INDIAN INSTITUTE OF TECHNOLOGY MADRAS
Department of Chemical Engineering

CH5115: Parameter and State Estimation

Assignment 4

Due: Thursday, December 02, 2020

---

1. The time-to-publication (in months) of randomly selected articles in a certain journal is provided in `qdist_data.mat`.
   (a) Fit an appropriate distribution for the observations. Provide suitable justifications for your answer.
   (b) Suppose the given observations are in error. Qualitatively discuss the impact of different error types, namely, (i) systematic error and (ii) random error on the difference between distributions of observed data and the error-free data.
2. A random sample of size N is drawn from a process governed by the p.d.f. $f(y;\theta) = 2y/\theta^2$.
   (a) Find the ML estimate of θ. Check for the bias and propose a modification, if needed, to make it unbiased.
   (b) Determine the (theoretical) median of Y and its ML estimator.
   (c) Check for the (theoretical) consistency of estimator in (2b).
   (d) Verify (2c) numerically by drawing random samples of different sizes. Finally, check if the estimates are asympotically Gaussian.
3. Consider the stack loss data provided in `stack_loss.mat`. This was published in an article in Technometrics journal. The response variable y is the amount of ammonia escaped during its oxidation to nitric acid. The regressors are air flow (ψ_1), temperature (ψ_2) and acid concentration (ψ_3).
   (a) Determine the correlation between response and each of the regressors. Is a linear model qualified between y and the regressors?
   (b) Fit a linear regression model accordingly and compute the goodness of model diagnostics, specifically, R^2, adjusted R^2 and significance of regression at $\alpha = 0.05$.
   (c) Test for significance of each regression coefficient at $\alpha = 0.05$.
   (d) Calculate the 95% CI on mean stack loss when $\psi_1 = 80$, $\psi_2 = 25$ and $\psi_3 = 90$.
   (e) Finally, compute the 95% PI for stack loss at the same values of regressors in (3d).
4. Consider the data provided in `engine_thrust.mat`. It contains a response variable (the thrust of a jet-turbine engine) and six regressor variables.
   (a) Perform a full regression of jet thrust on all regressors using the LS method and perform model diagnostics including residual analysis. Are all the assumptions made in using LS satisfied?

(b) Eliminate the terms with insignificant coefficients and redo the regression. Report your findings.

(c) Which among (4a) and (4b) is a better model? Perform a stepwise regression with $\alpha_{in} = 0.1$ and $\alpha_{out} = 0.15$. Does the resulting model agree with your choice of better model?

(d) Plot the residuals from the model of your choice against regressors and check for nonlinearities.

(e) Finally, fit a non-linear regression model with the speculated non-linearities. Is the resulting model more satisfactory than the linear regression model? Provide supporting arguments.

5. (a) Show that the ridge regression (Tikhonov regularization) for observations generated by $y[k] = \boldsymbol{\varphi}^T[k]\boldsymbol{\theta} + e[k]$ is equivalent to Bayesian estimation with a Gaussian prior $\boldsymbol{\theta} \sim \mathcal{N}(0, \sigma_\beta^2)$ and $e[k] \sim$ GWN$(0, \sigma_e^2)$. Further, determine a relation between the hyper parameter λ and the variances σ_β^2 and σ_e^2.

   (b) Show that the elastic-net optimization problem:

$$\min_{\boldsymbol{\theta}} ||\mathbf{y} - \Phi\boldsymbol{\theta}||_2^2 + \lambda[\alpha||\boldsymbol{\theta}||_2^2 + (1-\alpha)||\boldsymbol{\theta}||_1] \tag{1}$$

   can be cast as a LASSO problem using an augmented version of Φ and $\mathbf{y}$.

6. Consider a random sample of observations from DGP $y[k]|\theta \sim$ Poisson(θ), $\quad k = 1, 2, \cdots, N$ with $\theta > 0$. Suppose DataRaja assumes $\pi(\theta) \sim \theta^{-1/2}$.

   (a) Show that $\pi(\theta)$ is in the class of Jeffreys' priors.

   (b) Further, show that the posterior PDF of $2N\theta$ is the PDF of a $\chi^2(2N\bar{y} + 1)$ distribution.

   (c) Obtain the Bayesian estimate (MMSE) of θ.

   (d) Use the posterior PDF of (6b) to obtain a $(1?\alpha)100\%$ credible interval for θ.