Statistical modeling for bus network performance assessment

Sylvain Lassarre
GRETTIA/COSYS/IFSTTAR
• Spatio-temporal processes for buses
  – Deterministic
  – Stochastic
    • Travel time and dwell time
    • Daganzo’s model
• Trajectories and missing data
• Statistical models for processes
• Spatio-temporal processes for passengers
• Conclusion
• Most of the reliability indicators are derived from statistics on bus movements data

• For exemple
  – Coefficient of variation
  – Percentile

• without knowing the nature of the distribution of the stochastic variables such as travel time, headway, speed, … and the stochastic process which governs the dynamics
Spatio-temporal processes for buses

• The travelling of the bus along the route from origin to destination (#start and end terminal) as a moving particle \((x,t)\) which makes a pause at bus stops along the route

• Deterministic process = scheduled travelling = time-table
\[ t_{n,s+1} = t_{0,0} + nH + \sum_{0}^{s} p_{i} \]

- \( t_{n,s} \) time arrivals of bus \( n \) at bus stops \( s \) (or control points)
- \( H \) service headway
- \( t_{0,0} \) departure at origin
- \( p_{i} \) target travel time between \( i \) and \( i+1 \)

- Deterministic and stationary process

\[ t_{n,s+1} = t_{n,s} + c_{s} \]

- \( p_{s} = c_{s} \) average (global) travel time from \( s \) to \( s+1 \), sum of
  - \( \pi_{s} \) Inter-stop travel time between bus stops \( s \) and \( s+1 \)
  - \( d_{s} \) Dwell time or delay due to bus stop \( s \)
S stochastic processes and models (1)

• Of the arrival times $a_{n,s}$ because of random disturbances (traffic, passenger demand, driver behavior, ...) affecting $t_{n,s}$

• By means of (Hans et al., 2014)
  • $u_{n,s} = a_{n,s+1} - a_{n,s}$ global travel time for segment $s, s+1$
    – dwell time models (boarding and (alighting) passengers)
    – travel time models

\[
\begin{align*}
  u_{n,s+1} &= a_{n,s+1} - a_{n,s} = d_{n,s} + \pi_{n,s} \\
  a_{n,s+1} &= a_{n,s} + d_{n,s} + \pi_{n,s}
\end{align*}
\]
Dwell time models

- $L_{n,s}$ load of the bus
- $A_{n,s}$ number of alighting passengers off the bus
- $B_{n,s}$ number of boarding passengers in the bus

\[
L_{n,s+1} = L_{n,s} - A_{n,s} + B_{n,s} \quad \text{Infinite capacity}
\]

\[
d_{n,s} = \max(aA_{n,s}, bB_{n,s}) + c \quad \text{Times to alignto boardopening door}
\]

\[
d_{n,s} = aA_{n,s} + bB_{n,s}
\]

\[
A_{n,s} = d_s L_{n,s-1} \quad \text{Alighting proportion}
\]

\[
B_{n,s} = (\lambda_s - \frac{1}{\tau_s}) h_{n,s} \quad \text{Arrival rate of passengers}
\]
Travel time and speed models

• Gamma distribution
  skewed to the right, is equal to the exponential distribution when the scale is equal to 1 and tends towards the normal distribution when the scale gets infinite. The variance is proportional to the square of the mean $E(Y) = \nu \lambda$.

