

Question 2 (3' + 3' + 4' = 10')

The BPR function gives the link travel time as a function of the link flow rate q ,

$$TT(q) = TT_f \cdot \left[1 + \alpha \cdot \left(\frac{q}{c} \right)^\beta \right]$$

Here,

TT_f – travel time under free flow conditions,

c – design capacity,

α, β – constants.

(a) What practical phenomenon does the constant β signify? Give a typical value of β .

It signifies the delay caused by congestion (or: traffic volume close to or above design capacity). A typical value is 4 (or: at least >1).

(b) The BPR function is based on the assumption that flow arrives in a uniform pattern (i.e. same speed, equal headways). In reality it is usually not the case. How does this assumption affect the accuracy of the results?

Travel time can be much longer for periods with (temporary) high flow rate. To achieve higher accuracy, instead of using (average) flow rate for each hour, we separate the time horizon into shorter segments; in each time segment, the arrivals are (almost) uniform.

(c) In order to calibrate the BPR function for a particular network link, what procedures have to be taken?

- 1) Obtained from the designer: c, TT_f ; (can also be from experiment/observation)
- 2) Obtained from field observations: $q, TT(q)$; (a number of observations under different q are necessary)
- 3) Results of the regression: α, β

Question 4 (4' + 4' + 2' = 10')

The Webster formula gives the average delay per vehicle at a signalised intersection as

$$\bar{W} = \frac{C(1-u)^2}{2(1-y)}$$

Here,

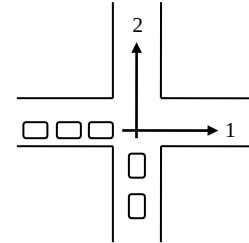
C – cycle time (s),

u – effective green ratio,

y – flow ratio, $y:=q/s$,

q – flow arrival rate (veh/hr),

s – flow departure rate (veh/hr).



Consider a particular intersection where only two movements are present (see Figure). The flow rates are given as

Movement	q	s
1	900	1800
2	600	1800

(a) Given that $C=60s$ and $u_1=u_2=0.5$, calculate the average delay for each movement.

$$y_1 = 1/2, y_2 = 1/3$$

$$W_1 = C (1-u_1)^2 = 15 s$$

$$W_2 = 3/4 C (1-u_2)^2 = 11.25 s$$

(b) Given that $C=60s$ and $u_1+u_2=1$, design a signal pattern so that the total delay is minimised.

$$\begin{aligned} \text{Total delay} &= 900 W_1 + 600 W_2 = 900 C (1-u_1)^2 + 450 C (u_1)^2 \\ &= 1350 C (u_1)^2 - 1800 C u_1 + 900 C \end{aligned}$$

$$\arg \min_{[u_1]} \{\text{total delay}\} = 2/3; \quad u_2=1-u_1=1/3$$

(c) Discuss how, according to the model as well as in practice, the duration of C affects the results of (a) and (b).

According to the model, average delay is proportional to C , as in (a); but the optimal signal pattern (u_1, u_2) is independent of C , as in (b). This is if we assume $u_1+u_2=1$.

In practice, all-red time means $u_1+u_2<1$. The value of $1-(u_1+u_2)$ is inversely proportional to C . In (a), delay would be very long if C is too long, or if C is too short. In (b), the optimal signal design would be dependent of C ; or, there is also an optimal C .

Question 5 ($5' \times 4 = 20'$)

A suburb road has only one lane in each direction. For safety reasons, overtaking is not allowed (double solid centrelines). On a particular day, traffic had been travelling steadily at speed 60km/h with headway 2s. At time t , a truck entered this road at junction n . The truck had a speed of 30km/h. Accordingly, upstream vehicles had to adjust to this speed and as such formed a queue behind the truck. It is known that traffic density in this queued area is x (veh/km). The truck left the road at junction $n+1$, which is 1km downstream of junction n . Afterwards vehicles in the queue started to resume their normal speed.

a) Calculate the shockwave speed for the end of the queue (as an expression of x).

Arrival flow: $q=1800$ veh/hr, $k=30$ veh/km, $v=60$ km/h;
In queue: $q=30x$, $k=x$, $v=30$.

Shockwave at the end of the queue: $w=(30x-1800)/(x-30)$

b) Analyse the movement of the queued area over time.

End of the queue:

If $x > 60$, moving forward (downstream);

If $x = 60$, stationary;

If $30 < x < 60$, moving backward (upstream).

[Note: x must be > 30 , otherwise there cannot be any queued area.]

Front of the queue:

Before the truck leaves, same as the truck: +30 km/h;

After the truck leaves, same as the end of the queue.

[Note: because the dispatching flow is the same as the arrival flow, the queue would actually never disappear.]

c) Discuss the practical relevance of shockwaves in general, and the implications for traffic management that we can draw from queue movements.

