Phase Analysis of Circadian-related genes in two tissues

Ashwini K Maurya, PhD Student, MSU

Supervised by Prof. Ashis Sengupta, Applied Statistical Unit, Indian Statistical Institute, Kolkata, India

Objective and Scope

- Recent circadian-clock studies of two different tissues of mouse has revealed some lag differences in phase expression times across two tissue in *vivo.*

- Earlier work by Shyamal et al.[2] has proposed clustering of genes based on differences of phase expression times across two tissues.

- We have prosed bivariate statistical modeling of phase expression times to better understand the joint characteristics of phase expression times of two tissues viz. heart and liver.

Data and Model Specification

- The phase angles for heart and liver tissues expression times has been estimated from cycling transcripts in Panda et al.[6]

Probe ID	Gene Description	Accession	Phase (heart)	Phase (liver)
100581_at	cystatin B	U59807	0.62	1.45
94145_at	interferon beta, fibroblast	K00020	2.35	-2.51

- Dataset consists of 48 genes and their phase angles in two tissues. [2]

- To analyze this interesting phenomenon, bivariate *von-mises* model is proposed, looking at the circular plot of observed phase angles.

Data and Model Specification

- Probability density function of Bivariate Von-Mises Sine model is given by :

$$f(\theta, \phi, \mu, \nu, \kappa 1, \kappa 2, \lambda 12) = C\mathcal{e}^{\kappa 1 Cos[\theta - \mu] + \kappa 2 Cos[\nu - \phi] + \lambda 12 Sin[\theta - \mu] Sin[\theta - \nu]}$$
$$\{-\pi \leq \theta, \phi < \pi, -\pi \leq \mu, \nu < \pi, \kappa 1 \geq 0, \kappa 2 \geq 0, -\infty < \lambda 12 < \infty\}$$

- (μ, ν) and $(\kappa 1, \kappa 2)$ are the mean directions and concentration parameters of respectively. $\lambda 12$ Accounts for statistical dependence between θ, ϕ. C is normalizing constant given by :

$$C = \frac{1}{4\pi^2 \sum_{m=0}^{\infty} 4^{-m} (\kappa 1 \kappa 2)^{-m} \lambda 12^{2m} \text{BesselI}[m, \kappa 1] \text{BesselI}[m, \kappa 2] \text{Binomial}[2m, m]}$$
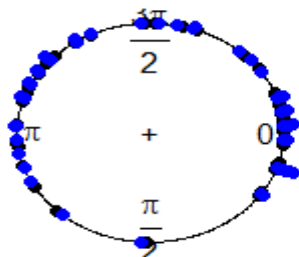
Univariate Analysis

- Heart Phase Angles Data :

  Mean Direction : -1.005 Rad, Circular Var. :0.748 Rad sq.
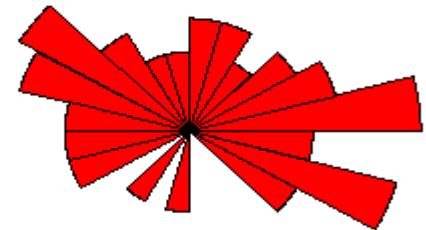
  MLE for *Von-Mises* Distr. :

  $\mu.\,\text{est}$= -1.005 (0.398), $\kappa.\,est$=0.521 (0.215)

- The Circular plot as well as the rose diagram suggests presence of bimodality in the data.

Circular plot of Heart phase angles

Rose Diagram
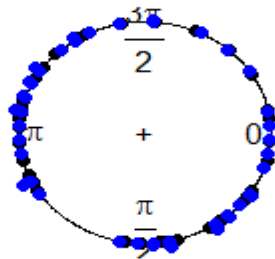
Univariate Analysis

- Liver Phase Angles Data :

  Mean Direction : 2.623 Rad, Circular Var. :0.860 Rad sq.
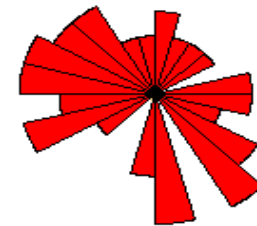
  MLE for *Von-Mises* Distr. :

  $\mu.\,est$= 2.623 (0.725), $\kappa.\,est$=0.282 (0.207)

- Similar to Heart phase Circular plot, bimodality is observed at about "–Pi/3" and "2 Pi/3" Radian.

Circular plot of Liver phase angles

Rose Diagram

Estimation of Parameters

- MLE as well as Pseudo likelihood method of estimation is employed to estimates the parameters. Both of these methods gives similar results.

- For the proposed BV Sine model, Mardia et al. [2] has given the condition of Uni-modality and Bimodality (Theorem 2 in [4]) of the proposed joint Sine density which can be checked for the fitted model.

  condition of bimodality is : $\lambda 12^2 > \kappa 1^* \kappa 2,$

Continued…

- The estimated parameters are given in table below :

Estimates/Method	ML Estimates	Pseudo Likelihood Estimates
Log-Likelihood	-163.66	-164.058
μ	-1.007	-0.894
ν	1.870	1.808
$\kappa 1$	0.577	0.565
$\kappa 2$	0.227	0.210
$\lambda 12$	-1.312	-1.127

- The condition of Bimodality for the joint density has been satisfied, therefore the fitted model is bimodal and is appropriate for this data.

