COWLES COMMISSION DISCUSSION PAPER: ECONOMICS NO. 2094

The Classification of Railway Cars: A Probability Approach[1/]

C. B. Winsten

December 17, 1953

Introduction

A problem is considered in the allocation of work between two different types of classification yard. To do this a probability model of the input is constructed, and costs of doing the work at the two types of yard are considered. A solution is given of a probability problem connected with the efficient allocation of work.

1. At a railway switching yard, cars are sorted, to be made up into trains. There are basically two types of yards, hump and flat, and in each of them the sorting is done by putting the cars on to a set of tracks called classification tracks; cars of different categories are put on different classification tracks. The two types of yards are mainly distinguished by the way this operation is performed. In a flat yard the cars are pushed and pulled on to their appropriate tracks by a switch engine. Though the yards are usually designed in a way calculated to reduce the to and fro motion of the

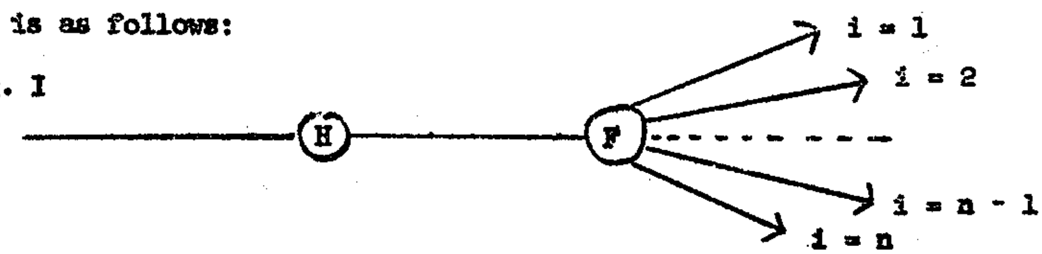
---

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

engine in doing this job, it does seem generally true that, the more scrambled up an incoming train is, the more work an engine will have to do, and the more time it will take to do it.* In a hump yard however, the incoming train is pushed steadily up and over the hump by a switch engine at the back. As the cars run down the slope the other side, they separate out, provided they have been disconnected. They can then be directed to their appropriate tracks by manipulation of switches. The operation is a much more continuous one, and the work of sorting depends very little on how scrambled up the train is. However the work at a hump yard can be increased by increasing the number of categories recognized in the sorting, for this will involve more classification tracks, and more work in assembling the new trains off the increased numbers of tracks.

2. Partly because of their different characteristics, it may be useful for a hump yard to do some sorting for a flat yard further down the line. This transfer of work is known as preclassification. The situation we will consider is as follows:

Fig. I



Traffic for a flat yard F passes through a hump yard H. At the flat yard n different categories are recognized, represented by $i = 1, 2 \ldots n$. The hump yard H is to do some preclassifying for F, and will recognize m different categories of traffic going beyond F. The problem is to find how much work this saves at F, so that this work saved can be balanced against increased costs at H. We also have the problem of choosing the categories to be recognized at H in order to save as much work as possible at F.

* See note N1 at end for further discussion.

We take as a measure of work at F the average number of cuts per car for traffic coming in to F.* A cut consists of an uninterrupted sequence of cars of the same category. (It is the same as a run in probability theory. For example; for the sequence of cars 2 2 1 3 1 1 1 3 3 2 there are 6 cuts, so cuts per car = 0.6.)

3. Suppose at first that the sequence of cars coming over the hump at H, which are bound for F, is a multinomial sequence, with probabilities p_i. (Later we suggest a more general hypothesis, which allows for the fact that shipments are often made in more than car load lots).

For a multinomial sequence, probability that a car is the first car of a cut of i's is $p_i q_i$

$\therefore$ probability that a car is first car of any cut is $\sum_{i=1}^{n} p_i q_i$ i.e. mean number of cuts per car = $\sum_{i=1}^{n} p_i q_i$ and this is an index of work if H does no preclassifying.

