Romi Goodman and Ophere Evan

Part 1: Theoretical Exercise (16 points)

Maximum Likelihood Estimation and Confidence Intervals

The Poisson distribution is a discrete porbability distribution over the non-negative integers with a single parameter $\lambda > 0$ and PMF:

$$\Pr[X = k] = \frac{\lambda^k e^{-\lambda}}{k!}$$

This distribution is useful in modeling the number of event occurrences in a fixed time interval when the probability of an event occurrence does not depend on the time since the last event. The parameter of this distribution, λ, is the expected number of events in one time interval. As a result, the expectation and variance of a $\mathrm{Pois}(\lambda)$ random variable are both λ.

Suppose that you have a set of samples $D = \{x_1, \ldots, x_n\}$. We propose a probability model in which these samples are taken from $X_1, \ldots, X_n$, which are independent $\mathrm{Pois}(\lambda)$ random variables.

1. Write the **log-likelihood** function of λ under D. Use natural log here (base e).

The log-liklihood function :
To write the log likelihood function of λ for a set of i.i.d samples $D = \{x1, x2,...., xn\}$ drawn from a poisson distribution, we first need to define the likelihood function and than take its natural logarithm.

H_0 defines the probability of observing data D given a parameter vector Θ. for i.i.d samples the likelihood function $L(\Theta; D)$ is the product of the probabilities of observing each individual sample:

$L(\Theta; D) = \prod_{i=1}^{n} p_{\Theta}(x_i)$

So in this case, our $\Theta = \lambda$

So the PMF for a poisson distribution is given as:

$p_{\lambda}(x_i) = Pr[X = x_i] = (\lambda^{x_i} e^{-\lambda})/(x_i!)$

So the likelihood function for the given sample D (under poisson model) is:

$L(\lambda; D) = \prod_{i=1}^{n} \frac{\lambda^{x_i} e^{-\lambda}}{x_i!}$

The log-likelihood function, $l(\lambda; D)$ is obtained by taking the natural logarithm of the likelihood function.

Hence:

$l(\lambda; D) = log(\prod_{i=1}^{n} \frac{\lambda^{x_i} e^{-\lambda}}{x_i!}) = \sum_{i=1}^{n} log(\frac{\lambda^{x_i} e^{-\lambda}}{x_i!}) = \sum_{i=1}^{n} \frac{log(\lambda^{x_i} e^{-\lambda})}{log(x_i!)} = \sum_{i=1}^{n} log(\lambda^{x_i}) + log(e^{-\lambda}) - log(x_i!) =$

$\sum_{i=1}^{n} log(\lambda^{x_i}) - \lambda - log(x_i!) = \sum_{i=1}^{n} log(\lambda^{x_i}) - \sum_{i=1}^{n} \lambda - \sum_{i=1}^{n} log(x_i!) =$

$(\sum_{i=1}^{n} x_i) \, log(\lambda) \; - n\lambda - \sum_{i=1}^{n} log(x_i!)$

And as we were asked to use natural logarithm:

$(\sum_{i=1}^{n} x_i) \, ln(\lambda) \; - n\lambda - \sum_{i=1}^{n} ln(x_i!)$

2. Use the **log-likelihood** function you derived above to find an expression for the maximum likelihood estimator (MLE) $\hat{\lambda}$ of λ.

To find the MLE $\hat{\lambda}$ we can take the derivative of the log=likelihood function with respect to λ , set it to 0 and solve for λ. and this way we will find the $\lambda = \theta$ that maximizes the function.

The log-likelihood function of λ for a set of samples $D = \{x1, x2,...., xn\}$ from a poisson distribution was derived as $(\sum_{i=1}^{n} x_i) \, ln(\lambda) \; - n\lambda - \sum_{i=1}^{n} ln(x_i!)$.

$\frac{\partial l}{\partial \lambda} = \frac{\partial}{\partial \lambda} ((\sum_{i=1}^{n} x_i) \, ln(\lambda) \; - n\lambda - \sum_{i=1}^{n} ln(x_i!)) = 0$

So now we'll differentiate each term:

$\frac{\partial}{\partial \lambda} (\sum_{i=1}^{n} x_i) ln(\lambda) = (\sum_{i=1}^{n} x_i) \frac{1}{\lambda}$

$\frac{\partial}{\partial \lambda} - n\lambda = - n$

$\frac{\partial}{\partial \lambda} \sum_{i=1}^{n} ln(x_i!) = 0$

Therefore we get:

$\frac{\partial}{\partial\lambda} l(\lambda; D) = (\sum_{i=1}^{n} x_i)(\frac{1}{\lambda}) - n$

So: $0 = (\sum_{i=1}^{n} x_i)(\frac{1}{\lambda}) - n \rightarrow n = (\sum_{i=1}^{n} x_i)(\frac{1}{\lambda}) \rightarrow \frac{n}{\sum_{i=1}^{n} x_i} = \frac{1}{\lambda} \rightarrow \lambda = \frac{\sum_{i=1}^{n} x_i}{n}$

So we get that $\hat{\lambda} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{1}{n}\sum_{i=1}^{n} x_i$

We now wish to find an expression for an intervalcontaining λ with probability at least 0.95 under the model we proposed. Namely, a 95% confidence interval (CI) for λ (around the estimated $\hat{\lambda}$) .

