Math 540: Project 4

Tilekbek Zhoroev

1. Consider the height-weight model

$$Y_i = \theta_0 + \theta_1(x_i/12) + \theta_2(x_i/12)^2 + \varepsilon_i$$

with the data compiled in Table 1. Using the parameter and covariance matrix estimates

$$\boldsymbol{\theta} = \begin{bmatrix} 261.88 \\ -88.18 \\ 11.96 \end{bmatrix}, \quad \boldsymbol{V} = \begin{bmatrix} 634.88 & -235.04 & 21.66 \\ -235.04 & 87.09 & -8.03 \\ 21.66 & -8.03 & 0.74 \end{bmatrix}$$

compute and plot 2σ and frequentist prediction intervals, along with the data, for heights ranging from 58 to 72 inches. Repeat this for the extrapolatory regime of 50 to 80 inches and discuss your results.

Height (in)	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72
Weight (lbs)	115	117	120	123	126	129	132	135	139	142	146	150	154	159	164

Table 1: Height-weight data

Solution The design matrix in this case

$$\mathbf{X} = \begin{bmatrix} 1 & x_1/12 & (x_1/12)^2 \\ 1 & x_2/12 & (x_2/12)^2 \\ \vdots & \vdots & \vdots \\ 1 & x_n/12 & (x_n/12)^2 \end{bmatrix}$$

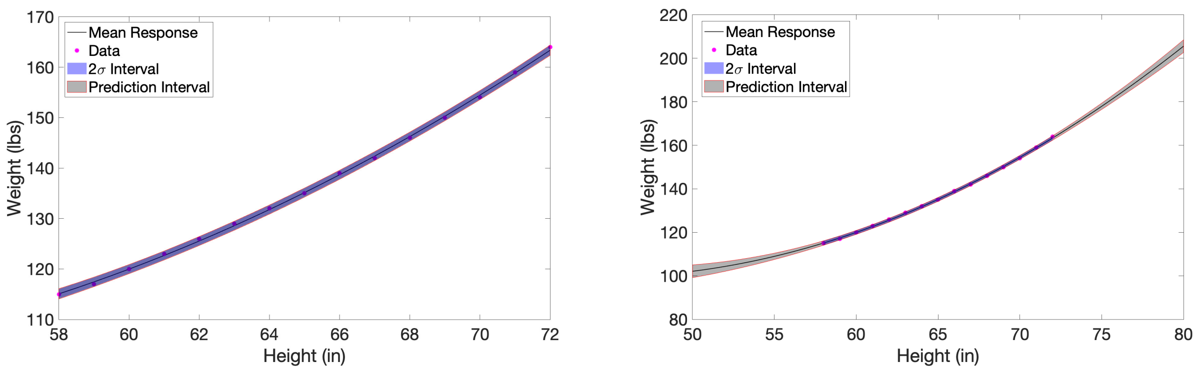
where x_i given data points for $i = 1, 2, 3, ..., n = 15$. Next, we can find the relative residuals and, by using them we can obtain the variance, σ^2 and covariance matrix, V. Since $\mathbb{E}[\mathbf{Y}] = \mathbf{X}\boldsymbol{\theta_0}$ and $\text{cov}[\mathbf{Y}] = \mathbf{X}\mathbf{V}\mathbf{X}^\top + V_{obs}$, where $\boldsymbol{V}_{obs} = \sigma^2\mathbf{I}_n$ the $\pm 2\sigma_{\boldsymbol{Y}}$ interval becomes

$$\left[\mathbf{X}_i\boldsymbol{\theta}_0 \pm 2\sqrt{(\text{cov}[\mathbf{Y}])_{ii}}\right]$$

where $\mathbf{X}_i$ is the i'th row of the design matrix. The $(1-\alpha)\times 100\%$ frequentist prediction interval at x_* is

$$\left[\hat{Y}_{x_*} \pm t_{n-p,1-\alpha/2}\hat{\sigma}\sqrt{1 + x_*(\mathbf{X}^\top\mathbf{X})^{-1}x_*^\top}\right].$$

Employing these formulas, we obtain the $\pm 2\sigma_{\mathbf{Y}}$ and frequentist 95% prediction interval, in Figure 1(a). From the graph, we observe that both the intervals are the same and all the given data points are inside of the intervals. Next, we redo the same analysis in the interval [50, 80] and the $\pm 2\sigma_{\mathbf{Y}}$ and frequentist 95% prediction interval, in Figure 1(b). Here we observe that the prediction interval for the extrapolation domain is wider than the calibration domain.



