KOIDALA SURYA PRAKASH : EE18BTECH11026
TATIPELLY VAMSHI : ES18BTECH11021

QUESTION : 03

SETUP

We considered the first 13 years as training data, aimed to fit a Poisson distribution over the training data. After learning the distribution, we tried to predict and compute the RMSE lossover the testing data set(i.e for the next 7 years). We showcased the errors accordingly. Features are the corps(14 features or dimensions). In order to fit the data to a distribution, we need to find the parameters of the distribution from the statistics of the data. Here λ is the parameter to be found. It is determined through MLE and MAP estimation techniques.

3.1 MAXIMUM LIKELIHOOD ESTIMATION

It turns out that for MLE, the parameter **λ** is the mean of the data over each feature (i.e; corps).
NOTE : We are trying to fit the data of each feature(corps) to a different Poisson distribution,this will lead to each corps having different **λ**.

Proof is attached here :

Problem 3.1

3.1. Using ML estimation to learn parameters

For poisson distribution

$$P(x/\lambda) = \frac{e^{-\lambda}(\lambda)^x}{x!}$$

↳ parameter 'λ'.

Estimating 'λ' from MLE :

Likelihood.

$$L(\lambda) = P(\cancel{y/\lambda}) = \prod_{i=1}^{n} P(y_i/\lambda)$$

$$= \prod_{i=1}^{n} \frac{e^{-\lambda}(\lambda)^{y_i}}{y_i!} = \frac{e^{-N\lambda}(\lambda)^{\sum y_i}}{\prod y_i!}$$

considering log-likelihood does not change anything.

$$J(\lambda) = \log(L(\lambda)) = \sum_{i=1}^{n} -\lambda + y_i \log(\lambda) - \log(y_i!)$$

To find the best 'λ' ⇒ need to minimise the log-likelihood

$$\Rightarrow \boxed{\frac{\partial J(\lambda)}{\partial \lambda} = 0}$$

$$\frac{\partial J}{\partial \lambda} = \sum_{i=1}^{n} -1 + \frac{x_i}{\lambda} = 0$$

$$\Rightarrow \quad -N + \frac{\sum_{i=1}^{n} x_i}{\lambda} = 0$$

$$\Rightarrow \quad \boxed{\lambda = \frac{\sum x_i}{N}}$$

$\therefore$ 'λ' is the mean of the data, in MLE.

PARAMETER ESTIMATION

$$lamda = \begin{bmatrix} 1.0 \\ 0.692 \\ 0.615 \\ 0.615 \\ 0.462 \\ 0.385 \\ 0.846 \\ 0.538 \\ 0.308 \\ 0.692 \\ 0.538 \\ 1.0 \\ 1.462 \\ 0.308 \end{bmatrix}$$

Observations : We can see that all the parameters are close to one, this is evident since most values are 0 or 1.

RMSE Calculations

Once we estimate the parameters, we can assume the predictions to be $\boldsymbol{\lambda}$,but it should bea count value(integer). Hence We rounded it and considered them as my predictions for the future. We calculated the RMSE error over the testing data using my estimates for each corps.

Here are the results :

$$Predictions = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 0 \\ 0 \\ 1 \\ 1 \\ 0 \\ 1 \\ 1 \\ 1 \\ 1 \\ 0 \end{bmatrix}, \qquad RMSE = \begin{bmatrix} 1.613 \\ 1.418 \\ 1.588 \\ 1.522 \\ 1.005 \\ 0.851 \\ 1.467 \\ 1.535 \\ 1.035 \\ 1.502 \\ 1.411 \\ 1.233 \\ 1.403 \\ 1.129 \end{bmatrix}$$

3.2 MAXIMUM A POSTERIORI ESTIMATION(MAP)

MAP estimation follows the principle of Bayes theorem, we will assume a prior distribution (Belief) across the wanted parameters (here λ) and the likelihood (Evidence), which is obtained from the data. We then use both and come up with the estimate.

3.2.a Assuming prior over `λ as gamma distribution` *and justification for it*

3.2 MAP estimation

⇒ a) Here, λ assumed a prior distribution of 'λ' parameter of each corps as a 'γ'-distribution.

→ Reasons:-

* ~~Let the~~ By seeing the data, we get a rough idea that values at every corps lie in $\{0, 1, 2, 3, 4\}$ values.