If H preclassifies into m classes, suppose these classes are formed by a partition of the categories 1,2...n into m mutually exclusive subsets $\mu_1, \mu_2 \ldots \mu_m$. This partition defines the preclassification policy at H. The cars now arrive at F in a series of runs, each run consisting of cars belonging to a μ_j $(j = 1 \ldots m)$.

Ignoring end effects of such runs, mean number of cuts per car for cars of μ_j is

$$\sum_{i \in \mu_j} p_i' q_i'$$

where $p_i' = \dfrac{p_i}{\sum_{i \in \mu_j} p_i}$ $\quad q_i' = 1 - p_i'$

But a proportion $\sum_{i \in \mu_j} p_i$ cars belong to μ_j $\therefore$ mean number of cuts per

---

* See note N1

car for all cars is $\sum_{j=1}^{m} \quad \sum_{i\epsilon\mu_j} p_i \sum_{i\epsilon\mu_j} p_i' \, q_i'$

$$= \sum_{j=1}^{m} \quad \sum_{i\epsilon\mu_j} p_i \;\; q_i'$$

$$(3.1) \qquad = 1 - \sum_{j=1}^{m} \frac{\sum_{i\epsilon\mu_j} p_i^2}{\sum_{i\epsilon\mu_j} p_i}$$

an index to the amount of work the given preclassification policy will have to H.

4. It is interesting to find what partition $\mu_1 \ldots \mu_m$ will leave the minimum work to h. The answer is provided by the following theorem.

Suppose $0 < p_1 \leq p_2 \leq \ldots \leq p_n$

Then for any partition $\mu_1 \ldots \mu_m$ of the numbers $(1, 2 \ldots n)$

$$(4.1) \qquad \sum_{j=1}^{m} \frac{\sum_{i\epsilon\mu_j} p_i^2}{\sum_{i\epsilon\mu_j} p_i} \leq p_n + p_{n-1} + \ldots + p_{n-m+1} + \frac{\sum_{i=1}^{n-m} p_i^2}{\sum_{i=1}^{n-m} p_i}$$

i.e. the partition which maximizes the left hand sum is $(1,2,\ldots,n-m)$ $(n-m+1) \ldots (n-1)$ (n).

First we will prove the result for $m = 2$

Suppose $n \in \mu_2$ $\quad \mu_2'$ is $\mu_2 - (n)$

Let $\sum_{i\epsilon\mu_1} p_i = \alpha \qquad \sum_{i\epsilon\mu_2'} pi = \beta$

$$\frac{\sum_{i\epsilon\mu_1} p_i^2}{\sum_{i\epsilon\mu_1} p_i} = a \qquad \frac{\sum_{i\epsilon\mu_2'} p_i^2}{\sum_{i\epsilon\mu_2'} p_i} = b.$$

(4.2) Note that $a \leq \alpha$, $b \leq \beta$ and $a, b \leq p_n$.

We have to prove that

(4.3)
$$a + \frac{p_n^2 + b\beta}{p_n + \beta} \leq \frac{a\alpha + b\beta}{\alpha + \beta} + p_n$$

If $a \leq b \leq p_n$ the result follows by comparing individual terms in 4.3. Suppose $b < a \leq p_n$

We have to prove that

$$p_n - \frac{p_n^2 + b\beta}{p_n + \beta} \geq a - \frac{a\alpha + b\beta}{\alpha + \beta}$$

i.e. that $\dfrac{p_n - b}{a - b} \geq \dfrac{p_n + \beta}{\alpha + \beta}$

(4.4) i.e. that $p_n b + p_n \beta + p_n \alpha \geq b\alpha + a\beta + p_n a$

If $\alpha \leq p_n$ the result follows by comparing individual terms, using 4.2

If $\alpha \geq p_n$ the result follows by rewriting (4.4) as

$$(\beta + p_n)(p_n - a) + (\alpha - p_n)(p_n - b) \geq 0$$

proving (4.1), for the case $m = 2$.