Test of Independence

- To test the marginal independence of the categories, we implement likelihood ratio test.

    H0 : Both the categories are independent, which

    is equivalent to testing $\lambda 12 = 0$ ***vs***

    H1 : $\lambda 12! = 0$

- (Since under H0, the parameter space does not include the boundary points, the usual LR test is valid.)

- Value of Chi-Square Test Statistic = 17.488

- P-value= 7.286×10^{-9}

- Therefore we reject the null Hypothesis and therefore both the categories are dependent.

Distribution of Marginal Density

- I have worked out the marginal densities and used Watson U^2 G-o-F test for fitted marginal models.

Parameters /Category	Heart	Liver
Log-Likelihood	-163.683	-172.203
μ	-1.402	-0.9898
$\kappa 1$	0.875	0.1007
$\kappa 2$	0.014	1.6×10^{-8}
$\lambda 12$	3.709	2.726

- In spite of lack of G-o-F test for Bivariate circular models, the G-o-F for marginal densities is useful as these are derived from the joint distribution.

G-o-F test for Marginal Density

- The Watson U^s G-o-F test implemented for marginal densities.

$$H0 : \theta \sim f(\theta, \mu, \kappa1, \kappa2, \lambda12)$$

$$H1 : H0 \text{ is not True.}$$

Category	Heart	Liver
Watson U^2 Test statistic	0.0414	0.0085
P-Value	0.823	0.347

- Since P-value is quite high for both the category, the null hypothesis is retained. Therefore we conclude that the marginal gives an appropriate fit to the data.

Continued…

- Singh[3] has been worked out (Theorem (3) of [3]) the condition of Uni-modality and Bimodality for the marginal densities derived from this joint density.

- The condition of bi-modality of marginal distributions is also satisfied, therefore the marginal distributions will have bimodality.
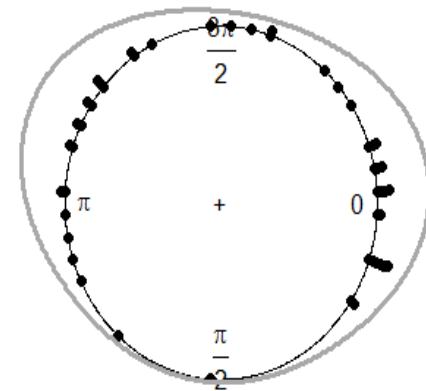
Marginal Density Plots (Heart)

- The plot confirms the bimodality of the marginal model, which is expected as the observed data exhibits bi-modality around the circular plot.

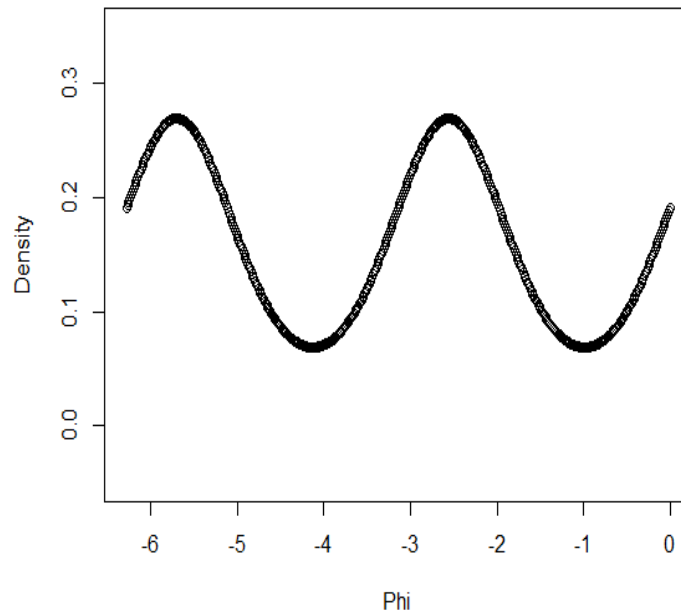**Estimated marginal Sine density for Heart Category**
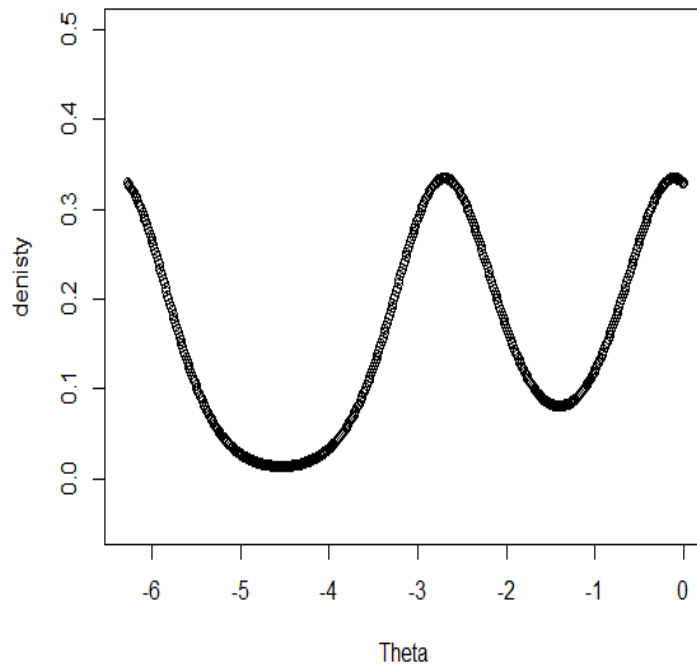
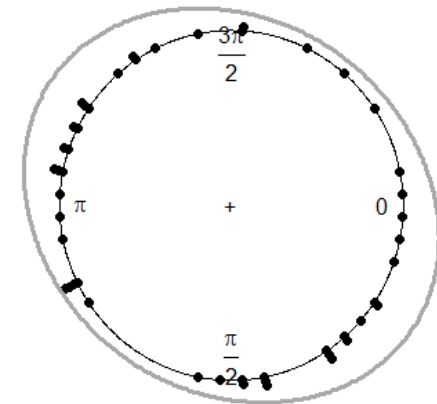**Estimated density estimate along the Circle**