→ Where, we see that ~~corps~~ no. of deaths for each corps is closer to '0'

i.e; $P(0) > P(1) > P(2) > P(3) > P(4)$

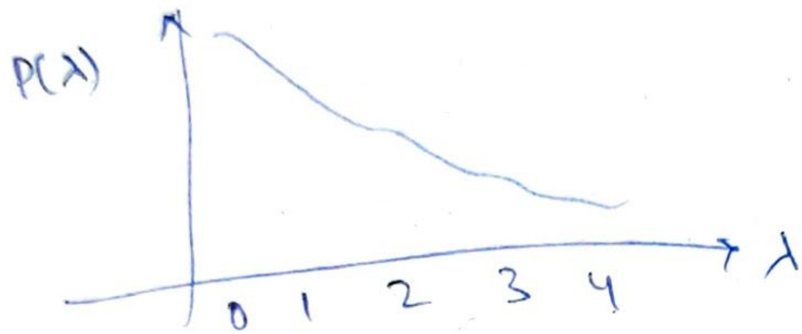
↳ Since, we are fitting a poisson distribution, whose mode (high probable value) is floor(λ); λ ↳ mean.

→ ∴ Predictions are same as 'λ' parameter

Hence, we assume that λ is

~~more~~ such that it will be

be equal to $\{0, 1, 2, 3, 4 \ldots\}$ with decreasing probability. Something like this



$\Rightarrow$ Mode of 'λ' $\rightarrow 0$

Variance should be more

For γ distribution

$$\text{mode} = \frac{\alpha - 1}{\beta} \rightarrow 0 \qquad \text{var} = \alpha/\beta^2$$

$\therefore \alpha \rightarrow 1,$

$\beta < 1 \rightarrow$ for variance.

(ii) Since counts $\geq 0 \Rightarrow \lambda \geq 0$.

→ For Gamma distribution:, this fits exactly.

$$\therefore P(\lambda/\alpha,\beta) = \frac{\beta^{\alpha}}{\Gamma(\alpha)} \lambda^{\alpha-1} e^{-\beta\lambda} \; ; \; \underline{\underline{\lambda > 0}}$$

(iii) Considering prior as Gamma, will lead to the aposterior being gamma as well.

⇒ Proof: $P(\lambda|\alpha,\beta) = \frac{\beta^{\alpha}}{\Gamma(\alpha)} \lambda^{\alpha-1} \cdot e^{-\beta\lambda} \; ; \; \lambda > 0$

Likelihood: $P(y/\lambda) = \frac{e^{-N\lambda} \lambda^{\sum_i x_i}}{\pi x_i!}$

$\therefore P(\lambda/y) \rightarrow$ Posterior.

$P(\lambda/y) \propto P(\lambda) P(y/\lambda) \propto$

$$\propto \left(\frac{\beta^{\alpha}}{\Gamma(\alpha) \pi x_i!}\right)\left(\lambda^{\alpha-1+\sum_i^N x_i} e^{-(N+\beta)\lambda}\right).$$

$\sim \Gamma(\tilde{\alpha}, \tilde{\beta})$, where $\tilde{\alpha} = \alpha + \sum x_i$

$\tilde{\beta} = \beta + N$

$\therefore$ Posterior $\sim \Gamma(\tilde{\alpha}, \tilde{\beta})$

∴ Since the posterior is $\Gamma(\tilde{\alpha}, \tilde{\beta})$

the mode of $\lambda = \dfrac{\tilde{\alpha}-1}{\tilde{\beta}} = \dfrac{\Sigma X_i + \alpha - 1}{N + \beta}$

→ Prediction:

→ $Y\_Pred = round(\lambda_{mode}) = \hat{Y}$

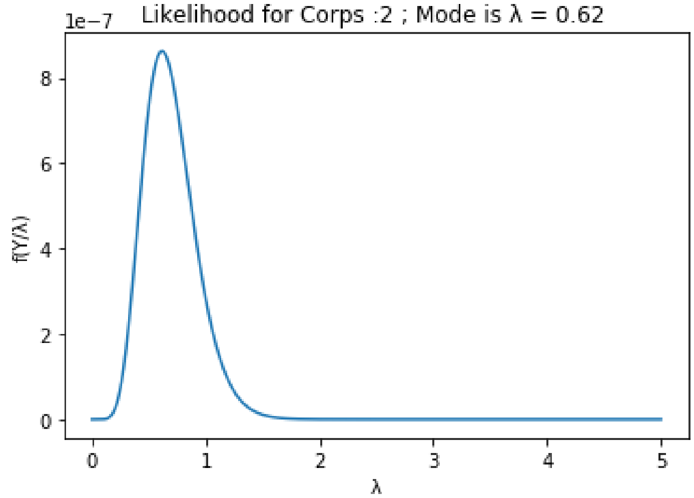
RMSE:

