6.

Intro

In this question we have to derive the maximum likelihood estimator λ_{MLE} for the Poisson distribution. Since the distribution has the e number and is composed by the multiplication of three terms, the use of the log likelihood will be very helpful. It is well known that the expected value of the Poisson distributions is equal to the rate parameter ($E[X] = \lambda$). So, it would be make sense if our MLE had something to do with the expected value.

Solution

The MLE of the Poisson distribution is the rate parameter λ_{MLE} that maximizes the likelihood (or the log likelihood). The log likelihood for the Poisson distribution is given by:

$$\begin{aligned} &l(\lambda) = \sum_1^N log(p(x_i|\lambda)) = \sum_1^N log(e^{-\lambda}\frac{\lambda^{x_i}}{x_i!}) = \\ &\sum_1^N -\lambda + \sum_1^N log\left(\frac{\lambda^{x_i}}{x_i!}\right) = -N\lambda + log(\lambda)\sum_1^N x_i - \sum_1^N log(x_i!) \end{aligned} \tag{1}$$

Differentiating (1) and equaling to zero, we get:

$$\begin{aligned} &\frac{dl}{d\lambda} = -N + \frac{1}{\lambda}\left(\sum_1^N x_i\right) = 0 \\ &\lambda_{MLE} = \frac{\sum_1^N x_i}{N} \end{aligned} \tag{2}$$

Therefore, the MLE is given by: $\lambda_{MLE} = \frac{\sum_1^N x_i}{N}$.

Conclusion

The MLE for the poisson distribution is $\lambda_{MLE} = \frac{\sum_1^N x_i}{N}$. In another words, the MLE is the sample average of our dataset, which makes a lot of sense, based on our previous discussion in the Intro. Remember that by the law of large numbers, the sample average converges almost surely to the expect value. What this means is that the if have a lot of points in our dataset, our MLE is a good estimator for the parameter of the distribution. We don't even have to go Bayesian this time!

Regarding the details of the optimization, since $\lambda > 0$ by definition and the log likelihood is a function of λ that starts out increasing (while $log(\lambda)$ is more relevant than $-N\lambda$), then goes decreasing indefinitely (when $-N\lambda$ becomes more relevant than $log(\lambda)$), we have certainty that (2) gives us a global maximum.