COWLES COMMISSION DISCUSSION PAPER: ECONOMICS NO. 2072

Some Models of Traffic Congestion [1/]

C. B. Winsten*

April 1, 1953

Introduction

1.1 This paper forms part of a study of the economics of highway systems. Recently Beckmann has discussed the effect of congestion in highways on the routing of traffic [2], and McGuire has made a survey of the work of highway engineers and others, to find what properties of congested systems could justifiably be assumed in such work [6]. In [2] the capacity functions used, had, for lack of information, to be very aggregative. The highway capacity was framed in terms of average speed; that of the intersection in terms of total flow, regardless of direction. They could not, therefore, show the social cost of a given vehicle's trip as a function of the driver's particular desired speed or course across an intersection. This function is of considerable economic interest. Consequently the phenomena which it seems most useful to examine in continuing this program are

(i) congestion at intersections

(ii) congestion of cars travelling with different desired speeds on the highway.

These are the topics with which we will be concerned below.

1/ Research undertaken under contract between the Cowles Commission for Research in Economics and the RAND Corporation.

* On leave from the Oxford University Institute of Statistics.

1.2 It is, of course, possible to carry out actual observation of congested systems; without an underlying theoretical model, however, the results of such observations are very difficult to interpret.

In this paper, therefore, we will be mainly concerned with the construction of models to represent traffic phenomena. These models will be very simplified, but they should be realistic enough to provide insight into actual traffic experience.

Their properties can be studied analytically; where the analysis proves too difficult, it is possible to perform "Monte Carlo" experiments. These correspond to performing controlled experiments with particular types of traffic streams; as they can be done on punched card or similar equipment they are often an economical method of research. The representations of traffic schemes make use of random variables; this is because such representations have been found realistic in many cases [1,7].

1.3 In section 2 we discuss traffic intersection models.

In section 3 we show how similar models can be used to describe a highway with cars travelling with just two desired speeds.

In section 4 we find the mean waiting time for one of the models described. Fuller analytic treatment is reserved for a later discussion paper.

1.4 Though the models are framed in terms of road traffic, they have other possible applications. We mention two examples.

(i) On railways, one is frequently faced with the situation in which slow and fast tracks merge for a stretch. Fast trains may be given priority so that slow trains would have to queue on their line till there was sufficient gap in the fast traffic. In considering what assumptions to make about the arrival of slow traffic, we have to remember that a railroad is a directed system, and so it is even more necessary to consider the network as a whole than in the traffic case.

(ii) At an airport, two types of planes may be handled, of which one may have priority. This leads to queuing of the non-priority traffic. To develop this analogy would require more detailed knowledge of airport operation than the present writer possesses.

An intersection problem.

2.1. We will discuss the minor road, or "stop sign" problem.[2/] In this case a minor road, intersects a major road. The minor road traffic is not supposed to interfere with the major road traffic at all; it must therefore wait till there is a sufficient gap in the major road traffic for it to cross. We suppose traffic in the minor road restricted to one lane for each direction; we can therefore consider one lane at a time. Since only one car can cross the major road at a time, a queue can form of cars waiting to cross. Our object is to examine the mean time cars in the minor road have to wait in this queue.

We suppose that cars in the minor road arrive with an interval of time of at least s secs. between each other. The car at the head of the queue can cross provided there is no car in the major road due to arrive within time less than T. Usually s is considerably less than T. Intervals during which a car would not be able to cross the major road are called blocks, those in which it would are called antiblocks. (see e.g., 7).

2/ Similar methods can be applied to a traffic light problem. See below.

A discussion of a rather similar problem to this is given by Tanner [8]. He considers the case of pedestrians waiting to cross a road, and assume that everybody who is waiting can cross the road as soon as there is an antiblock. Such an assumption is not realistic for cars waiting to cross. There is a minimum interval between two cars crossing the road during an antiblock, so that it may take some time for a queue to clear itself, and we must take account of this in our model. We will assume the minimum time between departures is s secs., the same as the minimum time between arrivals. Only if there was very sparse traffic in the minor road would Tanner's results apply. Then there would never be more than one car waiting to cross, so the departure spacing would not be relevant.