Now consider any partition $\mu_1, \mu_2 \ldots \mu_m$. Suppose $n_1 \in \mu_1$

The expression on the left hand side of (4.1) will not be diminished if we replace

$\mu_1, \ldots, \mu_m$ by $\mu_1', \mu_2', \ldots, \mu_m'$ where

$$\mu_1' = (n), \quad \mu_2' = \mu_1 + \mu_2 - (n), \quad \mu_3' = \mu_3, \ldots\ldots, \mu_m' = \mu_m$$

Suppose now $n - 1 \in \mu_i'$ $i \neq 1$

Replace μ_i', μ_j' $(i \neq j,\ i \neq 1,\ j \neq 1)$ by $(n - 1)$, $\mu_i' + \mu_j' - (n - 1)$. Again the sum is not diminished. The inequality 4.1 follows after m such steps. Thus it is best under these conditions to choose cut the $(m - 1)$ classes with highest probability, and preclassify them separately. Note though that if

$p_1 = p_2 = \ldots\ldots = p_n = \frac{1}{n}$, then it is quite immaterial what preclassification policy we use.

5. Extension to a more general type of sequence.

Although the multinomial sequence can often give quite a good representation of the incoming car sequence, it is rather restrictive. This can be seen especially if loads are shipped in more than carload lots. Then a run of cars is kept together throughout a trip, and in fact treated as a single car for sorting purposes. Thus mean cut length will tend to be longer than that expected from the multinomial distribution. However, the results of previous sections apply to a more general incoming sequence which will include this case.

This generalization is achieved by thinking of the incoming sequence of cuts rather than individual cars. Let A_i be a cut of incoming cars of type i. By the definition of a cut, no two A_i's of the same type can be together in the sequence. Suppose the sequence of A_i's is generated by a Marhoff chain with constant transition probabilities

$$(5.1) \qquad p_{ij} = \frac{p_j}{q_i} \qquad (i \neq j$$

$$= 0 \qquad (i \neq j).$$

and suppose the number of cars within the A_i is distributed in any manner whatever. Suppose now the A_i are preclassified, as in the previous sections by partitioning the i into m groups $\mu_1 \ldots \mu_m$. We can also suppose the A's keep their identity after sorting. It is thus possible to talk of cuts per A.

Let mean no. of cars per A $= x$

mean cuts per A after sorting $= y$

Then mean cuts per car after sorting $= \frac{y}{x}$

We can find y as follows:

If the original sequence is multinomial, then the cuts before sorting will be generated by a Marhoff chain of type (5.1). For a multinomial sequence, mean cuts per car after sorting =

$$1 - \sum_{j=1}^{m} \frac{\sum_{i \in \mu_j} p_i^2}{\sum_{i \in \mu_j} p_i} \qquad \text{from (3.1)}$$

For the multinomial sequence, mean no. of cars per A $= \dfrac{1}{\sum_{i=1}^{n} p_i q_i}$

$$\text{so} \quad y = \sum_{i=1}^{n} p_i q_i \qquad 1 - \sum_{j=1}^{m} \frac{\sum_{i \in \mu_j} p_i^2}{\sum_{i \in \mu_j} p_i}$$

and the same result must hold for the general sequence. So, for the general sequence mean cuts per car after sorting