Shockwave helps traffic management understand the formation and movement of congested areas. This has implications in network design (bottleneck vs. capacity), traffic signal design, speed control (e.g. variable speed limit), in order to avoid or mitigate the impact of traffic jams.

d) Shockwave is a macroscopic description of traffic flow. In the scenario above, we have made some implicit assumptions on a microscopic level. Discuss whether these assumptions are realistic or not and how they affect the macroscopic modelling results.

The major assumption is that flow is constant on both sides of the shockwave (boundary). This means the drivers change their 'state' (car following: speed, headway) right at the instant they cross the shockwave, i.e. infinite ac-/deceleration, no anticipation & no reaction time.

In practice, shockwave is not a line curve but a region, where traffic shift gradually from one state to another.

Question 5

Principal Component Analysis (PCA) and Cluster Analysis are two data editing methods.

a) What is the difference between PCA and clustering. You may give examples.

PCA: constructing components out of variables → combining variables and reducing the number of “relevant variables”

Clustering: constructing groups or clusters out of items → combining items and reducing the number of “relevant items”

Consider the following 6 data points (or items): [1,3], [2,4], [3,4], [5,2], [6,4] and [8,1].

b) Plot these points in a graph. Use euclidean distance to cluster these items according to the **nearest** neighbour method. For each iteration, give the clusters and the cluster size (of the new clusters). The latter is the distance between the data-points that are connected in the given iteration.

Graphic process which you need to do yourself

c) Plot the same points in a new graph. Use euclidean distance to cluster these items according to the **farthest** neighbour method. For each iteration, give the clusters and the cluster size (of the new clusters).

Graphic process which you need to do yourself

d) Make a dendrogram (number of clusters vs. cluster size) for both clustering methods. How many clusters would you choose according to the dendrogram? Suppose you choose to cluster the points in 2 clusters: use the answers from b) and c) to draw these two clusters in the respective graphs. Considering both clustering results, which clustering method would you use in this case? Explain your answers.

Nearest neighbour → clusters {1, 2, 3, 4, 5} and {6}

Furthest neighbour → clusters {1, 2, 3} and {4, 5, 6}

Nearest neighbour method construct clusters that are “line strings”. In this case, the furthers neighbour method is preferred.

Question 6

Consider observed route choices of two different OD pairs. For both OD pairs, 3 routes occur in the observations. These are denoted by r_1, r_2, r_3 .

- The first OD-pair yields: $[r_1, r_2, r_2, r_3, r_1, r_1, r_3, r_1, r_2, r_3]$
- The second OD-pair yields: $[r_1, r_1, r_2, r_1, r_1, r_1, r_3, r_1, r_1, r_1]$

We now consider two route choice models in which the model probabilities are the same for all travelers per OD pair.

- For the first OD pair, the first model yields: $p_1 = 0.40, p_2 = 0.35, p_3 = 0.25$
- For the second OD pair, the second model yields: $p_1 = 0.60, p_2 = 0.20, p_3 = 0.20$

Give the log likelihoods for the first model-OD pair and second model-OD pair. Also, give the maximum log likelihood for both OD pairs. These maximum likelihoods are the likelihoods for which the probabilities p_1, p_2, p_3 (of r_1, r_2, r_3 respectively) describe the observations in the best way. Comparing the log likelihoods of the models with the maximum likelihoods, which model do you think is better? Explain your answer.

First OD pair: $\ln L = \ln(p_1^{n_1} p_2^{n_2} p_3^{n_3}) = 4 \ln(0.40) + 3 \ln(0.35) + 3 \ln(0.25) = -11.0$
Second OD pair $\ln L = 8 \ln(0.60) + 1 \ln(0.20) + 1 \ln(0.20) = -7.3$

Maximum likelihoods first OD pair if $p_1 = 0.40, p_2 = 0.30, p_3 = 0.30$

$$\rightarrow \ln L = 4 \ln(0.40) + 3 \ln(0.30) + 3 \ln(0.30) = -10.9$$

Maximum likelihoods second OD pair if $p_1 = 0.80, p_2 = 0.10, p_3 = 0.10$

$$\rightarrow \ln L = 8 \ln(0.80) + 1 \ln(0.10) + 1 \ln(0.10) = -6.4$$

The first model is better, because the difference between the likelihood and the maximum likelihood is almost 0 (the likelihood ratio is almost 1), despite the fact that the “absolute” likelihoods are lower in the second case. The reason for this is that there is a clear preference for one route (route 1) in the second case.