(a) (b)

Figure 1: Data, $2\sigma_{\mathbf{Y}}$ and prediction intervals for (a) calibration domain [58, 72] and (b) extrapolation domain [50, 80]

2. Consider the steady state heat model

$$\begin{aligned} \frac{d^2\,u_s(x)}{dx^2} &= \frac{2(a+b)h}{kab}[u_s(x) - u_{amb}], \quad 0 < x < L \\ \frac{d\,u_s}{dx}(0) &= \frac{\Phi}{k} \\ \frac{d\,u_s}{dx}(L) &= \frac{h}{k}[u_{amb} - u_s(L)]. \end{aligned}$$

detailed in Example 3.9 and illustrated in Example 12.17 in the context of Bayesian model calibration. Here we employ a subset of the data from Table 2 to construct prediction intervals that extrapolate beyond the calibration domain. Employ the DRAM to construct densities for Φ, h, and ε using the data compiled in Table 3. You can use the thermal conductivity value k = 2.37 for aluminum. By sampling from the densities, construct credible and prediction intervals for $x \in [10, 66]$ and plot with the complete data set from Table 2.

x (cm)	10	14	18	22	26	30	34	38	42	46	50	54	58	62	66
Temp (°C)	96.14	80.12	67.66	57.96	50.90	44.84	39.75	36.16	33.31	31.15	29.28	27.88	27.18	26.40	25.86

Table 2: Steady-state temperatures measured at locations x for an aluminum rod.

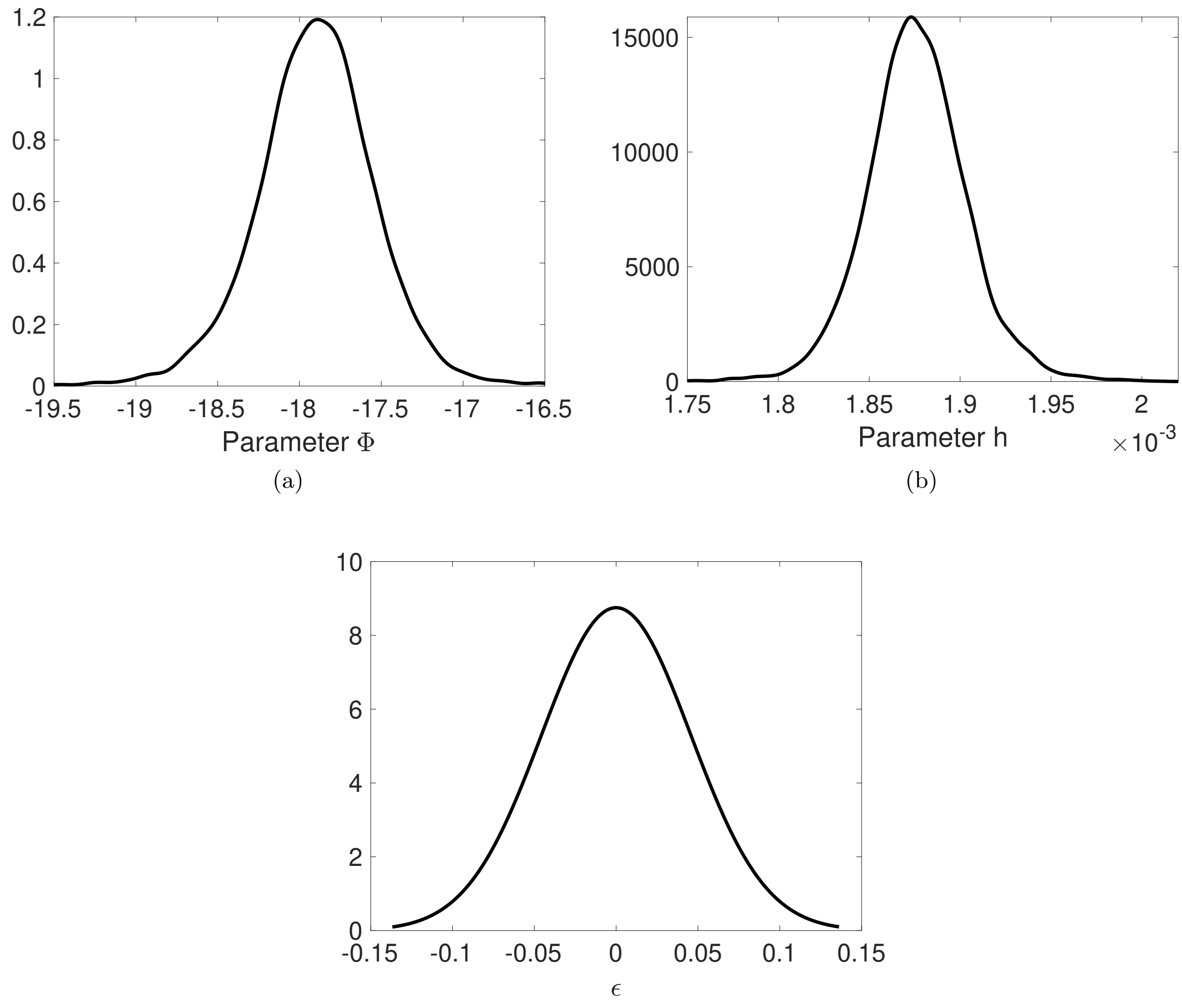
x (cm)	22	26	30	34	38	42	46	50	54
Temp(°C)	57.96	50.90	44.84	39.75	36.16	33.31	31.15	29.28	27.88