3. Consider $\hat{\lambda}$ as a random variable whose randomness stems from the randomness in $X_1, \dots, X_n$. Write an expression for the mean $\mathbb{E}[\hat{\lambda}]$ and the variance $\mathbf{Var}[\hat{\lambda}]$ as a function of n and λ.

We need to find an expression for an interval containing λ with 95% CI. that means we need to construct a 95% confidence interval for λ around the estimated $\hat{\lambda}$.

For a discrete random variable X, the expected value $E(x)$ is defined as the sum of each possible value multiplied by its probability. A key property of the Poisson distribution is that its mean is equal to λ. so for every $1 \leq i \leq n$, $E(x_i) = \lambda$

So we now need to compute $E(\hat{\lambda}) = E[\frac{1}{n}\sum_{i=1}^{n} x_i] \rightarrow \frac{1}{n}\sum_{i=1}^{n} E(x_i)$

So we sub $E(x_i) = \lambda$

$= E(\hat{\lambda}) = \frac{1}{n}\sum_{i=1}^{n} \lambda = \frac{1}{n} n\lambda = \lambda$

So now we find the variance:
The variance of a random variable X is defined as $Var(X) = E[(X - E[X])^2]$
Since X_1, ..., X_n are independent $Pois(\lambda)$ random variables, each X_i has a variance $Var[X_i] = \lambda$.

So we get: $Var[\hat{\lambda}] = Var[\frac{1}{n}\sum_{i=1}^{n} x_i]$

We use $Var[cX] = c^2 Var[X]$ and the fact that for i.i.d variables we get $Var[\Sigma X_i] = \Sigma Var[X_i]$
So we get :

$Var[\hat{\lambda}] = (\frac{1}{n})^2 Var[\sum_{i=1}^{n} x_i] = (\frac{1}{n^2}) \sum_{i=1}^{n} Var[x_i] - (Var[x_i] = \lambda) \rightarrow (\frac{1}{n^2}) \sum_{i=1}^{n} \lambda = \frac{1}{n^2} n\lambda = \frac{\lambda}{n}$

So $Var[\hat{\lambda}] = \frac{\lambda}{n}$ and $E(\hat{\lambda}) = \lambda$

4. Assume that your estimator $\hat{\lambda}$ is asymptotically normal, meaning that its **standardized** value apporaches a normal distribution as the number of data points goes to infinity:

$$\frac{\hat{\lambda} - \mathbb{E}[\hat{\lambda}]}{\sqrt{\mathrm{Var}[\hat{\lambda}]}} \sim \mathcal{N}(0, 1). \tag{1}$$

(Note: this is implied by the central limit theorem.)

Under this assumption, write expressions for the lower and upper boundaries of the 95% confidence interval for λ around the estimated $\hat{\lambda}$. Your two expressions should be specified as functions of $\hat{\lambda}$, n, and the inverse Gaussian CDF $\Phi^{-1}(p)$.

Given that the standardized value of $\hat{\lambda}$ follows a standard normal distribution for large n:

$Z = \frac{(\hat{\lambda} - E(\hat{\lambda}))}{\sqrt{(Var[\hat{\lambda}])}} \sim N(0, 1)$

So given our mean and variance we get $Z = \frac{\hat{\lambda} - \lambda}{\sqrt{\frac{\lambda}{n}}}$

So now we need to find for which z we get $P(-z \leq Z \leq z) = 0.95$

We know that $P(Z \leq z) = \phi(z)$ and $P(Z \leq -z) = \phi(-z)$

We know that the normal distribution is symmetric around 0 - so $\phi(-z) = 1 = \phi(z)$

So we get $P(-z \leq Z \leq z) = \phi(z) - (1 - \phi(z)) = 2\phi(z) - 1$

So setting $2\phi(z) - 1 = 0.95 \rightarrow 2\phi(z) = 1.95 \rightarrow \phi(z) = 0.975$

So we've seen in class that its 1.96

So:

$- 1.96 \leq (\frac{\hat{\lambda} - \lambda}{\sqrt{\frac{\hat{\lambda}}{n}}}) \leq 1.96 \rightarrow - 1.96(\sqrt{\frac{\hat{\lambda}}{n}}) \leq \hat{\lambda} - \lambda \leq 1.96(\sqrt{\frac{\hat{\lambda}}{n}}) \rightarrow - 1.96(\sqrt{\frac{\hat{\lambda}}{n}}) - \hat{\lambda} \leq \lambda \leq 1.96(\sqrt{\frac{\hat{\lambda}}{n}}) - \hat{\lambda}$

$= \hat{\lambda} - 1.96(\sqrt{\frac{\hat{\lambda}}{n}}) \leq \lambda \leq \hat{\lambda} + 1.96(\sqrt{\frac{\hat{\lambda}}{n}})$

So we get that the interval is:

$[\hat{\lambda} - 1.96(\frac{\sqrt{\hat{\lambda}}}{n}), \hat{\lambda} + 1.96(\frac{\sqrt{\hat{\lambda}}}{n})]$