In the analysis of this system, we have to make assumptions about the arrival time distribution and the block-antiblock sequence. We have attempted this with some of the more obvious assumptions, but the problem is untidy and difficult in the more realistic cases, and we have found the following device produces very important simplifications. Instead of thinking of time as a continuous variable, replace the model by one in which time is discrete. The meaning of this will become clearer below.

2.2. We take the discrete time points at which we observe the intersection as being at equal intervals of length s seconds apart. We will suppose that at each time point at most one car can arrive in the minor road (this corresponds to our previous assumption that there was an interval of at least s seconds between arrivals of cars.). At each instant, too, there is either a block or an antiblock in the major road.

If there is an antiblock at a time point, then just one car can cross the intersection, provided there is a car there to cross (i.e. provided a car has arrived at that instant, or a queue has been held over from earlier). A sequence through time of events at the intersection can be represented as follows:

Time	... t_{-2}	t_{-1}	t_0	t_1	t_2	t_3	.....
Arrivals in minor road	1	0	0	1	0	1	
Blocks caused by traffic in major road	x	x	x	∘	∘	x	

For any particular time interval t_r,

1 indicates the arrival of a car in the minor road

0 indicates no car arrives in the minor road

x indicates there is a block in the major road

∘ indicates there is an antiblock in the major road

From these two sequences, we can find the queue of cars waiting to cross at each time point, and the sequence of departures of cars from the intersection.

Suppose the queue of cars held over from t_r to t_{r+1} contains q_r cars. Let a_r be the number of cars arriving at t_r ($a_r = 0, 1$).

If there is an antiblock at t_r

$$q_r = q_{r-1} + a_r - 1$$

unless $q_{r-1} + a_r = 0$, when $q_r = 0$.

If there is a block at t_r

$$q_r = q_{r-1} + a_r$$

Thus, if we know the queue size at some point t_0 say, we can find the queue sequence for later time from the arrival and block sequence.

For work with actual sequences, such as we do in Monte Carlo experiments, it is useful to define

$$q_r' = q_{r-1} + a_r$$

the supply of cars waiting to cross at t_r.

We can now obtain the departure sequence, by comparing the q_r' sequence with the block sequence.

Let $d_r = 1$ if there is a departure at t_r

$d_r = 0$ if there is not.

Then $d_r = 1$ if (i) there is an antiblock at t_r

and (ii) $q_r' > 0$

If both these conditions are not satisfied $d_r = 0$.

Here is an example of the derivation of the queue and departure sequences from the block and arrival sequences. We omit the q_r' sequence.

Arrivals	a_r	0	0	1	1	0	1	0	1	0	0	1	0	1	0
Block		x	x	x	∘	∘	∘	x	x	∘	∘	x	x	x	∘
Queue	q_r	0	0	1	1	0	0	0	1	0	0	1	1	2	1
Departure	d_r	0	0	0	1	1	1	0	0	1	0	0	0	0	1

The reason for the simplification of this discrete time model, both for mathematical analysis and for Monte Carlo experiments may now be clearer. In generating the queue and departure sequences, we have to carry very little information over from one time instant to the next.

2.3. So far we have not considered any hypothesis about the nature or the mode of generation of either the arrival sequence, or the block-antiblock sequence. Obviously there are many possibilities, and practical usefulness should be a principal guide. Unfortunately there is not a great deal of empirical data on the structure of congested traffic flows. We have investigated analytically the following cases:

2.4. Arrival sequence

This is generated by a probability model. It is taken to be a binomial sequence; this means that the arrival of a car at t_r has a probability p which is independent of what happens at other t_r. p is the same for all t_r.

This model is an analogue in discrete time of the Poisson model in continuous time. The Poisson model (or a modification of it to allow for safe distance between cars) has been used quite often before to describe traffic under various conditions. In the discrete model we do not have to make special allowance for safe distance - that is done automatically by our formulation of the model.

The block, antiblock sequence

2.5 (i) The General/Geometric sequence.

In this model, the lengths of blocks b (measured by the number of time instants they cover) are independently drawn from some distribution $g(b)$. We leave this quite general for this discussion. In between each block there must, of course, be at least one antiblock point. With this restriction the lengths of antiblocks a are distributed in the geometric distribution, $\pi^a(1-\pi)$.

The geometric distribution in discrete time corresponds to the negative exponential distribution in continuous time.

