MATHEMATICS MATTERS

Train Delays: Staying on Track

Catching a train is part of daily life for many, whether they are commuting to work or heading for a night out. Mathematics could help negotiate the minefield of delays that often accompanies train travel, reducing the risk of arriving late, and so reducing the stress of travel.

Every year more than a billion journeys take place on Britain's railways. Together, UK train passengers annually travel a combined distance of over 50 billion kilometres – enough to take you on five return trips to Pluto. However, 1 in 10 trains do not arrive on time.

According to a 2008 estimate by the National Audit Office, 2006/7 saw passengers endure a total of 14 million minutes of delays, costing them around £1 billion in lost time. Train companies publish data on these delays; however, this is an overall percentage of trains on time for each train company and doesn't tell you anything about delays on specific routes or at particular times of day. In fact, the quickest route on paper might not always be the best route in practice - some routes may have delays more often or be more susceptible to delays at a particular times. Keith Briggs, a mathematician at BT Research, near Ipswich, is working on a system to give passengers the best chance of arriving on time.

His aim is to develop a website which allows passengers to enter the maximum permitted delay to their journey and how important it is to stay within this maximum. For example, they could say they wanted to be 95% sure of making it within five minutes of their intended arrival time. You can imagine that the importance of a journey varies, depending on whether you are travelling for a deal-clinching business meeting or ambling to the Cotswolds for a long weekend. This could also be represented as a traffic light system of red, amber and green allowing the passenger to select the importance of their journey.

> "2006/7 saw passengers endure a total of 14 million minutes of delays, costing them around £1 billion in lost time"

To do this, Briggs created software that records real-time delay information from publicly available websites. He has collected details of over two million individual journeys over several years. From this data a "delay profile" for each route was constructed. A "route" here means a pair of stations which have direct train connections. Each route's delay profile is a mathematical equation which gives the chances of being delayed by a train, starting from one of the end stations, by a certain amount of time. It could show, for example, that one in ten trains are delayed by 5 minutes, one in a thousand by 10 minutes and one in 10,000 by more than 45 minutes. Combining this probability distribution with timetable information, Briggs constructed a mathematical function - called a 'kernel' - for each station.

The mathematics Briggs uses is very similar to that used by physicists when describing the diffraction of waves around obstacles. Imagine a rock in a pond. As the waves of a ripple encounter the rock the wave spreads out. Similarly, each change of train by a passenger at a station is like an obstacle which, because of possible delays, spreads out the arrival time of the passenger at the next station on the route. For every leg of the journey the kernel for each station is applied in succession, giving the distribution of arrival time at the final destination.

Briggs's method selects up to five possible routes and calculates these final distributions for each, which are then compared to

Institute of mathematics & its applications
National HE STEM Programme

the desired outcome originally set by the passenger. Take a journey from Norwich to Manchester. There are several ways to make this trip, all with timetabled travel times within around ten minutes of each other. For example, you could head down to London and catch the fast train up to Manchester, or hop to Peterborough, then on to Leeds, before changing for a Manchester-bound service.

Delays along the way could mean you missing a connection, having to wait for the next one and subsequently arriving at your destination much later than you'd hoped. It could be that at the time you intend to travel, going via London may be less susceptible to delays - and therefore the most reliable route - despite being timetabled to take around ten minutes longer.

It is all about how much of a risk you want to take. If your journey isn't very important then you'd take your chances with the quickest route on paper, hoping you won't encounter delays. However, if the journey is for a job interview, say, then the risk isn't worth taking and a website backed by Briggs's equations could advise you to travel the more reliable route via the capital, because despite taking longer you'd have a better chance of arriving on time. The website could also suggest you leave earlier, and tell you exactly how much sooner to set off.

Train delays are not only a massive inconvenience, but set paying passengers back a considerable amount of time and money every year. With train travel back on the increase after the recession, and some predicting a doubling of passenger numbers by 2020, it is mathematics that is finding a way to help you reach your destination on time.

TECHNICAL SUPPLEMENT

Delay profile

For each station the data on delays from the publicly available websites were plotted on a graph of the size of the delay (up to a maximum of an hour) against the fraction of trains delayed by that amount. The mathematical function that best fitted the data was a continuous q-exponential law. Due to the fact that train timetables are not continuous - they only give departure times to the nearest minute - a discrete version of the q-exponential law was used when combined into each station's kernel.

Kernels

Each station has a 60 x 60 matrix for a particular time of day. It is 60 on one side because the maximum delay Briggs considers is an hour. The other side is 60 because that hour is divided up into discrete one minute intervals - the nearest value provided by the train timetables.

The matrix is populated with the probability that if you arrive at the station at minute i, that you depart at minute j. This is based on timetable information and the delay profile information obtained from the website data grab.
The matrices for each station are in turn applied to a column vector. The column vector contains the probability distribution of your arrival time at the next station with each value showing the probability of being 0, 1, 2, 3 minutes late etc. The total column vector sums to one. Before you depart, the first value in the column vector is 1 and the rest are zeros - a delta function. This is because you haven't had chance to be subjected to delays yet.

By applying your starting station's matrix to this column vector, a new one is generated containing the probability distribution of your arrival time at the next station.

The matrix for that station is then applied to the new column vector, and so on until you reach your destination. The final, resultant column vector provides the distribution of your probable arrival times. This can then be compared with the final column vector for other routes and the optimum route selected.

References

Briggs, K. M. & Beck, C. (2007) Modelling train delays with q-exponential functions. *Physica A*. 378, 498-504.

Briggs, K.M. & Kim Po Tan, P. (2010) Optimal trip planning in timetabled transport systems possessing random delays. Submitted to *Transportation Science*.

The IMA would like to thank Dr. Keith Briggs, of BT Research, for his help in the preparation of this document.