SMS Essay Competition 2017

Airline Overbooking with Mathematics and Big Data

Jovon Lim, Hu Xingyi, Xavier Goh

Victoria Junior College27/6/2017

Abstract

“Data will last longer than the systems themselves” (Lee, n.d.). Today, chunks of data collected over many years can be analysed. From data mining to machine learning, complemented by mathematics, big data is applied in many fields. The recent forced removal of a man by United Airlines has raised questions on the reliability of big data in allowing airlines to overbook flights to increase revenue. This essay engages in mathematics, using a binomial distribution model to investigate how overbooking raises revenue, and an evaluation of the feasibility of this model through the use of big data wraps up the essay.

Overbooking

As passengers fly, profits soar. In a bid to maximise revenue, airlines have been overbooking flights since the 1950.

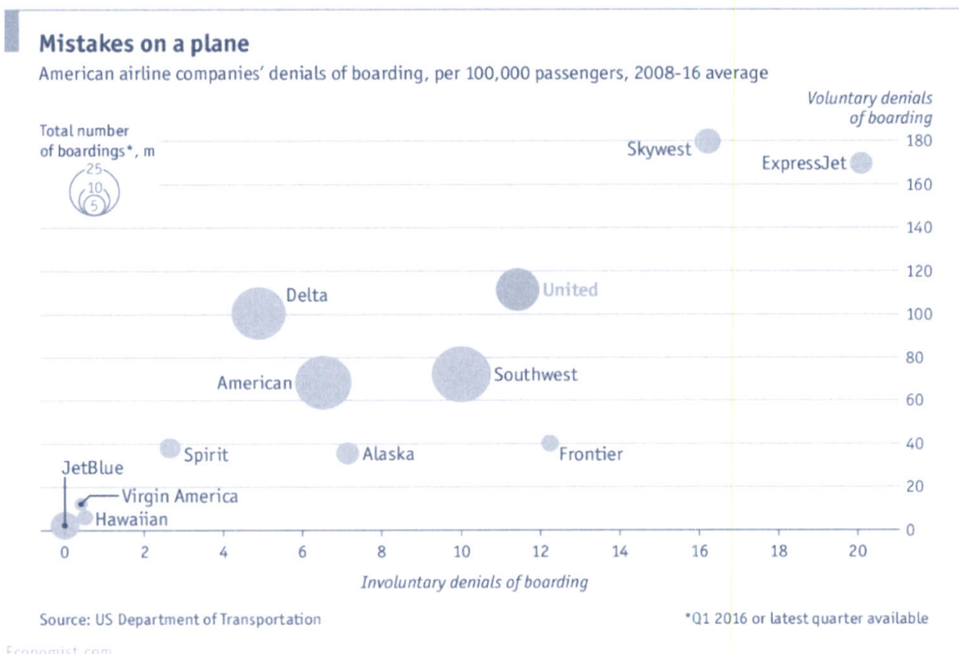

Figure 1: A graph illustrating the denials of boarding with respect to airline (Graph: TEAM, 2017)

Federal data reflects a broad pattern of aggressive overbooking (Figure 1). It was in this case of the United Airlines flight, where a paying customer was forcefully removed from the aircraft, that this issue of overbooking was brought to the spotlight. With widespread use of big data in the lucrative industry, it is pertinent to explore the rationale behind "overbooking" to then propose hypothetical distribution models on how airlines could use big data to overbook and thus maximize their profit. The feasibility of such models can then be evaluated in light of the airline industry.

Overbooking and why airlines do it

One could start by comparing the merits and demerits of overbooking. Assuming that an airline earns profit from a passenger only when the passenger takes the flight, the airline essentially loses potential profit from the ticket sale for every empty seat. Thus, airlines tend to sell more tickets than the plane's capacity, or overbook the flight, in an attempt to maximise profits (Simon, 2013).

There are, however, some risks involved. Such risks include more passengers showing up than anticipated. The airline has to then anger customers who must be bumped from the flight or look for volunteers who are willing to board the next flight. In both cases, vouchers are given as compensation or incentives.

Modelling with binomial distribution

Since the risks of overbooking can be simplified to one of whether passengers show up for their scheduled flight, we use a binomial distribution, in which 2 assumptions have to be made:

- the probability of a passenger showing up for the flight is constant
- the probability of a passenger showing up for the flight is independent of the probability that another passenger shows up for the flight

Consider an airline which opens up a one-way route on a Boeing 777 airplane from Singapore to Shanghai, with a seating capacity of 300 (50 Business Class seats and 250 Economy Class seats). During the busy travel season, the demand for the ticket exceeds the number of seats; hence, the airline attempts to overbook its flights.