Table 3: Steady-state temperatures at locations x for an aluminum rod.

Solution. Since the parameter estimations and covariance matrices are given in Example 12.17, we would use these results to start the DRAM algorithm. After obtaining all the chains, we observe that all the chains have converged. Moreover, the covariance of the chain is

$$\mathbf{V} = \begin{bmatrix} 1.3538 \times 10^{-1} & -1.0206 \times 10^{-5} \\ -1.0206 \times 10^{-5} & 7.9377 \times 10^{-10} \end{bmatrix}$$

and the variance $\sigma^2 = 0.0461$. Then using chains and **kde.m** we construct densities for Φ, h, and ε given in Figure 2.



(a) (b) (c)

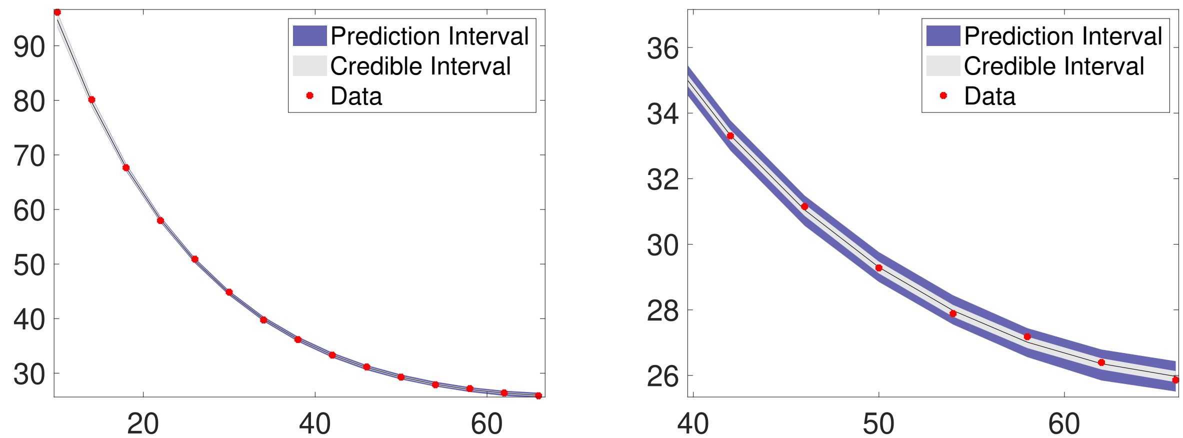
Figure 2: Densities for (a) Φ; (b) h; (c) ε

Next, we used the DRAM commands **mcmcpred** and **mcmcpredplot** to construct 95% credible and prediction intervals for the given data in Table 2. From Figure 2, we observed that the error variance and variance of the parameters are small. Hence, when we propagate the model response, the credible and prediction intervals are small. As a result, we obtained the zoomed version of part (a) in Figure 3(b) to demonstrate the difference between credible and non-credible prediction intervals.

3. Consider the SIR model

$$\begin{aligned} \frac{dS}{dt} &= \delta N - \delta S - \gamma IS, && S(0) = 900 \\ \frac{dI}{dt} &= \gamma IS - (r+\delta)I, && I(0) = 100 \\ \frac{dR}{dt} &= rI - \delta R, && R(0) = 0 \end{aligned}$$

where γ, r and δ are each in the interval [0, 1].



(a) (b)

Figure 3: (a) Data, 95% credible and prediction interval using data in Table 2; (b) Reduced x-axis to illustrate the difference of credible and prediction interval.

