Data-Driven Alerts in Airline Revenue Management

Nicola Rennie [1], Catherine Cleophas [1 2], Florian Dost [1], Jonas Rauch [3]

[1]Lancaster University, [2]Christian-Albrechts-Universität zu Kiel, [3]Deutsche Lufthansa AG

1. Introduction

- Revenue management maximises revenue through optimisation based on forecasted demand. **Unexpected demand** due to major sporting events, or carnivals, which deviates from forecasts needs to be identified.

- Aim to develop methods to highlight deviations between real-world booking data and revenue management system forecasts, and identify flights which will benefit from analyst intervention.

2. Simulated Passenger Arrivals

- Type 1 (Business) and Type 2 (Tourist) customers arrive according to two different piecewise homogeneous Poisson processes.

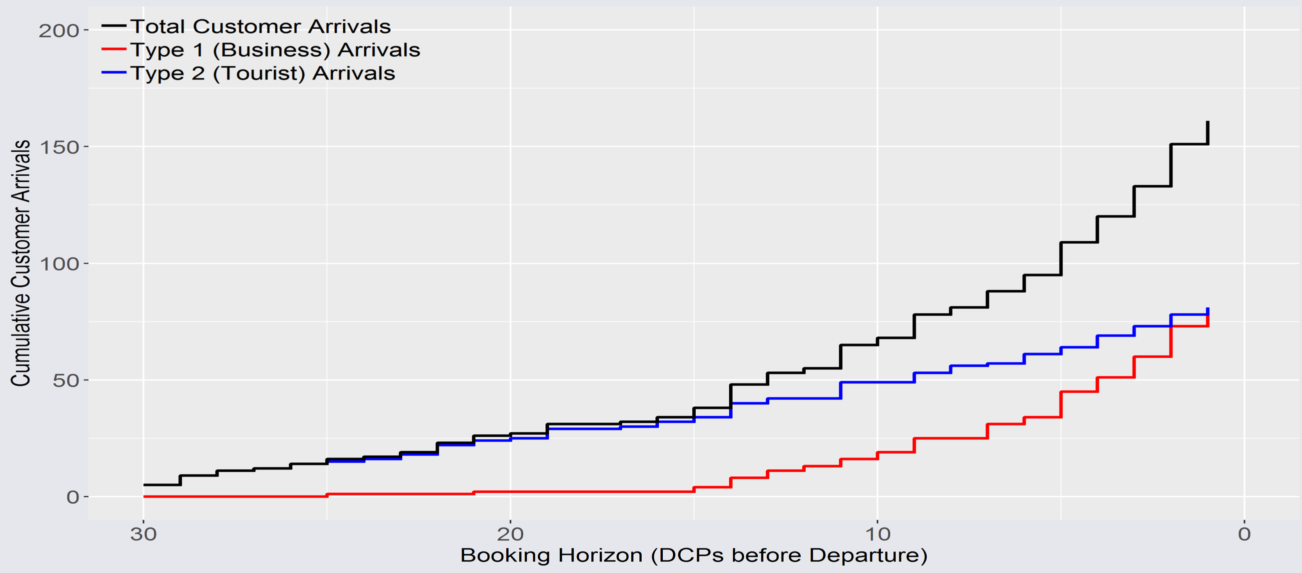


Figure 1: Customer Arrivals under Piecewise Homogeneous Poisson Processes

3. Simulated Passenger Bookings

- Customers prefer a fare class on offer (Y, M, or K) but buy-up with some probability. Fare classes close based on the **Seat Load Factor** (SLF).

