Exercise 1

(1)
$$p(x|\lambda) = \exp(-\lambda)\frac{\lambda^x}{x!} \quad , \quad x \in \mathbb{N}$$

→ Likelyhood function:

$$L(\lambda; x_1, \dots, x_N) = \prod_{j=1}^{N} \exp(-\lambda)\frac{\lambda^{x_j}}{x_j!}$$

→ Log-Likelyhood function:

$$\mathcal{L}(\lambda; x_1, \dots, x_N) = \ln\left[\prod_{j=1}^{N} \exp(-\lambda)\frac{\lambda^{x_j}}{x_j!}\right]$$

$$= \sum_{j=1}^{N} \ln\left(\exp(-\lambda)\frac{\lambda^{x_j}}{x_j!}\right)$$

$$= \sum_{j=1}^{N}\left[\ln(\exp(-\lambda)) - \ln(x_j!) + \ln(\lambda^{x_j})\right]$$

$$= \sum_{j=1}^{N}\left[-\lambda - \ln(x_j!) + x_j \cdot \ln(\lambda)\right]$$

$$= -N\lambda - \sum_{j=1}^{N}\ln(x_j!) + \ln(\lambda)\cdot\sum_{j=1}^{N} x_j$$

→ maximum likelyhood solution:

$$\hat{\lambda} = \underset{\lambda}{\operatorname{argmax}}\left(\mathcal{L}(\lambda; x_1, \dots, x_N)\right)$$

$$\frac{\partial}{\partial\lambda}\mathcal{L}(\lambda; x_1, \dots, x_N) = 0 = \frac{\partial}{\partial\lambda}\left(-N\lambda - \sum_{j=1}^{N}\ln(x_j!) + \ln(\lambda)\cdot\sum_{j=1}^{N} x_j\right)$$

$$0 = -N + \frac{1}{\lambda}\sum_{j=1}^{N} x_j \quad \Big|:N\Big|\cdot\lambda$$

$$0 = -\lambda + \frac{1}{N}\sum_{j=1}^{N} x_j$$

Poisson distribution:

$$\Leftrightarrow \quad \hat{\lambda} = \frac{1}{N}\sum_{j=1}^{N} x_j = \text{sample mean} \rightarrow \lambda = E(X) = Var(X)$$

Exercise 2

Task A

$f(x)$ is said to be a probability density function (PDF) if:

1. $f(x) \geq 0$ for all x
2. $\int_{-\infty}^{\infty} f(x)\mathrm{d}x = 1$.

1. Gaussian PDFs $N(x;\mu,\sigma)$ are ≥ 0 for all x values so the sum of two Gaussian PDFs is ≥ 0 for all x values.

2. $\int_{-\infty}^{\infty} p(x|\mu_1,\mu_2,\sigma_1,\sigma_2)\mathrm{d}x = \frac{1}{2}\int_{-\infty}^{\infty} N(x;\mu_1,\sigma_1)\mathrm{d}x + \frac{1}{2}\int_{-\infty}^{\infty} N(x;\mu_2,\sigma_2)\mathrm{d}x$.

The integral of each Gassian PDF is 1 therefore:

$$\int_{-\infty}^{\infty} p(x|\mu_1,\mu_2,\sigma_1,\sigma_2)\mathrm{d}x = \frac{1}{2}*1 + \frac{1}{2}*1 = 1.$$

Task B

A PDF can my maximised by maximising the log likelyhood $L(\Theta)$ of the PDF with

$$L(\Theta) = \sum_{n=1}^{N} \log(p(x^{(n)}|\Theta)).$$

To maximise $L(\Theta)$ the first derivative is equated to 0. The log likelyhood can be simplified easily if the PDF is a single Gaussian as $\log(ae^x) = \log(a) + x$. This simplification cannot be applied to $p(x|\mu_1,\mu_2,\sigma_1,\sigma_2)$ as it consists of the sum of two Gaussian PDFs.

$$L(\Theta) = \sum_{n=1}^{N} \log(p(x|\mu_1,\mu_2,\sigma_1,\sigma_2)) = \sum_{n=1}^{N} \log(\frac{1}{2}N(x;\mu_1,\sigma_1)\mathrm{d}x + \frac{1}{2}N(x;\mu_2,\sigma_2)\mathrm{d}x).$$

Therefore the derivative cannot be calculated as easily.