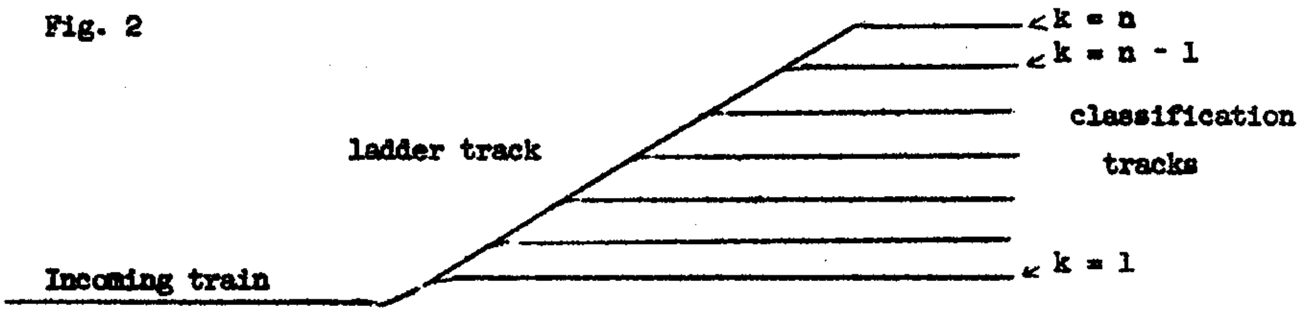
$$= \frac{\sum_{i=1}^{n} p_i q_i}{x} \left\{ 1 - \sum_{j=1}^{m} \frac{\sum_{i \in \mu_j} p_i^2}{\sum_{i \in \mu_j} p_i} \right\}$$

This provides the work function as before, and, of course, it is minimized for the same partition as before.

Notes

N 1 It may be useful to analyze the work function at a flat yard in a little more detail. Usually the yard is laid out rather like this:

Fig. 2



The diagonal track is called the ladder track.

The sorting operation is as follows. The switch engine pushes the train from behind along the ladder track. If necessary the train is backed up till the front cut is well behind its appropriate classification track. Then the train is pushed forward. Before the front cut reaches its appropriate classification track, it is disconnected from the rest of the train. Then the train is braked. The front cut runs forward ahead of the rest of the train, and the gap is sufficient to allow the switches to be set to divert it into its appropriate classification track, and then be reset to allow the train to proceed up the ladder track.

If successive cuts are to go into tracks higher and higher up the ladder track, a skillful crew can keep the train moving forward nearly all the time. To consider the work function in more detail, it seems useful to distinguish between cuts which only require the train to brake, and those which require it to back up.

S Suppose the sequence of cars is multinomial with probabilities $\pi_1 \ldots \pi_n$. The cars are now supposed to be numbered according to the position of their appropriate classification track on the ladder, as shown in Fig. 2.

We will only consider the number of reverses necessary, not the extent. Mean number of reverses per car

$$= \sum_{i=1}^{n} \sum_{j<i} p_i p_j$$

$$= \frac{1}{2} \sum p_i q_i$$

so that for just half the total number of cuts is backing up required, and the way in which classifications are assigned to tracks is immaterial. Thus whatever the relative costs of backing up or braking, the cost function is proportional to $\sum p_i q_i$.

This result generalizes immediately to the sequences discussed in section 5.

N 2 Note that the relative frequency of A_i in the A sequence defined by the Marhoff chain in (5.1) is $\lambda_i = \dfrac{p_i q_i}{\sum_{i=1}^{n} p_i q_i}$. It is not p_i.

(N 2.1)

This can be proved either by considering the multinomial distribution from which it is derived, or else by the following direct method.

If $n = 2$, the process is clearly cyclic, alternating between A_1 and A_2. Hence $\lambda_1 = \lambda_2 = \frac{1}{2}$.

If $n \geq 3$, the process is non-cyclic, and irreducible, so it is ergodic. The stationary state probabilities are the unique positive solution of the equations

$$\lambda_j = q_j \sum_{i \neq j} \lambda_i \frac{p_i}{q_i}$$

$$\sum_{i=1}^{n} \lambda_i = 1$$

These equations are satisfied by the expressions in N 2.1, so they are the stationary state solutions. By well known ergodic theorems, the relative frequency along a long series is equal to the stationary state solution.

N 3. If we have data on the relative frequencies λ_i of cuts generated by 5.1, a consistent method for estimating the p_i is to solve the equations

$$\text{N 3.1} \qquad \lambda_i = \frac{p_i (1 - p_i)}{\Sigma\, p_i (1 - p_i)} \qquad \sum_{i=1}^{n} p_i = 1$$

for the p_i.