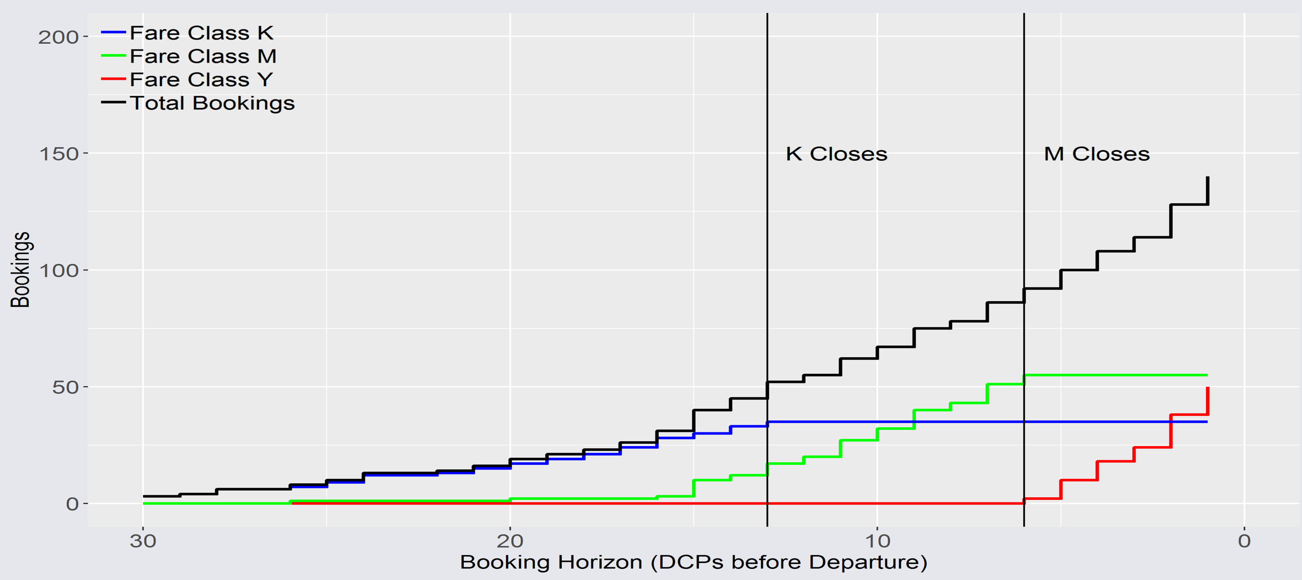
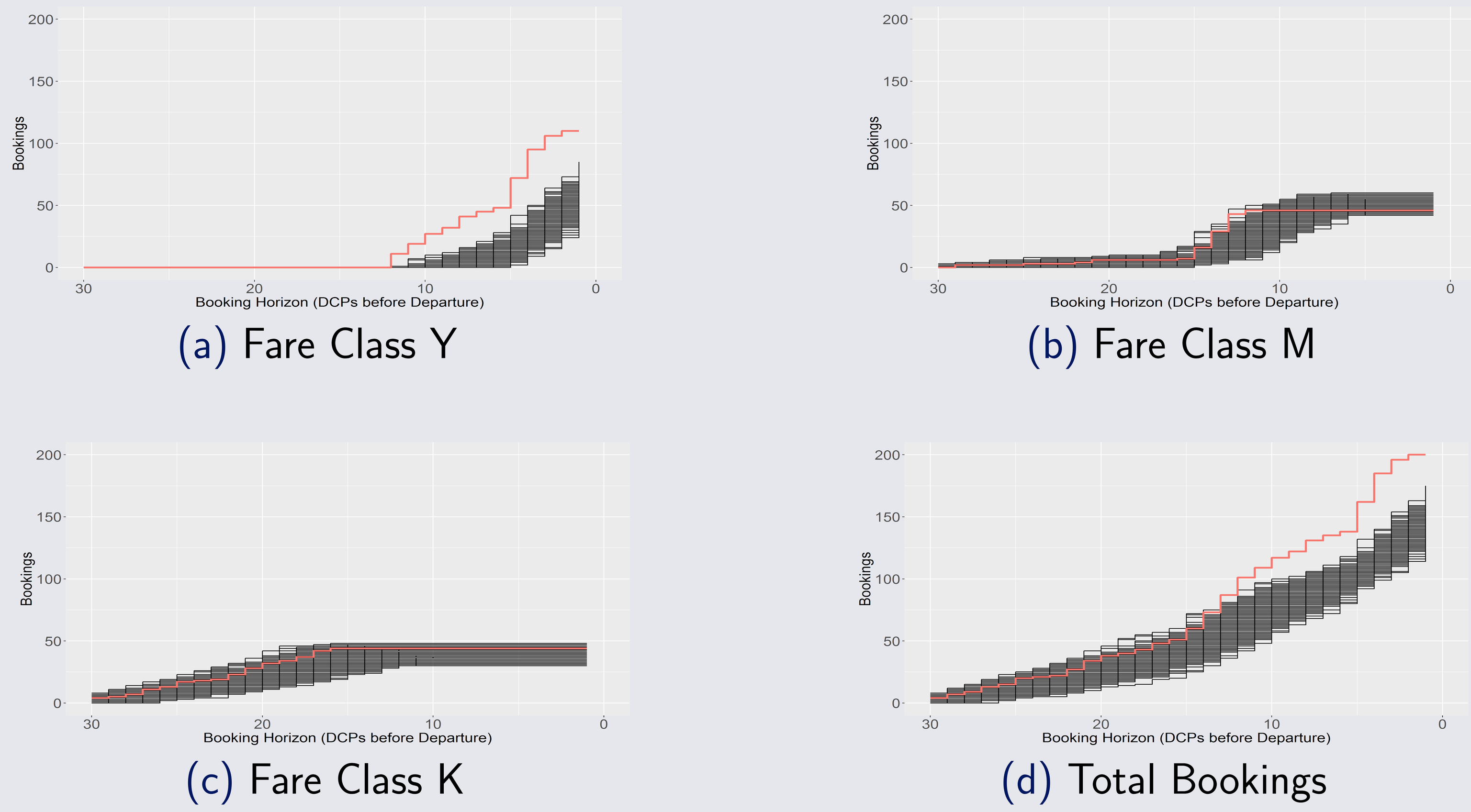