• an inverse-gamma distribution
  \( \Gamma(\nu_{s+1}, 1/L_{s+1} \lambda_{s+1}) \)
• Normal-exponential distribution
  – Compound of a normal distribution by an exponential distribution

Portland data At each stop door opening and closing times are recorded, and the APC records the number of boarding and alighting passengers
<table>
<thead>
<tr>
<th>Model</th>
<th>Description</th>
<th>Data requirement</th>
<th>Stochastic form</th>
<th>Deterministic form</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dwell time</td>
<td>1 Use distributions for dwell times</td>
<td>Archived bus dwell times from AVL</td>
<td>( d_{i,s} \sim \mathcal{E}(c + \beta_s h_{i,s}) )</td>
<td>( d_{i,s} = c + \beta_s h_{i,s} )</td>
</tr>
<tr>
<td></td>
<td>2 Use distributions for numbers of passengers</td>
<td>Archived bus dwell times from AVL</td>
<td>( B_{i,s} \sim \mathcal{P}(\lambda_s h_{i,s}) )</td>
<td>( B_{i,s} = \lambda_s h_{i,s} )</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Automatic passenger counters (APC)</td>
<td>( A_{i,s} \sim \mathcal{Bi}(L_{i,s}, \eta_s) )</td>
<td>( A_{i,s} = L_{i,s} \eta_s )</td>
</tr>
<tr>
<td>Travel time</td>
<td>1 Use travel time distributions</td>
<td>Archived bus travel times from AVL</td>
<td>(a) ( \pi_{i,s} \sim \mathcal{N}(\mu_s, \sigma_s) )</td>
<td>(a) ( \pi_{i,s} = \mu_s + \lambda_s )</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(b) ( \pi_{i,s} \sim \mathcal{L}(\mu_s, \sigma_s) )</td>
<td>(b) ( \pi_{i,s} = \mu_s + \frac{\lambda_s}{\lambda_s + \epsilon_s} )</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(c) ( \pi_{i,s} \sim \mathcal{G}(\mu_s, \sigma_s) )</td>
<td>(c) ( \pi_{i,s} = \mathcal{N}(\lambda_s, \sigma_s) )</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>Archived bus travel times from AVL</td>
<td>(a) Without a signal: ( \pi_{i,s} \sim \mathcal{NE}(\mu_s, \sigma_s, \lambda_s) )</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(b) Otherwise: ( \pi_{i,s} \sim \mathcal{NS}(\mu_s, \sigma_s, r_s, g) )</td>
<td>(b) ( \pi_{i,s} = \mu_s + \frac{\lambda_s}{\lambda_s + \epsilon_s} )</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3 Travel time based on synthetic and realistic trajectories</td>
<td>Signal location and settings</td>
<td>( \pi_{i,s} \sim \mathcal{N}\left(\frac{\lambda_s}{\theta_s}, \sigma_s\right) )</td>
<td>( \pi_{i,s} = \frac{\lambda_s}{\theta_s} )</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>Signal location and settings</td>
<td>( \pi_{i,s} \sim \mathcal{N}\left(\frac{\lambda_s}{\theta_s}, \sigma_s\right) )</td>
<td>( \pi_{i,s} = \frac{\lambda_s}{\theta_s} )</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Flow from loop detectors</td>
<td>( N_{i,s} \sim \mathcal{P}(q_s \Delta t) )</td>
<td>( N_{i,s} = q_s \Delta t )</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Velocity characteristic of buses</td>
<td>( N_{i,s} \sim \mathcal{P}(q_s \Delta t) )</td>
<td>( N_{i,s} = q_s \Delta t )</td>
</tr>
</tbody>
</table>
Particle filter as prediction method

- Generate K trajectories (particles) based on models

\[
t_{i,s_2|s_1}^{(k)} = \bar{t}_{i,s} + \sum_{s=s_1}^{s_2} \left( d_{i,s}^{(k)} + \pi_{i,s}^{(k)} \right)
\]

\[
t_{i,s_2|s_1} = \text{median}_{k \in [1:X]} \left( t_{i,s_2|s_1}^{(k)} \right)
\]
Headways and dwell times

- Headway distribution could be bimodal

- Typical dwell time distributions: Burr, log-logistic, log-normal
Stochastic processes and models (2)

