WEEKLY PARTICIPATION 3: MLE FOR POISSON REGRESSION

In class we have covered noise models that help us to model regression, binary classification, and multiclass classification models by modeling the "trend" with a parameterized machine learning model, using the noise model to model the likelihood of deviations from that trend, and then using MLE to learn the parameters of the model from data.

In this participation, we introduce the Poisson noise model that is commonly used for counting problems (where the count can be arbitrarily large, so we cannot use multiclass classification). Example applications include:

- Modeling the number of traffic accidents occurring at a specific intersection or along a stretch of road in a given period.
- Analyzing the number of new cases of a particular disease in different regions or over time, where the rate of occurrence might vary by location or demographic factors.
- Predicting the number of attendees at events or meetings based on factors like day of the week, weather conditions, or location.

A discrete nonnegative random variable z follows the Poisson distribution with parameter $\lambda \geq 0$, written $z \sim \text{Poisson}(\lambda)$, if its pmf is given by

$$p_z(k) = \mathrm{e}^{-\lambda}\frac{\lambda^k}{k!}, \quad \text{where} \quad k \in \{0, 1, 2, \ldots\}.$$

The mean of this distribution is λ, and the variance is also[1] λ. See the figure for a visualization of three Poisson pmfs.

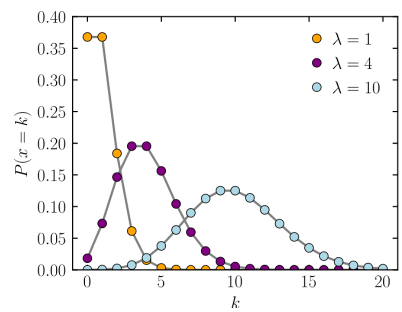

FIGURE 1. Poisson pmfs for different parameters. Taken from wikipedia.

Consider the problem of predicting the number of students present in lecture. We can predict this quantity more accurately if we take into account the time of year (are we before or after the drop date?), the weather, and so on. We will use Poisson regression to model this conditional dependence.

In Poisson regression, the target y is a nonnegative integer, and we model its dependence on the predictors $\mathbf{x}$ using a Poisson distribution:

$$y \,|\, \mathbf{x} \sim \text{Poisson}(\exp(\boldsymbol{\theta}^T \mathbf{x})).$$

The regression function $\lambda(\mathbf{x}) = \mathbb{E}[y \,|\, \mathbf{x}] = \exp(\boldsymbol{\theta}^T \mathbf{x})$ models the expected value of y as well as the variance of y, given the predictors $\mathbf{x}$.

The questions this week concern practical aspects of fitting and using Poisson regression models:

- Give the expression for $p_{\boldsymbol{\theta}}(y_i \,|\, \mathbf{x}_i)$.
- Given a fitted Poisson regression model $\boldsymbol{\theta}$ and a new data point $\mathbf{x}_{\text{new}}$, how would you predict the corresponding target y?
- State, in as simple a form you can manage, the optimization problem for finding an estimate $\hat{\boldsymbol{\theta}}$ of $\boldsymbol{\theta}$ by using the maximum likelihood principle for Poisson regression.

[1] There are other models for count data (e.g. negative binomial) that can model over- and under-dispersion, where the variance is smaller or greater than the mean.