Consider two cycle paths in opposite directions. For both paths, daily volumes were measured for 10 days. For cycle path 1 and 2, the volumes are given by q_1 and q_2 respectively:

- $q_1 = [299, 398, 871, 255, 987, 1065, 340, 342, 472, 192]$
- $q_2 = [313, 397, 877, 259, 1001, 1061, 345, 347, 479, 198]$.

a) Are the mean volumes $mean(q_1)$ and $mean(q_2)$ significantly different? Use a 2-sigma confidence level. Explain all steps.

The series are paired $\rightarrow$ define $z = q_2 - q_1$: is $average(z) = 0$?

$\rightarrow Abs(Average(z)) < 2*stdev(average(z))$

$Average(z) = 5.6$, $stdev(average(z)) = stdev(z) / \sqrt{(number\ of\ pairs)} = 5.6 / \sqrt{10} = 1.8$

$Average(z) = 5.6 > 3.6 = 2*stdev(average(z))$

$\rightarrow$ The mean volumes are significantly different

The volumes of cycle path 1, q_1 , are fitted by the following model, q_{model} :

$q_{model} = q_0 \exp[0.1(t_s - t_p)]$, with

- q_0 : the volume during “average” weather conditions
- t_s : the duration of sunshine (in hours)
- t_p : the duration of rain (in hours)

From the fit the values 0.1, and $q_0 = 500$ (per day) were obtained.

From the weather station, we got: $t_s = [2, 4, 7, 1, 9, 9, 3, 3, 5, 0]$ and $t_p = 10 - t_s$.

The total amount of hours during the day (excluding the night) was thus 10, and there is a correlation of -1 between t_s and t_p (note that in reality the correlation is less strong, but this was done for convenience).

The errors are: $err(q_0) = 10$, $err(t_s) = 0.2$ and $err(t_p) = -0.2$ (the negative sign indicating the negative correlation with $err(t_s)$). The errors in q_0 and errors in t_s (and in t_p) are uncorrelated. Finally, there is random variation in the observed q_1 . This variation is estimated to be $err(q_1) = \sqrt{q_{model}}$

b) Give an approximation of the model error, $err(q_{model})$, as function of $err(q_0)$, $err(t_s)$ and $err(t_p)$. Put the values of $err(q_0)$, $err(t_s)$ and $err(t_p)$ in your approximation

$err_{q_0}(q_{model}) = \exp[0.1(t_s - t_p)] * err(q_0) = 10 * \exp[0.1(t_s - t_p)]$

$err_{t_s}(q_{model}) = 0.1 * q_0 * \exp[0.1(t_s - t_p)] * err(t_s) = 0.02 * q_0 * \exp[0.1(t_s - t_p)]$

$err_{t_p}(q_{model}) = -0.1 * q_0 * \exp[0.1(t_s - t_p)] * err(t_p) = -0.02 * q_0 * \exp[0.1(t_s - t_p)]$

Negatively correlated $\rightarrow err_{t_s+t_p}(q_{model}) = err_{t_s}(q_{model}) - err_{t_p}(q_{model})$

Note: minus sign because individual errors in t_s and t_p are negatively correlated. In the expression, the errors in t_s and t_p are absolute “mean” errors (usually the standard deviations of the respective distributions of individual errors), both with a positive value.

$\rightarrow err_{t_s+t_p}(q_{model}) = 0.04 * q_0 * \exp[0.1(t_s - t_p)] = 20 * \exp[0.1(t_s - t_p)]$

Errors in t_s , t_p and q_0 are uncorrelated: quadratic errors add up

$\rightarrow err(q_{model}) = \sqrt{err_{q_0}(q_{model})^2 + err_{t_s+t_p}(q_{model})^2} = \sqrt{500^2 * \exp[0.2(t_s - t_p)]}$

c) Give the model values q_{model} , model errors $err(q_{model})$, “observation” errors $err(q_1)$ and total errors of model and observation combined for all 10 days.

$q_{\text{model}} = [274, 409, 746, 225, 1113, 1113, 335, 335, 500, 184]$
 $\text{err}(q_{\text{model}}) = [12, 18, 33, 10, 50, 50, 15, 15, 22, 8]$
 $\text{err}(q_1) = [17, 20, 27, 15, 33, 33, 18, 18, 22, 14]$
 $\text{err}_{\text{tot}}(q) = [21, 27, 43, 18, 60, 60, 24, 24, 32, 16]$

d) Perform a Chi-square goodness of fit test for the model. Should the model be rejected?
 Take a 99% confidence interval.

$\chi^2 = \sum [(q_{\text{model},i} - q_{1,i})^2 / \text{err}_{\text{tot}}(q_i)^2] = 19.05$
 10 items, and 2 coefficients fitted to the measurements $\rightarrow Df = 10 - 2 = 8$
 Chi-square table: if $Df = 8$, $\chi^2_{0.01} = 20.1$
 $19.05 < 20.1 = \chi^2_{0.01} \rightarrow \text{model is not rejected}$