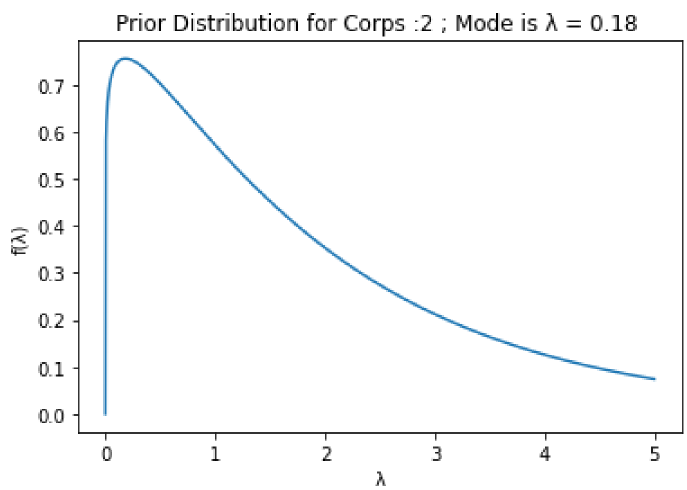
Root-mean square error:

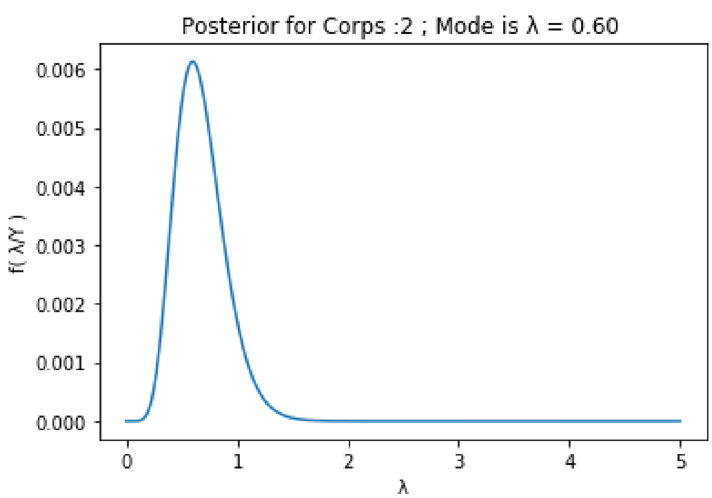
$$e = \sqrt{\dfrac{\sum_{i=1}^{N} (Y_i - \hat{Y}_i)^2}{N}}$$

