian distribution to rewrite this integral in terms of the standard normal distribution's cumulative distribution function (CDF).

Let $z = \frac{x-\mu}{\sigma}$ be the standard normal variable, where $\mu = 170$ and $\sigma = 7$. Then, $dx = \sigma\, dz$.

Substitute into the integral:

$$\begin{aligned}
\int_{-\infty}^{\infty} g(x)\, dx &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi(7)^2}} \exp\left(-\frac{(x-170)^2}{2(7)^2}\right) dx \\
&= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi(7)^2}} \exp\left(-\frac{(z\sigma)^2}{2}\right) \sigma\, dz \\
&= \frac{1}{\sqrt{2\pi(7)^2}} \sigma \int_{-\infty}^{\infty} \exp\left(-\frac{z^2}{2}\right) dz
\end{aligned}$$

Now, the integral $\int_{-\infty}^{\infty} \exp\left(-\frac{z^2}{2}\right) dz$ is a known result from the standard normal distribution, and it equals $\sqrt{2\pi}$.

$$= \frac{1}{\sqrt{2\pi(7)^2}}\sigma \cdot \sqrt{2\pi}$$

$$= \frac{1}{\sqrt{2\pi(7)^2}} \cdot 7 \cdot \sqrt{2\pi}$$

$$= \frac{1}{\sqrt{2\pi}} \cdot \sqrt{2\pi}$$

$$= 1$$

Therefore, the integral of the Gaussian PDF over its entire range equals 1, which means that the Gaussian PDF integrates to 1.