• By means of (Daganzo, 2009)
  • \( u_{n,s} = a_{n,s+1} - a_{n,s} \) global travel time for segment \( s, s+1 \)
  • \( h_{n,s} = a_{n,s} - a_{n-1,s} \) inter-arrival headway between buses \( n-1 \) and \( n \) at stop \( s \)

\[
\begin{align*}
    u_{n,s} &= c_s + \beta_s (h_{n,s} - H) + \nu_{n,s+1} \\
    a_{n,s+1} &= a_{n,s} + c_s = a_{n,s} + \beta_s (a_{n,s} - a_{n-1,s} - H) + \nu_{n,s+1}
\end{align*}
\]

• With \( \beta_s \) marginal increase in expected bus delay from a unit increase in headway
• \( \nu_{n,s+1} \) Random noise detected at \( s+1 \)
• Origin of bus bunching (instability of the dynamical system)
\[ u_{n,s} = c_s + \beta_s (h_{n,s} - H) + \nu_{n,s+1} \]

\[ a_{n,s+1} = a_{n,s} + c_s = a_{n,s} + \beta_s (a_{n,s} - a_{n-1,s} - H) + \nu_{n,s+1} \]

\[ h_{n,s+1} = h_{n,s} - u_{n-1,s+1} + u_{n,s+1} \]

\[ h_{n,s+1} - h_{n,s} = u_{n,s+1} - u_{n-1,s+1} \]
Observed trajectories

Time sampling
Frequency 10s.

Departure and arrival times to be estimated on observed data

Dwell times missing on scheduled trajectory

Ref: Data as resource, Nair et al, 2014
Missing data

• Checking data to face the following anomalies
  – **Bus overtaking** (BO), succeeding bus passes the predecessor
  – **Technical failures** (TF) in the recording of GPS data. Results in missing data and temporal technical gaps
  – **Incorrect Operations in the Service** (IOS) consequences of unpredicted missed trips and unexpected breakdowns. Results in missing data and temporal technical gaps
Estimation of arrival times from the bus trajectories (Lakhotia et al., 2017)

• Bus stop as the unit of reference. The bus is defined to be at a bus stop if the distance between the stop and the bus is less than a certain value, which is called buffer radius (30m. radius).

• Bus as the unit of reference with route matching
• Select a timestamp only if the distance to nearest station is less than the buffer radius.
  – Since the GPS data was at a frequency of 10 seconds, it is possible that we have multiple timestamps for a particular stop. In this case, the first time stamp is designated as arrival time and the last is departure time. It is also to be noted that it is not uncommon for drivers to stop at bus stops for less than 10 seconds in case there is no demand, in which case we have no timestamp for arrival (Missing values).

• Order of arrival of buses at intermediate stations may not be the same as overtaking is possible
  – Complicate the calculation of headways
Estimation of statistical models

• The model is a system of equations (n=1 to N bus) on the first differences between the arrival times on a sample of days (j = 1,l)

\[ \Delta^s Y_{nsj} = c_s + \beta_s (Y_{ns-1,j} - Y_{n-1s-1,j} - H) + \nu_{nsj} \]

\[ \Delta^s Y_{nsj} = c_s + \beta_s (\Delta^n Y_{ns-1,j} - H) + \nu_{nsj} \]

• which can be written

\[ U_{nsj} = c_s + \beta_s (h_{ns-1,j} - H) + \nu_{nsj} \]

• The first statistical approach is to use a gamma distribution for the global travel time with two parameters shape and scale \( G(\nu_s, \lambda_s) \) and a gamma regression knowing the departure times of the previous bus.
• Then we have to change the formulation of the model to adapt to the form of a generalized mixed linear model with a link function which could be a logarithm (or the inverse)

\[
\text{Log}E(U_{nsj}) = c_s + \beta_s (h_{ns-1j} - H)
\]

\[
E(U_{nsj}) = e^{c_s} e^{\beta_s (h_{ns-1j} - H)}
\]