3.2.b PLOTTING DISTRIBUTIONS FOR CORPS: 2,4,6

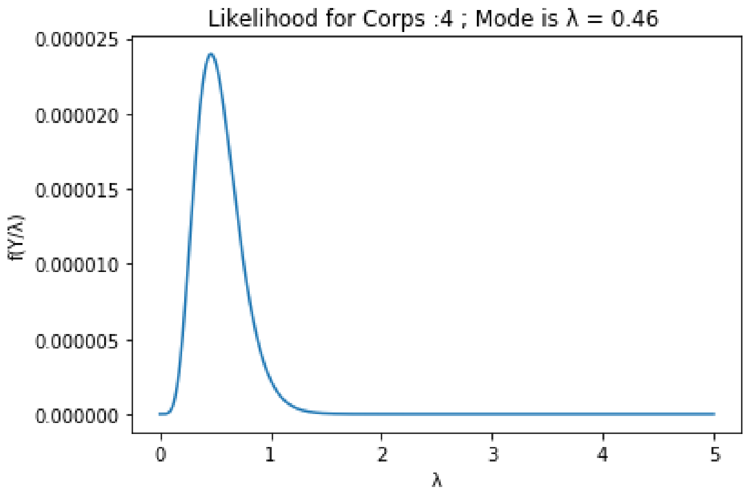
CORPS - 2



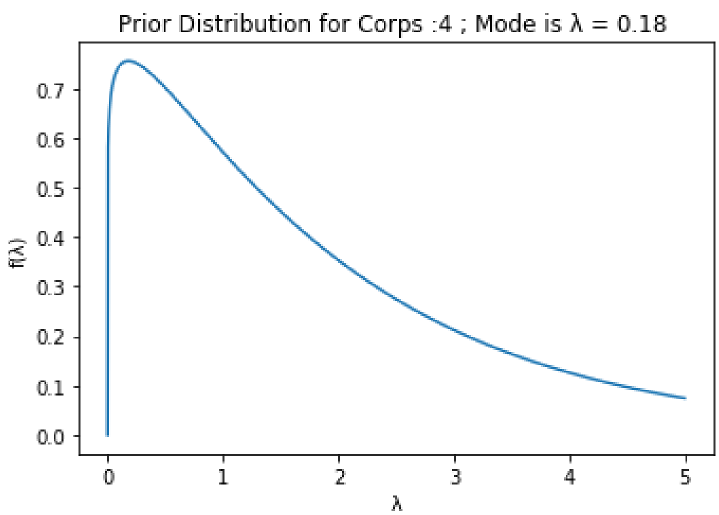




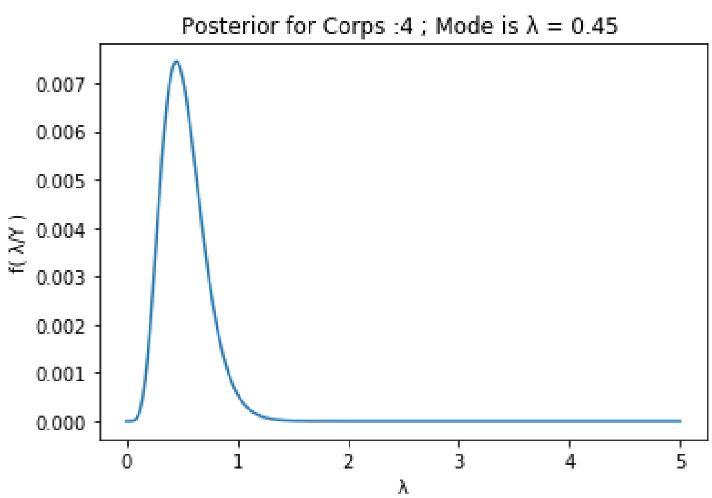
CORPS - 4



Likelihood for Corps :4 ; Mode is λ = 0.46
0.000025
0.000020
0.000015
0.000010
0.000005
0.000000
f(Y/λ)
0
1
2
3
4
5
λ

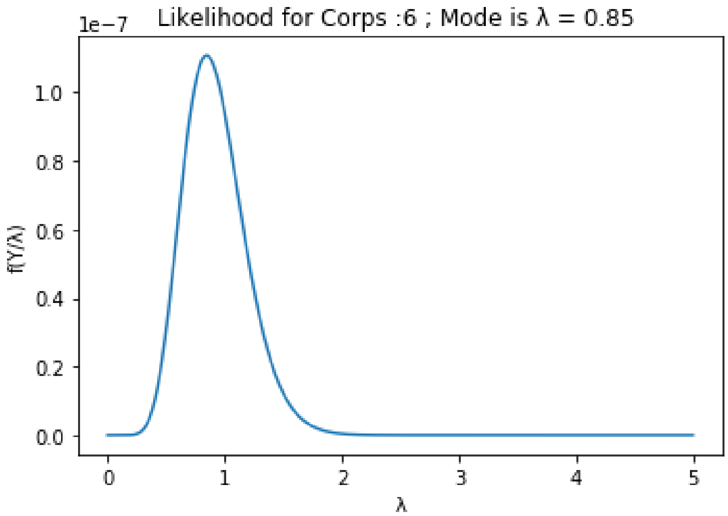


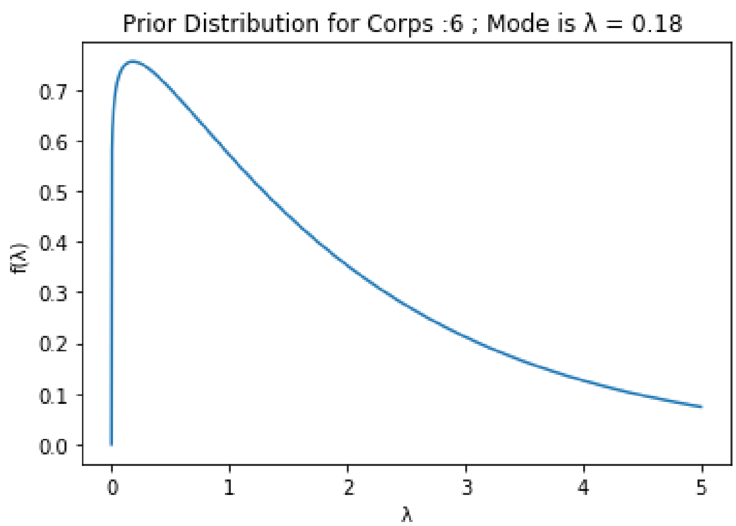
Prior Distribution for Corps :4 ; Mode is λ = 0.18
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
f(λ)
0
1
2
3
4
5
λ