Marginal Density Plots (Liver)

- The similar plot for Liver category establishes the appropriateness of the model as observed in the circular plot. The density is bimodal and fits very well to the data.

Estimated marginal Sine density for Liver Category

Estimated density estimate along the Circle

Cluster 1 Analysis

- Shyamal et al.[2] has modeled the differences between the phase angles and clustered the genes in two categories.

- They claim that the genes in cluster 1 has proximity in the phase angles for both the categories.

- To test their claim, we fit the Bivariate von-Mises density to the cluster 1 data and implement test of hypothesis of equality of means for both the groups.(Cluster1 has 38 genes,Shyamal et al.[2]

Continued…

- The estimated parameters for cluster 1 data are given in below :

Parameters	ML Estimates
Log-Likelihood	-116.047
μ	-1.209
ν	1.996
$\kappa 1$	0.659
$\kappa 2$	0.240
$\lambda 12$	-2.438

- The condition of Bimodality for the joint density has been satisfied. Also well the condition of bimodality of Marginal densities. Therefore the model for cluster 1 is similar to full model.

Continued…

- The marginal densities are fitted and Watson U^2 test is implemented for G-o-F.

- The P-values for Marginal of heart and liver category are 0.744 and 0.781 respectively.

- High P-value suggests that the marginal fits are appropriate for the both the categories.

Test of Equality of Means

- To test their claim that the difference in the phase angles of heart and liver tissues in cluster 1 are close, we implement the following test of hypothesis.

$$H0: \mu = \nu \quad \textbf{\textit{vs}} \quad H1: \mu \mathrel{!}= \nu$$

- Value of test statistic $-2Ln\,\lambda = 5.89$

- Assuming the asymptotic distribution of test statistic is Chi-Square with 1df

$$P\text{-Value}=0.0163$$

- Therefore null hypothesis of equality of mean directions is rejected at 2% level of significance.

Summary

- The proposed model works very well and it's appropriateness has been accomplished up to certain extent by testing the G-o-F of derived marginal distributions.

- The Marginal plots of the densities suggests further the suitability of the model it captures the bimodality present in the data.

- I also tried BV von-Mises model with Cosine component as well Bivariate cardioid but the those models were not able to capture the peaked-ness and the variation in the data.

Continued…

- The model for cluster 1 is very much similar to the full model, which characterizes the robustness of model even with a sample of size 38.

- The approach of testing the marginality has not been discussed in many instances in similar kind of research, which could be implement to advocate the appropriateness of the model.

Future work…

- Address the identification of further clusters of genes having similar expression times within cluster 1.

References

[1] S. Rao Jammalamadaka, A. SenGupta: **Topics in Circular Statistics,** World Scientific, 2001.

[2] Delong Liu, Shyamal D Peddada, Leping Li and Clarice R Weinberg : **Phase analysis of circadian-related genes in two tissues, BMC Bioinformatics, 2006.**

[3] Harshinder Singh, Vladimir Hnizdo and Eugene Demchuk : **Probabilistic Model for Two Dependent Circular Variables, Biometrica, Sep,2002.**

[4] Kanti V. Mardia, Charles C. Taylor, and Ganesh K. Subramaniam: P**rotein Bioinformatics and Mixtures of Bivariate von Mises Distributions for Angular Data, Biometrics, June 2007.**

[5] GRACE S. SHIEH I AND RICHARD A. JOHNSON: **INFERENCES BASED ON A BIVARIATE DISTRIBUTION WITH VON MISES MARGINALS, Ann. Inst. Stat. Mathematics.2005.**

[6] Panda S, Antoch MP, Miller BH, Su AI, Schook AB, Straume M, Schultz PG, Kay SA, Takahashi JS, Hogenesch JB: **Coordinated transcription of key pathways in the mouse by the circadian clock.** *Cell* 2002, **109:**307-320.

Questions and Suggestions

Thank You!