10.007: Modelling the Systems World

SUTD 2020 1D Project Class 19F07 Group 4

Andaramanage Charini Amanda Jinadasa
James Raphael Tiovalen
Kadir Bahajjaj
Mun Jern Wei Ivan
Ng Yu Yan
Peter Santosh David

PART A

A0) Our group members list is as follows:

Member Name	Student ID
Andaramanage Charini Amanda Jinadasa	[REDACTED]
James Raphael Tiovalen	[REDACTED]
Kadir Bahajjaj	[REDACTED]
Mun Jern Wei Ivan	[REDACTED]
Ng Yu Yan	[REDACTED]
Peter Santosh David	[REDACTED]

Thus, our average student ID value is 1004382.333 and our value of $X \approx 219.1$.

A1) Taking the natural logarithm of both sides of the equation $F = \alpha e^{\beta C}$, we would obtain:

$$\ln(F) = \beta C + \ln(\alpha)$$

To obtain the values of α and β, we would use regression. By forming the matrices:

$$A = \begin{bmatrix} 1 & 8 \\ 1 & 10 \\ 1 & 12 \\ 1 & 14 \\ 1 & 16 \\ 1 & 18 \end{bmatrix}, \text{ and } \vec{b} = \begin{bmatrix} \ln(8.2) \\ \ln(19.1) \\ \ln(37.6) \\ \ln(70.4) \\ \ln(145.9) \\ \ln(219.1) \end{bmatrix},$$

$$\text{then } \left(A^\intercal A\right)^{-1} \approx \begin{bmatrix} 2.5810 & -0.1857 \\ -0.1857 & 0.0143 \end{bmatrix},$$

$$\text{and } \begin{bmatrix} \ln(\alpha) \\ \beta \end{bmatrix} = \begin{bmatrix} c_0 \\ c_1 \end{bmatrix} = \vec{c} = \left(A^\intercal A\right)^{-1} A^\intercal \vec{b} \approx \begin{bmatrix} -0.416 \\ 0.331 \end{bmatrix}.$$

Therefore, $\ln(\alpha) \approx -0.416$ and $\beta \approx 0.331$. Hence, $\alpha \approx 0.660$ and $\beta \approx 0.331$.

A2) By substituting the values of α and β into the equation, we would get the model as:

$$\ln(F) \approx 0.331C - 0.416$$

Substituting in $C = 15$, we would obtain $\ln(F) \approx 4.55$ and thus, $F \approx 94.3$.

A3) From the graph, M is the 'plateau position' where F_i is levelling off, which is 450.0. By substituting $C_i = d$, we would obtain $F_i = \frac{M}{2}$. By using the graph to find the value of C_i where $F_i = \frac{450.0}{2} = 225.0$, we would obtain $d = C_i \approx 18.3$. Hence, $M = 450.0$ and $d \approx 18.3$.

A4) We have the original equation as

$$F_i \approx \frac{M}{1 + e^{-k(C_i - d)}}, \text{ for } i \in \{1, 2, 3, \ldots, 15\}.$$

We bring over the denominator from the right-hand side to the left-hand side, obtaining

$$F_i \left(1 + e^{-k(C_i - d)}\right) \approx M$$

We would then rearrange the equation to obtain

$$e^{-k(C_i-d)} \approx \frac{M}{F_i} - 1, \text{ where (expression 1)} = \frac{M}{F_i} - 1 \text{ and (expression 2)} = -k\left(C_i - d\right)$$

Taking the natural logarithm of both sides, we would obtain

$$-k\left(C_i - d\right) \approx \ln\left(\frac{M}{F_i} - 1\right)$$

Hence, we would get $\epsilon_i(k) = (c_i - d)\,k + \ln\left(\frac{M}{F_i} - 1\right)$, where $a_i = c_i - d$ and $b_i = \ln\left(\frac{M}{F_i} - 1\right)$.

A5) We could reformulate the non-linear constraint $z \geq |f(k)|$ as two linear constraints, $z \geq f(k)$ and $z \geq -f(k)$. Hence, expression 1 is $f(k)$ and expression 2 is $-f(k)$.

A6) A general non-linear objective would be

$$\min \sum_{i=1}^{n} |f_i(k)|\,,$$

which could be linearized as

$$\left.\begin{array}{lrl} & \min\ z = & \displaystyle\sum_{i=1}^{n} f_i(k) \\ & \text{subject to:} & \\ \forall\, j \in \{1,2,3,\ldots,n\} & z & \geq f_j(k) \\ \forall\, j \in \{1,2,3,\ldots,n\} & z & \geq -f_j(k) \end{array}\right\}$$

Hence, we could rewrite the optimization problem $\mathcal{P}$ as the linear program:

$$\left.\begin{array}{lrl} & \min\ z = & \displaystyle\sum_{i=1}^{15} \epsilon_i(k) \\ & \text{subject to:} & \\ \forall\, j \in \{1,2,3,\ldots,15\} & z & \geq \epsilon_j(k) \\ \forall\, j \in \{1,2,3,\ldots,15\} & z & \geq -\epsilon_j(k) \\ & k & \geq 0 \end{array}\right\}$$

By using Excel with $M = 450.0$ and $d \approx 18.87$, we would obtain $k \approx 0.333$, with a minimum total absolute error value of $z \approx 0.457$.

