MP4: Bayesian Estimation

Report

Group No. 18

Rohith D	Siddharth DP	Jay Shah	Aravint Annamalai	Sayan Mitra
EE18B148	EE18B072	EE18B158	EE18B125	EE18B156

26 May, 2021

Question 1: Maximum Likelihood Estimation

We estimate the covariance matrix using MLE.

$$\hat{\Sigma} = \frac{1}{n} \sum_{i=1}^{n} (\mathbf{y}_\mathrm{i}) * (\mathbf{y}_\mathrm{i})^T$$

We obtain the following results for different sample sizes:

n	10	100	1000
Covariance	$\begin{bmatrix} 1.51131988 & 0.53082332 \\ 0.53082332 & 1.5139427 \end{bmatrix}$	$\begin{bmatrix} 1.06716844 & -0.09420016 \\ 0.09420016 & 2.15856406 \end{bmatrix}$	$\begin{bmatrix} 1.00031613 & 0.00125482 \\ 0.00125488 & 1.97069350 \end{bmatrix}$
MSE	0.26531163098	0.0118503751695	0.000215530025

We see that the Mean Square Error (MSE) reduces with increase in n.

Question 2: Using Conjugate Priors (Inverse Wishart Distribution)

- We are using maximum aposteriori probability (MAP) as the point estimate of the covariance.
- Here, MAP is nothing but the mode (most frequently occurring value) of the Inverse Wishart distribution.

We obtain the following results for different sample sizes:

n	10	100	1000
Covariance	$\begin{bmatrix} 1.061844 & 0.294901 \\ 0.294901 & 1.118857 \end{bmatrix}$	$\begin{bmatrix} 1.02515596 & -0.08722237 \\ -0.08722237 & 2.04496672 \end{bmatrix}$	$\begin{bmatrix} 0.996345362 & 0.00124486 \\ 0.00124486 & 1.9600133 \end{bmatrix}$
MSE	0.003291743633299	0.00068019320454	0.000105840402691

MSE reduces as number of samples increases. Order of MSE remains same for higher n values, but reduces a lot for low n value.

Question 3: Using Non-informative prior (Non-informative Jeffrey's prior & the Independence-Jeffrey's prior)

We estimate the covariance matrix using Non-informative and Independence priors.

We obtain the following results for different sample sizes:

n	10	100	1000
Covariance,k=1.5	$\begin{bmatrix} 1.16255 & 0.40832 \\ 0.40832 & 1.1645 \end{bmatrix}$	$\begin{bmatrix} 1.0795142 & 0.37915 \\ 0.37915 & 1.08138 \end{bmatrix}$	$\begin{bmatrix} 1.03608 & -0.09145 \\ -0.09145 & 2.09569 \end{bmatrix}$
MSE,k=1.5	0.26445612	0.00679699	0.00031235
Covariance,k=2	$\begin{bmatrix} 1.0261 & -0.09057 \\ -0.09057 & 2.07554 \end{bmatrix}$	$\begin{bmatrix} 0.997324153 & 0.00125107 \\ 0.00125107 & 1.96479910 \end{bmatrix}$	$\begin{bmatrix} 0.996330 & 0.0012498 \\ 0.0012498 & 1.96284213 \end{bmatrix}$
MSE,k=2	0.28442376	0.00569937	0.00034932

Non-informative prior gives worse results compared to conjugate prior as no information is given before estimation. The effect is more pronounced for lower n as less data is available.

Question 4: Monte-Carlo Bayesian Estimation, for 2 different priors

We perform the Monte-Carlo Bayesian estimation for two different prior matrices $\begin{bmatrix} 4 & 0 \\ 0 & 5 \end{bmatrix}$ and $\begin{bmatrix} 2 & 0 \\ 0 & 4 \end{bmatrix}$. The results are tabulated as follows:

For prior as $\begin{bmatrix} 2 & 0 \\ 0 & 4 \end{bmatrix}$

n	10	100	1000
Covariance, m=1000	$\begin{bmatrix} 1.44497 & 0.47629 \\ 0.47629 & 1.66424 \end{bmatrix}$	$\begin{bmatrix} 1.06154 & -0.053703 \\ -0.053703 & 2.256013 \end{bmatrix}$	$\begin{bmatrix} 0.998362 & 0.104509 \\ 0.104509 & 2.051804 \end{bmatrix}$
MSE,m=1000	0.1911109	0.0187746	0.00613265
Covariance,m=10000	$\begin{bmatrix} 1.58739 & 0.43925 \\ 0.43925 & 1.68135 \end{bmatrix}$	$\begin{bmatrix} 1.10734 & -0.11773 \\ -0.11773 & 2.18193 \end{bmatrix}$	$\begin{bmatrix} 0.99445 & -0.03094 \\ -0.03094 & 1.97031 \end{bmatrix}$
MSE,m=10000	0.208117	0.0180857	0.0007067
Covariance,m=100000	$\begin{bmatrix} 1.42619 & 0.43725 \\ 0.43725 & 1.58938 \end{bmatrix}$	$\begin{bmatrix} 1.06487 & -0.08980 \\ -0.08981 & 2.15452 \end{bmatrix}$	$\begin{bmatrix} 1.01069 & -0.00483 \\ -0.00483 & 1.99082 \end{bmatrix}$
MSE,m=100000	0.18315712	0.01105417	6.1339462e-05