Figure 2: Booking Data Under Seat Load Factor Simulation

4. Outlier Generation

- **Outlier**: 'an observation which deviates so much from the other observations as to arouse suspicions that it was generated by a different mechanism' (Hawkins, 1980).



(a) Fare Class Y (b) Fare Class M

(c) Fare Class K (d) Total Bookings

Figure 3: Generation of a Single Outlier Caused by Type 1 (Business) Demand Increase

- Generate an outlier in the demand for a flight:
  - Establish *normal* behaviour through 500 runs of SLF simulation.
  - Introduce outliers by changing mean of customer arrival processes.
  - **Aim**: Adapt and apply outlier detection algorithms in order to highlight which flights' booking behaviour results from unexpected demand.

5. Initial Distance-based Outlier Detection Results

- Distance-based outlier detection approaches are unsupervised methods based on the idea that observations which are further away from others can be deemed outliers.
- For a time series $x_i = (x_{i1}, x_{i2}, \dots, x_{iN})$, if:

$$T_{x_i} = \sum_{k=1}^{K} \left(\sum_{n=1}^{N} (x_{in} - x_{kn})^2 \right)^{\frac{1}{2}} \geq \tau$$

then x_i can be defined as an outlier, for some pre-defined threshold τ, where K is the total number of time series being considered.

Fare Class	Avg. No. False Positives	% Runs Outlier Identified
Y	1.98	100
M	6.15	47
K	5	0
Total	3.88	99