We use this formulation since it covers a rather wide class of cases. Amongst these are the block distribution formed from a binomial stream of traffic in the major road, when there is a safe interval T in front of each car during which a minor road car will not cross. For the discrete model we may suppose T to be a constant multiple of the safe time between cars; this safe time we suppose the same on both roads.

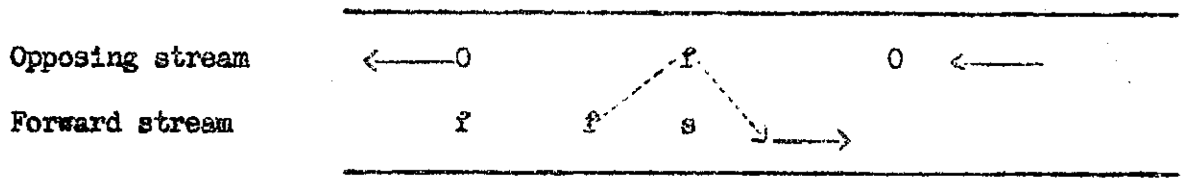
The case also covers streams of cars in the major road which have been through some congestion process, so that the cars tend to arrive in clusters. It also covers the blocks caused by two way traffic consisting of two streams of this type. It is not difficult to find the relevant properties of the $g(b)$ distribution, and we will give them in a later paper with more detailed calculations. The derivation of similar distributions in the continuous case has been discussed for Poisson streams of traffic [4, 7, 8]; a formally equivalent problem connected with Geiger counters has also received a great deal of attention. [3]

2.6. (ii) The fixed block and antiblock length sequence.

This is important for consideration of traffic lights, which are often made to work this way, and also for equally spaced cars in the major road. This problem requires rather more elaborate methods for solution than the case given above and we will give a discussion of it in a separate paper.

3.1. <u>A model for overtaking on a two lane highway</u>

In this model, we consider two streams of traffic, the forward and opposing streams.



The opposing stream must not be disturbed by the forward stream.

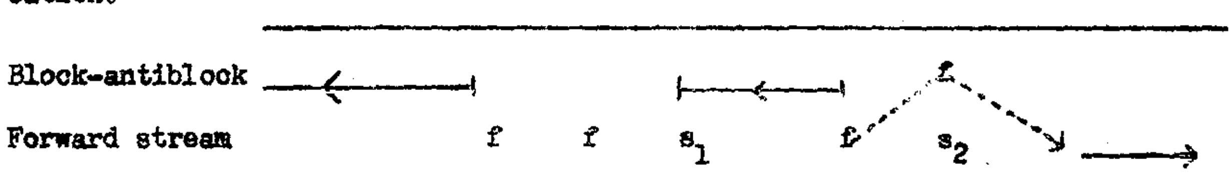
In the forward stream is a slow car, and a stream of fast cars, each with the same desired speed, say. The fast cars can only overtake the slow car if there is a sufficient gap in the opposing stream. Thus queue of fast cars can form behind the slow car.

If we imagine we are looking at this system from the slow car, it will be seen that this problem is formally similar to the "stop sign" intersection problem. The fast cars in the forward stream take the place of the minor road traffic. The opposing stream takes the place of the major road traffic. Passing the slow car is equivalent to crossing the intersection. The analysis we use for the intersection model also serves for this case.

3.2. Complications arise if there is more than one slow car. Suppose, for example, there are two slow cars, far apart on the road, but travelling at the same speed. If the arrival distribution of cars behind the first slow car is binomial, then for the types of block and antiblock distribution we have considered, the sequence of departures, i.e. of cars which have passed the first slow car, will not be binomial. If the fast cars travel at the same speed this will be the arrival distribution behind the second slow car. The distribution will be modified again on passing the second car, so there will be a further problem if there is a third slow car in the stream.

3.3. These cases can be examined by Monte Carlo methods, but a possible analytic approach which is being explored follows: It seems plausible that, after passing sufficient slow cars, the arrival and departure distributions will settle down to stability, i.e. the arrival distribution behind successive slow cars will have the same form. This distribution should be to a considerable extent independent of the arrival distribution behind the first car. This has not yet been proved though, and it seems worthwhile to examine the question by Monte Carlo methods, which would in any case also show how quickly the stability is reached, if it is reached.