A7) We have the optimization problem

$$\left.\begin{array}{lr} \min & \left(\max\{|\epsilon_1(k)|\,, |\epsilon_2(k)|\,, \ldots, |\epsilon_n(k)|\}\right) \\ \text{subject to:} & k \geq 0 \end{array}\right\},$$

which could be converted to the linear program

$$\left.\begin{array}{rll} \min & \epsilon & \\ \text{subject to:} & & \\ \forall\, i \in \{1,2,3,\ldots,n\} & \epsilon & \geq \epsilon_i(k) \\ \forall\, i \in \{1,2,3,\ldots,n\} & \epsilon & \geq -\epsilon_i(k) \\ & k & \geq 0 \end{array}\right\},$$

where ϵ is just a random variable that we have defined, instead of it being the sum of all $\epsilon_i(k)\ \forall\, i \in \{1,2,3,\ldots,n\}$, which is subtly different from the value of z in the case of question **A6**.

A8) From the graph, we could estimate the value of γ to be the value of F_i where the value of $\frac{dF_i}{dC_i}$ is at its maximum, which happens to be our value of X in this case. Hence, $\gamma \approx 219.1$.

We could rearrange the equation

$$F = (\alpha C + \beta)^{1/3} + \gamma$$

to become

$$F - \gamma = (\alpha C + \beta)^{1/3}$$

Raising both sides of the equation to the third power will yield us

$$(F - \gamma)^3 = \alpha C + \beta$$

To obtain the values of α and β, we would use regression. By using our estimate of the value of γ and by forming the matrices:

$$X = \begin{bmatrix} 1 & 8 \\ 1 & 10 \\ 1 & 12 \\ 1 & 14 \\ 1 & 16 \\ 1 & 18 \\ 1 & 20 \\ 1 & 22 \\ 1 & 24 \\ 1 & 26 \\ 1 & 28 \\ 1 & 30 \\ 1 & 32 \\ 1 & 34 \\ 1 & 36 \end{bmatrix}, \text{ and } \vec{y} = \begin{bmatrix} (8.2-\gamma)^3 \\ (19.1-\gamma)^3 \\ (37.6-\gamma)^3 \\ (70.4-\gamma)^3 \\ (145.9-\gamma)^3 \\ (219.1-\gamma)^3 \\ (296.7-\gamma)^3 \\ (367.8-\gamma)^3 \\ (378.2-\gamma)^3 \\ (403.3-\gamma)^3 \\ (424.5-\gamma)^3 \\ (436.6-\gamma)^3 \\ (445.1-\gamma)^3 \\ (446.8-\gamma)^3 \\ (448.0-\gamma)^3 \end{bmatrix},$$

$$\text{then } \left(X^\intercal X\right)^{-1} \approx \begin{bmatrix} 0.498810 & -0.019643 \\ -0.019643 & 0.000893 \end{bmatrix},$$

$$\text{and } \begin{bmatrix} \beta \\ \alpha \end{bmatrix} = \begin{bmatrix} c_0 \\ c_1 \end{bmatrix} = \vec{c} = \left(X^\intercal X\right)^{-1} X^\intercal \vec{y} \approx \begin{bmatrix} -15068952.7 \\ 809891.1 \end{bmatrix}.$$

Therefore, $\alpha \approx 809891.1$, $\beta \approx -15068952.7$ and $\gamma \approx 219.1$.

PART B

B1) Let A_i and C_i be the binary variables which is 1 if the i-th base of the 'toy' primer is A and C respectively, and 0 otherwise. Let S be the set $\{1, 2, 3, \ldots, 10\}$. The integer linear program would be

$$\left.\begin{array}{rll} \text{max} & 0 & \\ \text{subject to:} & & \\ \text{(Each base is either an } A \text{ or a } C\text{)} & A_i + C_i = 1 & \forall\, i \in S \\ \text{(The number of } C\text{'s is between 4 and 6)} & 4 \leq \sum_{i=1}^{10} C_i \leq 6 & \\ \text{(There are 2 } C\text{'s among the last 5 bases)} & \sum_{i=6}^{10} C_i = 2 & \\ \text{(Primer has 10 bases)} & A_i, C_i \in \{0, 1\} & \forall\, i \in S \end{array}\right\}$$

An example solution of the aforementioned integer linear program is $CCAAAAACAC$, which is attached in the Excel file.