In this note we show when an admissible solution exists, and suggest a numerical method for obtaining it:

Given a set of $\lambda_i > 0$, $(i = 1,2....n)$, such that $\Sigma \lambda_i = 1$, the N and S condition for the equations N 3.1 to have a solution such that $0 < p_i < 1$ for all i is that $\lambda_i < 1/2$ $(i = 1,2....n)$. Also if such a solution exists it is unique.

Proof: Suppose that $\lambda_1 \leq \lambda_2 \leq \ldots\leq \lambda_n$. Then if $(p_1, p_2 \ldots p_n)$ is a solution of N 3.1, then $p_1 \leq p_2 \leq \ldots \leq p_n$ and $p_i < 1/2$ for $i = 1,2\ldots,n - 1$.

To find a solution of N 3.1 with these properties consider the equations

$$\text{N 3.2} \qquad p_i (1 - p_i) = \frac{\lambda_i}{\lambda_n}\, p_n (1 - p_n)$$

For any p_n, $0 \leq p_n \leq 1$ these equations give a unique solution for p_i in the range $0 \leq p_i < 1/2$. Consider this solution for p_i as a function of p_n

$$p_i = p_i (p_n)$$

and consider the function

$$s\,(p_n) = \sum_{i=1}^{n-1} p_i\,(p_n) + p_n$$

a single valued function of p_n $0 \le p_n \le 1$

Then each solution of N 3.1 of the required type corresponds to a solution of $s\,(p_n) = 1$ $0 < p_n < 1$.

$s\,(p_n)$ is a twice differentiable function for $0 \le p_n \le 1$, and $s\,(0) = 0$, $s\,(1) = 1$. From N 3.2

N 3.3 $$(1 - 2p_i)\,\frac{dp_i}{dp_n} = \frac{\lambda_i}{\lambda_n}\,(1 - 2p_n)$$

N 3.4 $$(1 - 2p_i)\,\frac{d^2p_i}{dp_n^{\,2}} = 2\left(\frac{dp_i}{dp_n}\right)^2 - 2\,\frac{\lambda_i}{\lambda_n}$$

From N 3.3, N 3.4, $\dfrac{d^2p_i}{dp_n^{\,2}} < 0$ for $0 \le p_n \le 1$.

so $\dfrac{d^2s}{dp_n^{\,2}} < 0$ for $0 \le p_n \le 1$

and there can be at most one solution for $s\,(p_n) = 1$ in the range $0 < p_n < 1$.

Since $s\,(1) = 1$, there will be a solution if and only if $\left(\dfrac{ds}{dp_n}\right)_{p_n = 1} < 0$,

From N 3.3

$$\left(\frac{dp_i}{dp_n}\right)_{p_n = 1} = \frac{-\lambda_i}{\lambda_n}$$

so $$\left(\frac{ds}{dp_n}\right)_{p_n = 1} = 1 - \frac{1}{\lambda_n}\sum_{i=1}^{n-1} \lambda_i$$

$$\text{Thus} \quad \left(\frac{ds}{dp_n}\right)_{p_n = 1} < 0 \text{ if and only if } \sum_{i=1}^{n-1} \lambda_i > \lambda_n$$

i.e. if and only if $\lambda_n < 1/2$, since $\sum_{i=1}^{n} \lambda_i = 1$

Thus if $\lambda_1, \lambda_2, \ldots \lambda_n < 1/2$ the equations N 3.1 will have a unique solution, with $0 < p_i < 1$. Otherwise they will have no such solution. If $n = 2$, any p_1, p_2 with $p_1 + p_2 = 1$ $0 < p_1 < 1$, $0 < p_2 < 1$ will satisfy the equations.

The simplest way of solving the equations numerically seems to be to solve $s(p_n) = 1$ by an iterative method.

The statistical properties of this and the various other possible estimates of the p_i have still to be investigated.

Acknowledgements:

Discussion with Mr. M. Beckmann and Mr. C. B. McGuire has been extremely helpful.