3.4. We have postulated that the slow cars should be sparsely scattered on the road. This is because we wished to avoid the following complication:



In this diagram there are two slow cars x_1 and s_2; suppose the opposing cars generate a block-antiblock distribution as shown. Then if an antiblock passess s_2, allowing a fast car to pass s_2, it will shortly afterwards pass s_1. Thus fast cars passing s_1 will normally find a shorter queue behind s_2 than if there were no relation in the block-antiblock distributions passing the two slow cars, which would be true if the slow cars were far enough apart. The effect of having the slow cars close shows the waiting time behind the second slow car is decreased on the average.

4.1. Derivation of the mean waiting time

General / geometric case, with binomial arrivals.

We use the discrete model formulated above, and take the interval between time points (which we have associated with the minimum safe interval between cars) as the unit of time.

Let the probability of arrival at any time point be α.

Suppose the distribution of block length b is $g(b)$

$$(b = 1, 2, 3 \ldots.)$$

The distribution of antiblock length is geometric with parameter π say. Thus the probability that an antiblock selected at random will be of length a is $\pi^{a-1}(1 - \pi)$. The important property of this distribution is that, at any antiblock point, the probability the next point will be an antiblock point is π, regardless of how long the antiblock has already lasted.

We suppose, for the present derivation at least, that the process has settled down to a state of statistical equilibrium.

Note that a queue of length q_r at time t_r implies that q_r cars must wait from t_r to t_{r+1}.

Thus the average time waited per time point is $E(q_r)$, so the average time waited per car is $\frac{1}{\alpha} \; E(q_r)$.

4.2. We will find $E(q_r)$ in two stages, first taking averages over antiblock points, and then over block points.

Suppose t_s is the sequence of antiblock points. The points are supposed numbered consecutively, regardless of whether there is a block between them or not. Because antiblocks have a geometric distribution, the probability of a block occuring between any two consecutive time points, t_s and t_{s+1} say, is $(1 - \pi)$, and is independent of where other blocks are interspersed in the sequence.

Define the variables

b_s = no. of block points between t_s and t_{s+1}.

β_x = no. of arrivals at block points between t_s and t_{s+1}

α_s = no. of arrivals at t_s $(\alpha_s = 0, 1)$

If $b_s = 0$, then necessarily $\beta_s = 0$.

Then $q_{s+1} = \max(q_s + \alpha_s + \beta_s - 1, 0)$

or we may write

(1) $$q_{s+1} = q_s + \alpha_s + \beta_s - 1 + \delta_s$$

(2) where $$\begin{aligned} \delta_s &= 0 \quad \text{if } q_s + \alpha_s + \beta_s > 0 \\ \delta_s &= 1 \quad \text{if } q_s + \alpha_s + \beta_s = 0 \end{aligned}$$

$E(q_s)$ can now be found by a device used for a similar problem by Kendall [5].

In equilibrium, $E(q_{s+1}) = E(q_s)$

(3) $$\therefore \; E(\delta_s) = 1 - E(\alpha_s) - E(\beta_s)$$

(4) also $$E(\delta_s^2) = E(\delta_s) \quad \text{since} \quad \delta_s = 0 \text{ or } 1.$$

(5) By (2) $$E\left\{\delta_s (q_s + \alpha_s + \beta_s)\right\} = 0.$$

Squaring (1), and using (3), (4), (5)

$$(6) \qquad \left\{2\,E\,(q_s) + 1\right\} \left\{1 - E(\alpha_s) - E(\beta_s)\right\} = E\,(\alpha_s + \beta_s - 1)^2 .$$

To find $E(\beta_s)$ note that the probability that there is a block between t_s and t_{s+1} is $(1 - \pi)$, and the mean no. of arrivals, if there is a block, is $\alpha E(b)$

$$(7) \qquad \therefore E(\beta_s) = \alpha(1 - \pi)\,E(b).$$

In the same way

$$(8) \qquad E(\beta_s^{\,2}) = (1 - \pi)\left\{\alpha^2\,E(b^2) + \alpha(1 - \alpha)\,E(b)\right\}$$

by successively taking expectations over blocks of length b, and then over all blocks.

