Laboratory Assignment 3
Exploring the Normal Distribution

CSC372-M72: Optimisation

2022-23

1 Objectives.

- To generate random samples from a Normal and a truncated Normal distributions.
- To plot histograms and scatters of the samples.
- To generate samples from multi-dimensional distributions.

2 Background

A Normal, or Gaussian distribution is a continuous probability distribution for *real-valued* random variable. The probability distribution function can be defined as:

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

Here, μ is the mean of the distribution, and σ is the standard deviation. As we learned that a distribution allows us to define the behaviour of a system, as in if we query the system how it may behave. This means that, if we randomly sample from a system, and then take the average of the values produced by the system, we may find that most of the values are near the average or **mean**. The variance σ^2 is the average of squared errors from the mean across all samples, and finally, the standard deviation σ is the square root of variance, which represents the average deviation from the mean.

Generally speaking a Normal distribution is omnipresent. This is specifically true when we encounter the Central Limit Theorem, which states that, under special circumstances, the average of a random variable is itself a random variable that is Normally distributed with a mean and a standard deviation. For example, if we ask lots of random people about their heights, then the average height would somewhere around 163 cm for females and 171 cm for males, and the distribution of their heights would be roughly Normally distributed; see the following link for more information on this data: `ourworldindata.org/human-height`.

If we were to truncate a Normal distribution with lower and upper bound, we can do so by redefining the probability distribution function as follows:

$$\rho(x) = \begin{cases} \frac{1}{\sigma\sqrt{2\pi}} \exp^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}. & \text{for } l_i \leq x \leq u_i \\ 0 & \text{otherwise.} \end{cases} \tag{2}$$

In this lab, we are going to investigate the Normal distribution: in particular, we will concentrate on how to sample from uni- and multi-dimensional distributions, and when they are truncated.

3 Tasks

1. For a mean $\mu = 163$ and and a standard deviation of $\sigma = 10$, generate 10000 samples from the Normal distribution. You are given that this is a distribution over height h.

   > Hint: You may want to take a look at the output of the following code:
   >
   > `print(scipy.stats.norm.__doc__)`

2. Plot the histogram of the samples with appropriate axes labels. You should think about the following queries:

   - What should be the label for the horizontal axis?
   - What should be the label for the vertical axis?

   > Hint: You may want to take a look at the output of the following code:
   >
   > `print(matplotlib.pyplot.hist.__doc__)`

3. Now, code a Python function for the equation (1) that accepts three parameters: x, μ and σ. Using the following array `x = numpy.linspace(mu - 3 * sigma, mu + 3 * sigma, 1000)`, evaluate your function, and plot the response. You should think about the following queries:

   - What should be the label for the horizontal axis?
   - What should be the label for the vertical axis?
   - What does this plot represent?

4. For a mean $\mu = 100$ and and a standard deviation of $\sigma = 15$, generate 10000 samples from the Normal distribution like before, and plot the relevant histogram with appropriate labels. You are given that this is a distribution over intelligence quotient (IQ) q.

   > Please note there is no reason why h and q should be related, i.e. knowing what someone's height is should give you no information about their IQ. So, we can say that these quantities are independent of each-other.

5. Generate a scatter plot (see `matplotlib.pyplot.scatter.__doc__` for more information) from he samples where the horizontal axis is the height h and the vertical axis is the intelligence quotient q.

6. Study this discussion about the truncated Normal distribution. Based on your analyses, and what you have done thus far, perform the following tasks:

   (a) Now, generate a set of 1000 two-dimensional random vectors with the following specifications:

Mean vector: $\boldsymbol{\mu} = (163, 100)^\top$
Standard deviation vector: $\boldsymbol{\sigma} = (10, 15)^\top$
Lower bound: $\mathbf{l} = (140, 80)^\top$
Upper bound: $\mathbf{u} = (180, 150)^\top$

(b) Generate a scatter plot of the samples from task 6a.