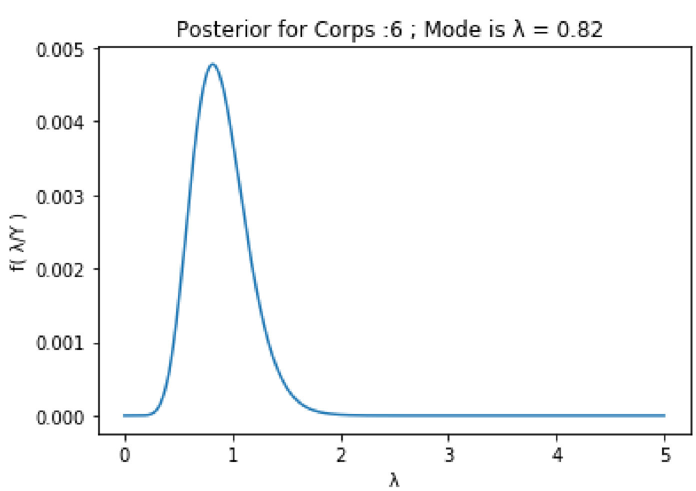


Posterior for Corps :4 ; Mode is λ = 0.45
0.007
0.006
0.005
0.004
0.003
0.002
0.001
0.000
f(λ/Y)
0
1
2
3
4
5
λ

CORPS - 6


Likelihood for Corps :6 ; Mode is λ = 0.85
1e−7
1.0
0.8
0.6
0.4
0.2
0.0
f(Y/λ)
0
1
2
3
4
5
λ


Prior Distribution for Corps :6 ; Mode is λ = 0.18
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
f(λ)
0
1
2
3
4
5
λ


Posterior for Corps :6 ; Mode is λ = 0.82
0.005
0.004
0.003
0.002
0.001
0.000
f(λ/Y)
0
1
2
3
4
5
λ

OBSERVATIONS

1. We can see that the posterior distribution lies between prior and likelihood distributions.
2. Mode of posterior distribution shifts towards the left when compared to the likelihood.
3. As explained in 3.2(a) we will consider that **λ** takes in the value of the mode of the posterior distribution.
4. The predictions are thus made similar as in 3.1, i.e we round the values of **λ** and consider them as predictions.

PREDICTIONS AND RMSE

Estimation of λ for every corps :
We estimated λ using MAP estimation through training data of every corps :

Here are the estimates :
We considered prior distribution of gamma (alpha = 1.1, beta = 0.5)

Here are the estimates of λ :

$$lamda = \begin{bmatrix} 0.903 \\ 0.628 \\ 0.559 \\ 0.559 \\ 0.421 \\ 0.352 \\ 0.766 \\ 0.49 \\ 0.283 \\ 0.628 \\ 0.49 \\ 0.903 \\ 1.317 \\ 0.283 \end{bmatrix}$$

RMSE CALCULATIONS FOR MAP ESTIMATES

Once we estimate the parameters, we can assume the predictions to be **λ**,but it should be a count value(integer) . Hence We rounded it and considered them as our predictions for the future. We calculated the RMSE error over the testing data using our estimates for each corps.

Here are the results :

$$Predictions = \begin{bmatrix}1\\1\\1\\1\\0\\0\\1\\0\\0\\1\\0\\1\\1\\0\end{bmatrix}, \qquad RMSE = \begin{bmatrix}1.613\\1.418\\1.588\\1.522\\1.005\\0.851\\1.467\\1.138\\1.035\\1.502\\1.336\\1.233\\1.403\\1.129\end{bmatrix}$$

<u>Comments on MLE vs MAP</u> :

1. We can observe that both MAP , MLE estimates (**λ**) are different, but when we predict using them , they tend to have almost the same RMSE values.
2. This occurs due to non-confident estimation of prior distributions, which is usually done by domain -experts. Here we randomly picked a few values of parameter and showcased the results.
3. MAP will always outperform MLE, because MAP consists of an additional source of information i.e; the prior(also known as Belief) , , the likelihood is called as evidence , since it is derived from the data.
4. We then combine our evidence knowledge and beliefs and come up with better estimates.
5. In this case due to lack of much data, and less domain knowledge about finding a good prior distribution our MAP performs the same as MLE (in terms of predictions (since we are rounding the lambda parameter), but the parameters obtained differ.