(a) The file **SIR.txt** contains times t_i in the first column and corresponding values $I(t_i)$ in the second, and the parameters $\boldsymbol{\theta} = [\gamma, r, \delta]$. Employ DRAM to compute parameter chains in the manner investigated in Exercise 12.8. Use the DRAM commands **mcmcpred** and **mcmcpredplot**, to construct 95% credible and prediction intervals for $I(t)$ and plot with the data from **SIR.txt**

(b) Consider the influenza, as detailed in Exercise 11.7(d), you can employ $\delta = 0$ and the initial values $S(0) = 730$, $I(0) = 3$ and $R(0) = 0$ so that $N = 733$. Employ DRAM to compute parameter chains in the manner investigated in Exercise 12.8(c). Construct 95% credible and prediction intervals for $I(t)$ and plot with the data in Table 4.

Day	0	1	2	3	4	5	6	7	8	9	10	11	12	13
Confined to Bed	3	8	26	76	225	298	258	233	189	128	68	29	14	4

Table 4: Influenza data

Solution.

(a) Since we performed Bayesian model calibration using DRAM in the previous project, we skipped that part. It's worth noting that the chains are converged. Hence, we can use these results to propagate parameter uncertainty for the model response. Then, using the DRAM commands **mcmcpred** and **mcmcpredplot**, we construct 95% credible andprediction intervals for given data in **SIR**.**txt**. Figure 4(a) shows the graphs of the credible and prediction intervals.

(b) We used the DRAM algorithm for this problem in Project 3. Here we observed that the chains have all converged. Since we have a converged chain, we obtained the 95% credible intervals using posterior samples. The graph in Figure 4 was obtained using the **mcmcpred** and **mcmcpredplot** commands.

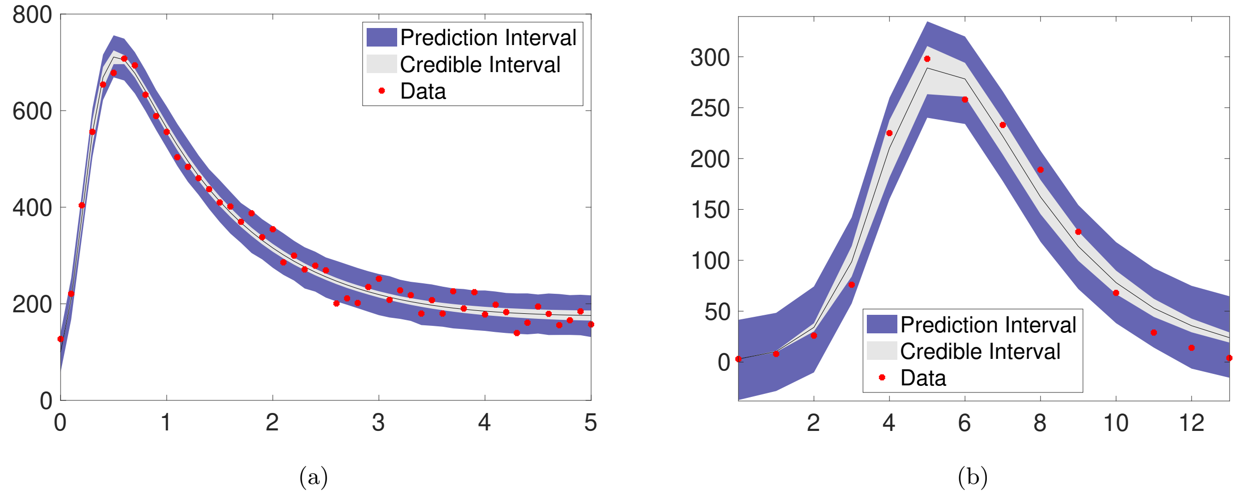


Figure 4: Data, 95% credible and prediction interval using (a) data in **SIR.txt**; (b) Influenza data, Table 4.

4. Consider the SIR model from Exercise 13.5 with parameters $\boldsymbol{\theta} = [\gamma, r, \delta]$. The file **SIR.txt** contains times t_i and corresponding values $I(t_i)$. Employ the complex-step or sensitivity equation code posted for Example 8.9 to estimate the sensitivity matrix $\mathbf{S}$ having entries $[\mathbf{S}]_{ij} = \frac{\partial I}{\partial \theta_j}(t_i, \bar{\boldsymbol{\theta}})$ for the nominal parameter values $\bar{\boldsymbol{\theta}} = [0.0100; 0.7970; 0.1953]$ estimated in Exercise 11.7. You can employ the estimate $\sigma^2 = 426.8$ from that example. Use the sensitivity matrix to estimate the parameter covariance matrix $\mathbf{V}$ and construct the response and observation matrices

$$\mathrm{var}[\boldsymbol{f}(\boldsymbol{\theta})] = \mathbf{SVS}^\top$$

$$\mathrm{var}[\mathbf{Y}] = \mathbf{SVS}^\top + \sigma^2 \mathbf{I}_{n\times n}.$$

Let $\sigma_{\boldsymbol{f}}(t_i)$ and $\sigma_{\mathbf{Y}}(t_i)$ denote the square roots of the diagonal elements. Plot $I(t_i)$ and the $\pm 2\sigma_{\boldsymbol{f}}(t_i)$ and $\pm 2\sigma_{\mathbf{Y}}(t_i)$ intervals and compare with the 95% credible and prediction intervals computed in Exercise 13.5. What do you conclude about the accuracy of the linearization?

