X_i is the number of seedlings observed in quadrat #i.

$X_1, \dots, X_n \overset{iid}{\sim} \text{Poisson}(\lambda)$

Goal: a confidence interval for λ based on the large-sample normal approx. to the sampling distribution of

$$\hat{\Lambda}^{MLE} = \bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

The normal approx. is:

$$\underline{\hat{\Lambda}^{MLE}} \sim \text{Normal}\left(\lambda, \frac{1}{\underbrace{J(\hat{\lambda}^{MLE})}}\right)$$

observed Fisher information evaluated at $\hat{\lambda}^{MLE}$ (estimate based on observed sample data)

$$\underline{J(\lambda^*)} = -\frac{d^2}{d\lambda^2} \log\left\{\prod_{i=1}^{n} f_{X_i}(x_i \mid \lambda)\right\}\Bigg|_{\lambda=\lambda^*}$$

$$= -\frac{d^2}{d\lambda^2} \sum_{i=1}^{n} \log\left\{\frac{e^{-\lambda}\lambda^{x_i}}{x_i!}\right\}\Bigg|_{\lambda=\lambda^*}$$

$$= -\frac{d^2}{d\lambda^2} \sum_{i=1}^{n} \left\{-\lambda + x_i \log(\lambda) - \log(x_i!)\right\}\Bigg|_{\lambda=\lambda^*}$$

$$= -\frac{d}{d\lambda} \sum_{i=1}^{n} \left\{-1 + \frac{x_i}{\lambda}\right\}\Bigg|_{\lambda=\lambda^*}$$

$$= \cancel{-} \sum_{i=1}^{n} \cancel{-}\frac{x_i}{\lambda^2}\Bigg|_{\lambda=\lambda^*}$$

$$= \frac{1}{(\lambda^*)^2} \cdot \sum_{i=1}^{n} x_i$$

Approximately, if n large,

$$\hat{\Lambda}^{MLE} \sim \text{Normal}\left(\lambda, \frac{1}{\frac{1}{\bar{x}^2} n \cdot \underbrace{\frac{1}{n}\sum_{i=1}^{n} x_i}}\right)$$

$$\boxed{\hat{\Lambda}^{MLE} \sim \text{Normal}\left(\lambda, \frac{\bar{x}}{n}\right)}$$

Need a pivotal quantity:

- can involve data and parameters
- its distribution can't depend on unknown parameters

Many options; most common:

$$\frac{\hat{\lambda}^{MLE} - \lambda}{\sqrt{\bar{x}/n}} \sim \underline{\text{Normal}(0,1)}$$

(same thing)

① $$\boxed{\frac{\bar{X} - \lambda}{\sqrt{\bar{x}/n}}} \sim \underbrace{\text{Normal}(0,1)}$$

This is an approximation

The pivotal quantity in green does not exactly follow a normal distribution

② If n large,

$$P\left(z\left(\frac{\alpha}{2}\right) \le \frac{\hat{\lambda}^{MLE} - \lambda}{\sqrt{\bar{x}/n}} \le z\left(1-\frac{\alpha}{2}\right)\right) \approx 1-\alpha$$

← because based on a large-sample normal approx.

where $z(\frac{\alpha}{2})$ and $z(1-\frac{\alpha}{2})$ are the $\frac{\alpha}{2}$ and $1-\frac{\alpha}{2}$ quantiles of a Normal(0, 1) distribution.

③ Continuing from above:

$$P\left(z\left(\tfrac{\alpha}{2}\right)\sqrt{\tfrac{\bar{x}}{n}} \le \hat{\Lambda}^{MLE} - \lambda \le z\left(1-\tfrac{\alpha}{2}\right)\sqrt{\tfrac{\bar{x}}{n}} \right) \approx 1-\alpha$$

$$\Rightarrow P\left(-\hat{\Lambda}^{MLE} + z\left(\tfrac{\alpha}{2}\right)\sqrt{\tfrac{\bar{x}}{n}} \le -\lambda \le -\hat{\Lambda}^{MLE} + z\left(1-\tfrac{\alpha}{2}\right)\sqrt{\tfrac{\bar{x}}{n}} \right) \approx 1-\alpha$$

$$\Rightarrow P\left(\hat{\Lambda}^{MLE} - z\left(\tfrac{\alpha}{2}\right)\sqrt{\tfrac{\bar{x}}{n}} \ge \lambda \ge \hat{\Lambda}^{MLE} - z\left(1-\tfrac{\alpha}{2}\right)\sqrt{\tfrac{\bar{x}}{n}} \right) \approx 1-\alpha$$

$$\Rightarrow P\left(\hat{\Lambda}^{MLE} - z\left(1-\tfrac{\alpha}{2}\right)\sqrt{\tfrac{\bar{x}}{n}} \le \lambda \le \hat{\Lambda}^{MLE} - z\left(\tfrac{\alpha}{2}\right)\sqrt{\tfrac{\bar{x}}{n}} \right) \approx 1-\alpha$$

So an <u>approximate</u> $(1-\alpha)*100\%$ C.I. for λ is:

$$\left[\hat{\Lambda}^{MLE} - z\left(1-\tfrac{\alpha}{2}\right)\sqrt{\tfrac{\bar{x}}{n}},\ \hat{\Lambda}^{MLE} - z\left(\tfrac{\alpha}{2}\right)\sqrt{\tfrac{\bar{x}}{n}} \right]$$

This will not have a actual confidence level of $(1-\alpha)*100\%$ because it's based on a normal approximation to the distribution of the pivotal quantity, not the actual distribution of

$$\frac{\hat{\Lambda}^{MLE} - \lambda}{\sqrt{\bar{x}/n}} \; !$$

Def.: The nominal confidence level (or the nominal coverage rate) of a confidence interval is the claimed proportion of samples for which the interval contains the actual value of the parameter being estimated.

$$(1-\alpha) * 100\%$$

Comment: If the confidence interval was derived based on an approximate sampling distribution, the actual coverage rate (or confidence level) may be different from the nominal coverage rate.