For prior as $\begin{bmatrix} 4 & 0 \\ 0 & 5 \end{bmatrix}$

n	10	100	1000
Covariance, m=1000	$\begin{bmatrix} 1.56173 & 0.43278 \\ 0.43278 & 1.73114 \end{bmatrix}$	$\begin{bmatrix} 1.08616 & -0.08228 \\ -0.08228 & 2.21012 \end{bmatrix}$	$\begin{bmatrix} 0.93449 & -0.06055 \\ -0.06055 & 2.07872 \end{bmatrix}$
MSE,m=1000	0.190606	0.016279	0.004456
Covariance,m=10000	$\begin{bmatrix} 1.44613 & 0.42539 \\ 0.42539 & 1.58753 \end{bmatrix}$	$\begin{bmatrix} 1.06540 & -0.07157 \\ -0.07157 & 2.12243 \end{bmatrix}$	$\begin{bmatrix} 0.98527 & -0.03577 \\ -0.05377 & 2.03838 \end{bmatrix}$
MSE,m=10000	0.18277	0.007379	0.001062
Covariance,m=100000	$\begin{bmatrix} 1.59405 & 0.44337 \\ 0.44337 & 1.67931 \end{bmatrix}$	$\begin{bmatrix} 1.08913 & -0.09024 \\ -0.09024 & 2.16486 \end{bmatrix}$	$\begin{bmatrix} 1.01281 & 0.00569 \\ 0.00569 & 1.99081 \end{bmatrix}$
MSE,m=100000	0.2122217	0.01285266	7.834064e-05

Monte Carlo Method approximates the MMMSE estimate (expected value) the posterior distribution.
The estimate with prior as $\begin{bmatrix} 2 & 0 \\ 0 & 4 \end{bmatrix}$ performs better here. This is because the inverse Wishart distribution is probability distribution of the maximum-likelihood estimator (MLE) of the covariance matrix of a multivariate normal distribution.. Here the mean is $\Delta/(\nu$-d-1). So, when the prior Δ is twice the actual value of Δ, then the mean is the actual value of Δ. This is the reason why $\begin{bmatrix} 2 & 0 \\ 0 & 4 \end{bmatrix}$ gives the best result. This also justifies the need of modelling the prior properly.

Question 5: Hierarchical Bayes estimation and Gibbs sampling:

We obtain the following results for different sample sizes:

n	10	100	1000
Covariance	$\begin{bmatrix} 0.8061 & 0.2793 \\ 0.2793 & 0.8074 \end{bmatrix}$	$\begin{bmatrix} 0.9915 & -0.0864 \\ -0.0864 & 1.9929 \end{bmatrix}$	$\begin{bmatrix} 1.0038 & 1.243e-03] \\ 1.243e-03 & 1.9656 \end{bmatrix}$
MSE	0.403957	0.003765	0.000298

Question 6: Empirical Bayes:

n	10	100	1000
Covariance	$\begin{bmatrix} 1.26963062 & 0.445934 \\ 0.44593441 & 1.271833 \end{bmatrix}$	$\begin{bmatrix} 1.03767138 & -0.09159642 \\ -0.09159642 & 2.09890031 \end{bmatrix}$	$\begin{bmatrix} 0.99734048 & 0.00125109 \\ 0.00125109 & 1.9648312 \end{bmatrix}$
MSE	0.250160350746	0.0069950530989	0.000311760606128

As we see above, having a conjugate prior gives better results than a hierarchical prior, which in turn, performs better than a non-informative prior prior. This is because inverse Wishart is the PDF of the covariance matrix estimated by the MLE and having a prior information is always beneficial.Having no prior info gives the maximum MSE.

Conclusion

- It is suggested to use Monte Carlo Bayesian Estimation since it is an mmse estimator and gives the best (lowest mse results).
- If we had used MMSE instead of MAP for Conjuagate prior then it would have also performed along the similar lines as Monte Carlo
- For problems of a higher dimension, it is harder to form a prior, so **Empirical Bayes**, which makes a prior can be used