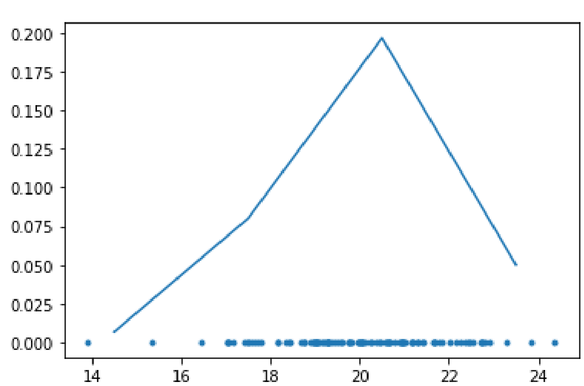
Exercise 3

Task A

A dataset of $N = 100$ normally distributed datapoints is drawn via the *numpy.random.normal()* function, with the mean $\mu = 20$ and the variance $\sigma^2 = 4$. Subsequently, the probability distribution is approximated via the sampling formula

$$p(x_i) = \frac{n_i}{N \cdot \Delta_i}$$

Below, the sample is visualized via a scatterplot and via a probability distribution approximated via the sampling formula.



Task B

The following steps are taken to compute the max likelihood parameters of the distributed datapoints.

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

$$\begin{aligned}
\mathcal{L}(\Theta) &= \sum_{i=1}^{N} \log\left(\mathcal{N}(x_i;\mu,\sigma^2)\right) = \sum_{i=1}^{N} \log\left(\frac{1}{\sqrt{2\pi\sigma^2}} \cdot \exp\left(-\frac{(x_i-\mu)^2}{2\sigma^2}\right)\right) \\
&= \log\left(\frac{1}{(2\pi\sigma^2)^{N/2}} \cdot \exp\left(-\frac{\sum_{i=1}^{N}(x_i-\mu)^2}{2\sigma^2}\right)\right) \\
&= \log\left(\frac{1}{(2\pi\sigma^2)^{N/2}}\right) + \log\left(\exp\left(-\frac{\sum_{i=1}^{N}(x_i-\mu)^2}{2\sigma^2}\right)\right) \\
&= -\frac{N}{2}\log(2\pi\sigma^2) + \left(-\frac{\sum_{i=1}^{N}(x_i-\mu)^2}{2\sigma^2}\right)
\end{aligned}$$

Now take derivative of $\mathcal{L}(\Theta)$ with respect to μ, σ^2 and set equal 0

$$\begin{aligned}
\frac{\partial \mathcal{L}(\Theta)}{\partial \mu} = \frac{\partial}{\partial \mu}\left(-\frac{N}{2}\log(2\pi\sigma^2) + \left(-\frac{\sum_{i=1}^{N}(x_i-\mu)^2}{2\sigma^2}\right)\right) &\overset{!}{=} 0 \\
\frac{\partial}{\partial \mu}\left(-\frac{\sum_{i=1}^{N}(x_i-\mu)^2}{2\sigma^2}\right) &= 0 \\
\frac{1}{\sigma^2} sum_{i=1}^{N}(x_i-\mu) &= 0 \\
\mu &= \frac{1}{N}\sum_{i=1}^{N} x_i.
\end{aligned}$$

$$\begin{aligned}
\frac{\partial \mathcal{L}(\Theta)}{\partial \sigma^2} = \frac{\partial}{\partial \sigma^2}\left(-\frac{N}{2}\log(2\pi\sigma^2) + \left(-\frac{\sum_{i=1}^{N}(x_i-\mu)^2}{2\sigma^2}\right)\right) &\overset{!}{=} 0 \\
\frac{\partial}{\partial \sigma^2}\left(-\frac{N}{2}\log(2\pi)\right) + \frac{\partial}{\partial \sigma^2}\left(-\frac{N}{2}\log(\sigma^2)\right) + \frac{\partial}{\partial \sigma^2}\left(-\frac{\sum_{i=1}^{N}(x_i-\mu)^2}{2\sigma^2}\right) &= 0 \\
-\frac{N}{2}\frac{1}{\sigma^2} + \left(\sum_{i=1}^{N}(x_i-\mu)^2 \cdot \frac{1}{2(\sigma^2)^2}\right) &= 0 \\
\frac{1}{2\sigma^2}\left(\sum_{i=1}^{N}(x_i-\mu)^2 \cdot \frac{1}{\sigma^2} - \frac{N}{2}\right) &= 0 \\
\sum_{i=1}^{N}(x_i-\mu)^2 \cdot \frac{1}{\sigma^2} - \frac{N}{2} &= 0 \\
\sigma^2 &= \frac{1}{N}\sum_{i=1}^{N}(x_i-\mu)^2.
\end{aligned}$$

From the max likelihood approach we get the same formulas as for the sample mean and variance. For the generated sample, they are computed as:

$$\mu* = 19.99,\ \sigma* = 3.37$$

Task C

No the generating parameters are not recovered exactly, as the amount of datapoints is too little for a good estimation of the statistical quantities.

Task D

For smaller N the dispersion of the recovered mean and variance around the generating parameters increases, whereas for larger N they tend to be closer to the generating parameters.