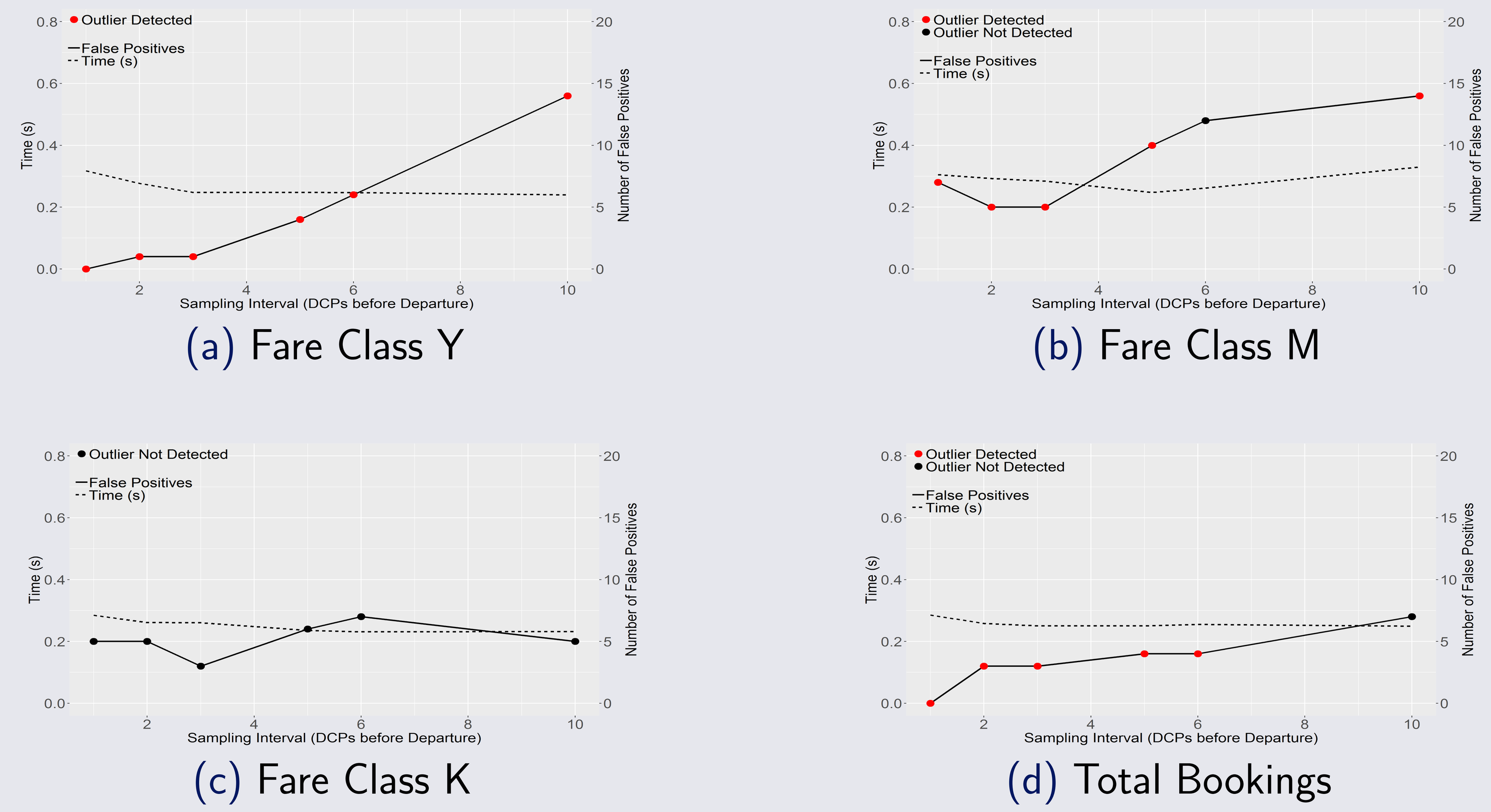
Table 1: Initial Results of Euclidean Distance Outlier Detection for 100 Runs

✓ Fairly low false positive rate.

✗ Variable detection rate and computationally intensive.

6. Computational Feasibility

- One of the main concerns about any outlier detection technique is its computational complexity. One potential solution is to reduce the sampling frequency, which raises the question of: how does the sampling frequency affect the computational time, false positive rate, and detection rate?



(a) Fare Class Y (b) Fare Class M

(c) Fare Class K (d) Total Bookings

Figure 4: Sampling Frequency Effect on Euclidean Distance Single Outlier Detection

✓ Slightly decreased computational time, as expected.

✗ Effects on detection and false positive rates are less clear.

7. Future Work

- Build a more complex simulation, including more fare classes and verify whether initial results still hold.
- Collect empirical data and label historical outliers.
- Consider semi-supervised learning approaches to outlier detection.

Contact Details

- Web: http://lancaster.ac.uk/~rennien/
- Email: n.rennie@lancaster.ac.uk