Solution

Using complex step approximation,

$$f'(x) \approx \frac{Im(f(x+ih))}{h},$$

and the optimal parameters obtained by frequentist approach we get the sensitivities of the response, $I(t)$, to the each parameter. By utilizing it, we obtained the sensitivity matrix $\mathbf{S}$ having entries $[\mathbf{S}]_{ij} = \frac{\partial I}{\partial \theta_j}(t_i, \bar{\boldsymbol{\theta}})$. Employing estimated variance and sensitivity matrix we obtained the covariance matrix, $\mathbf{V}$. Next using sensitivity matrix and covariance matrix we obtained the response and observation matrices

$$\mathrm{var}[\boldsymbol{f}(\boldsymbol{\theta})] = \mathbf{SVS}^\top$$

$$\mathrm{var}[\mathbf{Y}] = \mathbf{SVS}^\top + \sigma^2 \mathbf{I}_{n\times n}$$

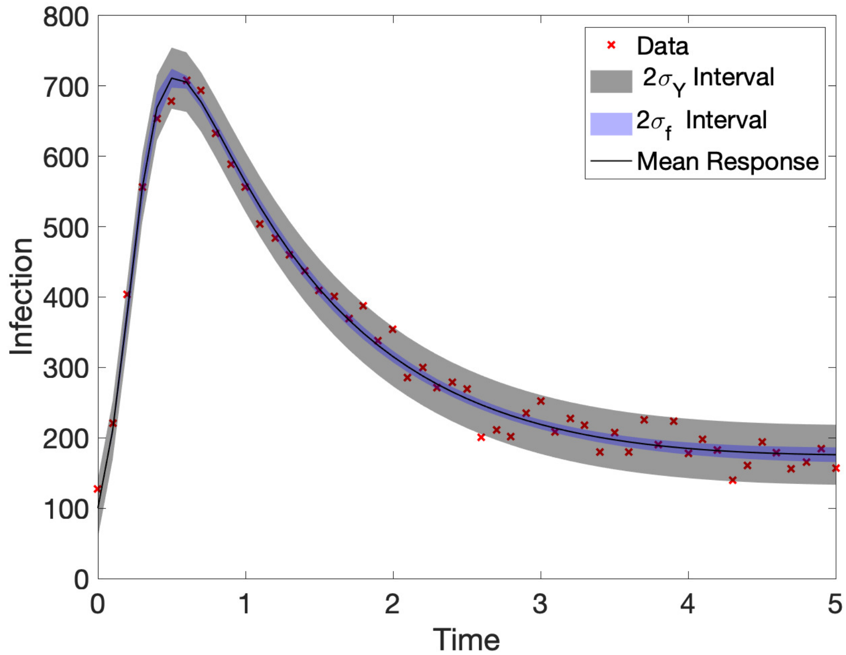


Figure 5: Data, $\pm 2\sigma_{\mathbf{f}}$ and $\pm 2\sigma_{\mathbf{Y}}$ interval using data in **SIR.txt**.

where both of them are 51×51 matrices, since we have $n = 51$ data points. In this case the mean response would be $I(t_i, \bar{\boldsymbol{\theta}})$, the $\pm 2\sigma_{\boldsymbol{f}}(t_i)$ interval is

$$\left[I(t_i, \bar{\boldsymbol{\theta}}) \pm 2\sqrt{(\mathrm{var}[\boldsymbol{f}(\boldsymbol{\theta})])_{ii}} \right]$$

and the $\pm 2\sigma_{\mathbf{Y}}(t_i)$ interval is

$$\left[I(t_i, \bar{\boldsymbol{\theta}}) \pm 2\sqrt{(\mathrm{var}[\mathbf{Y}])_{ii}} \right].$$

The intervals are given in Figure 5. We can observe that this intervals are narrower than Bayesian 95% credible and prediction interval given in Figure 4. But the difference is not big, so we can conclude that accuracy of the linearization is high.