(9) Also $\quad E(\alpha_s) = E(\alpha_s^{\,2}) = \alpha$

Using these relations and the independence of α_s, β_s, we find from (6) that

$$(10) \qquad E(q_s) = \frac{\alpha^2(1 - \pi)\left\{E(b^2) + E(b)\right\}}{2\left\{1 - \alpha - \alpha(1 - \pi)\,E(b)\right\}} .$$

4.3. Now we must consider the queue length over block points.

Suppose a block of length b comes between t_s and t_{s+1}. We can number the points of the block t_{s1}, t_{s2} ....t_{sb}. The queue length just before the start of the block will be q_s, and these cars will remain throughout the block. We can therefore limit ourselves to considering additions to the queue during the block.

If there is an arrival at t_{s1}, it will be in the queue at $t_{s1}, t_{s2}, \ldots, t_{sb}$.

It will thus add b to the sum of the queues over the block.

Similarly, an arrival at t_{s2} adds $b - 1$ to the sum.

Thus mean addition to queues per block of length b caused by arrivals during such blocks $= \frac{b(b+1)}{2}\alpha$.

$\therefore$ mean addition per block point in blocks of length b, caused by arrivals in those blocks $= \frac{b+1}{2}\alpha$.

A proportion $\frac{b\,g(b)}{E(b)}$ of block points lie in blocks of length b.

$\therefore$ mean addition per block point caused by arrivals during

$$\text{blocks} = \Sigma \frac{b(b+1)}{2} \frac{g(b)}{E(b)} \alpha$$

$$= \frac{\alpha}{2\,E(b)} \quad E(b^2) + E(b)$$

Ratio of block points to antiblock points is

$$(1 - \pi)\,E(b) : 1 \tag{11}$$

$\therefore$ mean addition per point

$$= \frac{(1-\pi)\alpha}{2} \quad \frac{E(b^2) + E(b)}{1 + (1-\pi)\,E(b)} \tag{12}$$

Thus mean queue size per point, from (1) and (12) is

$$\frac{\alpha(1-\pi)\,\{E(b^2) + E(b)\}}{2\,\{1 + (1-\pi)\,E(b)\}\,\{1 - \alpha - \alpha(1-\pi)\}} \tag{13}$$

and, as we showed before, mean waiting time per car is $\frac{1}{\alpha} E(q_r)$.

Note that $\left\{1 - \alpha - \alpha(1 - \pi) E(b)\right\}$ has an obvious intuitive meaning. The average no. of arrivals per block point is α. The average no. of points during an antiblock where there are no arrivals is $(1 - \alpha)$, and at these points the queue has a chance to lessen. Since the ratio of block points to antiblock points is

$$(1 - \pi) E(b) : 1$$

the ratio of no. of arrivals in a block to no. of possible queue reduction points is

$$\frac{\alpha(1 - \pi) E(b)}{1 - \alpha} .$$

I have discussed this work with Mr. C. B. McGuire at all stages, and he drew my attention to many of the problems. I am also indebted to Professor W. Feller, Mr. D. G. Kendall and to Professor T. Koopmans and other members of the Cowles Commission for valuable discussion.

Bibliography

[1] Adams, W. F., "Road traffic considered as a Random Series," Journal of the Institute of Civil Engineers, p: 121-130 (1936).

[2] Beckmann, M., "Efficient transportation in networks," Cowles Commission Discussion Paper No. 2049A.

[3] Feller, W., "On probability problems in the theory of counters," Studies and Essays presented to R. Courant, p. 105-115, New York 1948.

[4] Garwood, F., "An application of the theory of probability to the operation of vehicularly controlled traffic signals," Journal of the Royal Statistical Society (Supplement) 7:65-77 (1940-41).

[5] Kendall, D. G., "Some problems in the theory of queues," Journal of the Royal Statistical Society (Series B), 13 pp. 151-185 (1951).

[6] McGuire, C. B., "Highway capacity and traffic congestion," Cowles Commission Discussion Paper No. 2048.

[7] Raff, M. S., "The distribution of blocks in an uncongested stream of automobile traffic," Journal of the American Statistical Association 46:114 (1951).

[8] Tanner, J. C., "The delay to pedestrians crossing a road," Biometrika 38:383-392 (1952).