B2) Let A_i, C_i, T_i and G_i be the binary variables which is 1 if the i-th base of the forward primer is A, C, G and T respectively, and 0 otherwise. Let S be the set $\{1, 2, 3, \ldots, 18\}$. The integer linear program would be

$$\left.\begin{array}{rll} \text{max} & 0 & \\ \text{subject to:} & & \\ \text{(Each base is either an } A\text{, a } C\text{, a } G \text{ or a } T\text{)} & A_i + C_i + G_i + T_i = 1 & \forall\, i \in S \\ \text{(Total } C\text{'s and } G\text{'s is between 7 and 11)} & 7 \leq \left(\sum_{i=1}^{18} C_i + \sum_{i=1}^{18} G_i \right) \leq 11 & \\ \text{(At most 3 } C\text{'s and } G\text{'s in the last 5 bases)} & \left(\sum_{i=14}^{18} C_i + \sum_{i=14}^{18} G_i \right) \leq 3 & \\ \text{(Similar annealing temperature)} & 4 \times \left(\sum_{i=1}^{18} C_i + \sum_{i=1}^{18} G_i \right) + 2 \times \left(\sum_{i=1}^{18} A_i + \sum_{i=1}^{18} T_i \right) = 58 & \\ \text{(Primer has 18 bases)} & A_i, C_i \in \{0, 1\} & \forall\, i \in S \end{array}\right\}$$

A solution forward primer would be $CCCCCCCCCCCAAAAAAA$, which is attached in the Excel file.

B3) We decided to implement Constraints 1 and 3 in Excel, with a 'lucky bonus' of Constraint 2 being automatically satisfied without explicitly stating it in Excel's Solver.

Constraint 1 dictates that the forward primer has no 'runs' of length more than 4. We can impose this constraint in Excel by setting every set of 5 consecutive bases to be less than or equal to 4. Let K be the set $\{1, 2, 3, \ldots, 14\}$. Thus,

$$\sum_{i=k}^{k+4} B_i \leq 4 \ \forall k \in K, \forall B \in \{A, C, G, T\}$$

This would yield us an additional 56 constraints.

Constraint 2 dictates that the forward primer has no alternating patterns of the form $XYXY$. Although we did not use this constraint in Excel's Solver (cells highlighted in yellow), we could implement it by choosing two types of bases at a time for a specific base pattern and impose that the sum of every 4 consecutive bases alternating between the two types of bases has to be less than or equal to 3. For example, to eliminate the pattern $ACAC$, we choose A and C and we set the constraints

$$A_i + C_{i+1} + A_{i+2} + C_{i+3} \leq 3 \ \forall i \in \{1, 2, 3, \ldots, 15\}$$

Similar constraints are set up for the patterns $CACA$, $GTGT$, $AGAG$, $ATAT$, $CGCG$, $CTCT$, $GAGA$, $GCGC$, $TATA$, $TCTC$ and $TGTG$.

This would yield us an additional 180 constraints, which is impossible to be considered by Excel. Thus, we do not explicitly add these constraints to Excel's Solver.

Constraint 3 dictates that when the forward and reverse primers are aligned side-by-side, the number of alignments between A's from one primer and T's from the other is less than or equal to 4, and the same goes for C's and G's. Since the reverse primer is a constant, we can use it to build our constraints for the forward primer. First, we note down the positions of A and T bases in the reverse primer. Then, we find the sum of the binary variables of the A bases in the forward primer in the same positions as the T bases in the reverse primer, as well as the sum of the binary variables of the T bases in the forward primer in the same positions as the A bases in the reverse primer. We add these sums together which would give us the total number of A-T alignments between the forward and reverse primers. Then, we impose that this final sum should be less than or equal to 4. Repeat for G-C alignments.

Hence, we would obtain

$$A_5 + A_8 + A_{12} + A_{14} + T_3 + T_{11} + T_{17} \leq 4$$

and

$$C_1 + C_2 + C_6 + C_9 + C_{15} + G_4 + G_7 + G_{10} + G_{16} + G_{18} \leq 4$$

A possible solution for the forward primer under these 3 constraints would be $CCCCACCCCTCCCTAAAA$.

B4) Instead of having the constraint $A_{18} + C_{18} + G_{18} + T_{18} = 1$, replace it with $A_{18} + C_{18} + G_{18} + T_{18} \leq 1$.

B5) Let Ar_i, Cr_i, Tr_i and Gr_i be the binary variables which is 1 if the i-th base of the reverse primer is A, C, T and G respectively, and 0 otherwise.

To form a constraint for this model, firstly we multiply the binary decision variables for A in the forward primer with the binary decision variables for T in the reverse primer at the respective positions. Then, we sum up the products at each position. For this, we would get

$$\sum_{i=1}^{18} (A_i Tr_i)$$

Similarly, we multiply the binary decision variables for T in the forward primer with the binary decision variables for A in the reverse primer at the respective positions. Then, we sum up the products at each position. For this, we would get

$$\sum_{i=1}^{18} (T_i Ar_i)$$

We would then require the sum of the two sums produced to be less than or equal to 4, i.e.:

$$\sum_{i=1}^{18} (A_i Tr_i) + \sum_{i=1}^{18} (T_i Ar_i) \leq 4$$

These steps are similarly repeated for the alignment between the C and G bases.

Thus, to model the alignment constraint, we would have to implement the following two constraints:

$$\sum_{i=1}^{18} (A_i Tr_i) + \sum_{i=1}^{18} (T_i Ar_i) \leq 4$$

$$\sum_{i=1}^{18} (C_i Gr_i) + \sum_{i=1}^{18} (G_i Cr_i) \leq 4$$