**Lineup Forecasting Model 1**

We have a snapshot of the “active schedule”; i.e., a snapshot of the trains that are scheduled to run as of the current time. We know that this schedule won’t be executed exactly, so we want to augment this with history to forecast the schedule that will actually run. Time is “bucketized” into ∆ hr intervals, and the granularity required on the forecasted schedule is the number of trains by departure-arrival bucket pair.

Observe that departures in the snapshot can differ from actual departures for several reasons:

1. A train in the snapshot is dropped from the schedule between now and when it is scheduled to depart (drop)
2. A train is added to the schedule between now and when it actually departs (add)
3. A train delays or advances in the schedule between now and when it actually departs (shift)

Note also that transit times can also be longer or shorter than what is shown in the snapshot.

**Assumptions**

* Transit time deviations are independent of departure time deviations
* Day of week is the only seasonal effect (no hour of day, week of month, etc. effects)

**Parameters**

*T* = set of time buckets

*W* = set of days of the week = {M, Tu, W, Th, F, Sa, Su}

*dij* = number of trains departing & arriving in buckets *i* and *j* in the snapshot; *i*,*j*∈*T*

*wi* = day of week of bucket *i* for this snapshot; *i*∈*I*

**Estimated Parameters**

Compare historical snapshots to reality to estimate the following parameters (using some tbd models):

*pii’*(*k*) = probability a train departing in bucket *i* on day of week *k* in the snapshot actually departs in bucket *i’*; *i*,*i’*∈*T*, *k*∈*W*

*qij*(*k*) = probability a train departing in bucket *i* on day of week *k* with a transit time of *j* bucketsin the snapshot never departs (is dropped); *i*,*i*+*j*∈*T*, *k*∈*W*

*rij*(*k*) = probability a train actually departing in bucket *i* on day of week *k* with a transit time of *j* bucketswasn’t in the snapshot (is added); *i*,*i*+*j*∈*T*, *k*∈*W*

*tijj’*(*k*) = probability a train departing in bucket *i* on day of week *k* with a transit time of *j* bucketsin the snapshot has an actual transit time of *j’* buckets; *i*,*i*+*j*,*i*+*j’*∈*T*, *k*∈*W*

**Random Variables**

*Xij* = number of trains that actually depart & arrive in buckets *i* & *j*; *i*,*j* ∈ *T*

*Sii’jj’* = number of trains departing & arriving in buckets *i* & *j* in the snapshot that actually depart & arrive in buckets *i’* & *j’*; *i*,*i’*,*j*,*j’*∈*T*

*Dij* = number of trains departing & arriving in buckets *i* & *j* in the snapshot that never depart (are dropped); *i*,*j*∈*T*

*Aij* = number of trains that actually depart & arrive in buckets *i* & *j* that weren’t in the active snapshot (are added); *i*,*j*∈*T*

**Model**

(1)

Equivalently,

(2)

Because transit time deviations are assumed to be independent of departure time deviations,

*E*[*Sii’jj’*] = *dijpii’*(*wi*)*ti,j-i,j’-i’*(*wi*) *i*,*i’*,*j*,*j’*∈*T* (3)

Because we expect *E*[*Aij*](1-*ri,j-i*(*wi*)) = *dij*,

*E*[*Aij*] = *dij*/(1-*ri,j-i*(*wi*)) *i*,*j*∈*T* (4)

Also,

*E*[*Dij*] = *dijqi,j-i*(*wi*) *i*,*j*∈*T* (5)

**Extensions**

* Incorporate train type
* Incorporate eastbound vs. westbound