• It is an accelerated failure model. Or

\[
\text{Log}E(U_{nsj}) = c_s + \beta_s \log(h_{ns-1j} - H)
\]

\[
E(U_{nsj}) = e^{c_s} (h_{ns-1j} - H)^{\beta_s}
\]

for a multiplicative model

• We can also use the canonical link which is the inverse function (so the travel speed)

\[
1/E(U_{nsj}) = c_s + \beta_s (h_{ns-1j} - H)
\]

\[
E(U_{nsj}) = 1/(c_s + \beta_s (h_{ns-1j} - H))
\]
Spatio-temporal processes for passengers

• Bulk queue models (serving more than one customer at a time) and Markov chain
  – Stock waiting at stop s at run n
    \[ S^n_s = S^{n-1}_s - B^{n-1}_s + \beta_s (a^n_s - a^{n-1}_s) \]
  – Hypothetical number of passengers boarding
  – Hypothetical departure time
    \[ \overline{B}^n_s = S^n_s + \beta_s (\overline{d}^n_s - a^n_s) \]
    \[ \overline{d}^n_s = a^n_s + ad_s L^n_s + b\overline{B}^n_s \]

L is the number of passengers in the bus

\[ \overline{d}^n_s = a^n_s + \frac{1}{1 - b \beta_s} (ad_s L^n_s + bS^n_s) \]
\[ L_{n,s+1} = L_{n,s} - dL_{n,s} + B_{n,s} \]

\[ \overline{B}^n_s = S^n_s + \frac{\beta_s}{1 - b \beta_s} (ad_s L^n_s + bS^n_s) \]

hypothesis: all passengers can board
• Then look at capacity
  – Bus can accommodate the hypothetical number

If residual capacity + alighting > waiting, then
Check that the departure time is not earlier than the scheduled departure (if not hold)

  – Bus cannot accommodate the hypothetical number then
  boarding = residual capacity and departure time = min (theoretical, feasible)
Figure 1: number of passengers waiting at a station, function of time

Fig. 3. Scheme of the change of bus filling with time.
Performance indicators

• Total passenger travel time

\[
\sum_{s=1}^{S-1} \left(a_s^n - d_{s-1}^n\right) L_s^n + \sum_{s=1}^{S-2} \left(L_{s+1}^n - B_s^n\right) \left(d_s^n - a_s^n\right)
\]

• Waiting time

• Delay

Figure 2: Area representing total waiting time at a station
User cost component between i and j

- $C(t)$ travel disutility when leaving at time $t$ with schedule delay cost or head start (time allocated to waiting)

- $t^*$ preferred arrival time, $T$ travel time

\[
C(t) = \alpha_w T_w + \alpha_v T_v + \beta(t^* - t - T_w)^+ \\
+ \gamma(t + T_w - t^*)^+
\]

- Excess waiting time at the origin stop i (probability to miss the targeted departure)

- Potential travel time called the travel time buffer $A_{j^{95}} - E(A_j)$

- Mean riding time $E(A_j) - E(D_i)$
• Excess waiting time: On(s) or boarding
• Potential travel time: OFF(s) or alighting
• Mean riding: OD(i,j) but in fact ON(i) and OFF(j)
Conclusion

• Well known stochastic processes for bus and passenger movements. Interesting non stable dynamics with bus bunching.

• Fit of Statistical distributions for travel and dwell times
  – Data problems to solve with GPS (missing data)
  – More than one candidate for travel times, dwell times, headways: non standard distributions (very skewed)

• Use of statistical models to fit Daganzo’s models

• Comprehensive transition to performance indicators
Bibliography


• Nair, R., Bouillet, E., Gkoufas, Y., Verscheure, O., Mourad, M., Yashar, F., Perez, R., Perez, J., Bryant, G. and Transit, M.D., 2014. Data as a resource: real-time predictive analytics for bus bunching.


• Combes F., Benezech V., 2014. Quality of service and congestion in urban public transport: a simulation approach. Conference AET.