Let X be the number of passengers out of 250 that turn up for their Economy Class flight. Since "7-8% of customers are no-shows" (Bailey, n.d.), the probability that a passenger turns up is 0.92. Hence $X \sim B(250,0.92)$.

Let Y be the number of passengers out of 50 that turn up for their Business Class flight. It is assumed that the probability that a passenger turns up is 0.98 due to its higher cost. Hence $Y \sim B(50,0.98)$

The two distributions are assumed to be independent. Figure 2 below provides a summary of the total revenue earned.

		Airline does not overbook its flight	Airline overbooks its flight
Economy	Number of tickets sold	250	250 + *p*
	Expected number of passengers turning up	E(X) = 250*0.92=230	E(X') = (250+*p)**0.92
	Price of 1 ticket / $	400	
	Expected revenue / $	400*230=92000	∑R(x)P(X=x) 97991 when p=21
Business	Number of tickets sold	50	50 + *q*
	Expected number of passengers turning up	E(Y) = 50*0.98=49	E(Y')=(50+*q*)*0.98
	Price of 1 ticket / $	1400	
	Expected revenue / $	49*1200=58800	58834 when q=1
	Total expected revenue / $	92000+58800 = 150800	97991+58834= 156825

Figure 2: A table indicating the expected revenue with and without overbooking

Now consider the case where the airline overbooks an additional *p* tickets for Economy and *q* tickets for Business, with the cost of bumping one passenger is at $800 for Economy and at $2000 for Business. The new binomial distributions are as follows:

$$X' \sim B\,(250+p,\ 0.92)$$
$$Y' \sim B\,(50+q,\ 0.98)$$

$$\mathrm{P}(\mathrm{X}' = \mathrm{x}) = \binom{250+\mathrm{p}}{\mathrm{x}}(0.92)^{\mathrm{x}}(0.08)^{250+\mathrm{p}-\mathrm{x}}$$

If $x \leq 250$, then revenue gained $= R(x) = 400x$

If $y > 250$, then revenue gained $= R(y) = 400\times250 - 800(x - 250)$

Therefore, the expected revenue gained from economy class=$\sum_{x}^{250+p} R(x)P(X' = x)$

By plotting the graph of total revenue against p, the airline can obtain the maximum revenue, as indicated by the turning point, and hence the optimal number of tickets to overbook.

An example would be the case when p=21 (Figure 2), where the expected revenue would be $\sum_{x}^{250+21} R(x)P(X' = x)$

$=\sum_{1}^{250}(400x)P(X = x) + \sum_{251}^{271}(400\times250 - 800(x - 250))P(X = x)$

$= 58422 + 39570 = 97991$

This value of 97991 is more than the expected revenue of 92000 when there is no overbooking.

Similarly, for business class:

$$\mathrm{P}(\mathrm{Y}' = \mathrm{y}) = \binom{50 + \mathrm{q}}{\mathrm{y}}(0.98)^{y}(0.02)^{50+\mathrm{q}-\mathrm{y}}$$

If $y \leq 50$, then revenue gained$= R(y) = 1200y$

If $y > 50$, then revenue gained$= R(y) = 1200y - 2000(y - 50)$

Therefore, the expected revenue gained from business class$=\sum_{y}^{50+q} R(y)P(Y' = y)$

By plotting the graph of total revenue against q, the airline can obtain the maximum revenue, as indicated by the turning point, and hence the optimal number of tickets to overbook.

An example would be the case when q=1, where the expected revenue would be $\sum_{y}^{50+1} R(y)P(Y' = y)$

$=\sum_{1}^{50}(1200x)P(X = x) + (1200\times50 - 2000\times1)P(X = 51)$

$= 38134 + 20699 = 58834$

This value of 58834 is more than the expected revenue of 58800 when there is no overbooking.

Therefore, it can be shown that the airline indeed gains more profit when they oversell tickets.

Feasibility of proposed models complemented with big data

The degree to which a binomial distribution model simulates actual passenger behaviour has to be consistently checked and reviewed in light of new data and changing trends, since there is virtually no mathematical model that completely models consumer behaviour in social sciences in relation to empirical evidence. Nevertheless, mathematical models used by the airlines, including the binomial model, are becoming more accurate with the increasingly widespread use of big data.

The first assumption of the binomial model that the probability of a passenger showing up for the flight is constant would be more tenable with the use of big data. Airlines are using technology to chunk data and organise information collected from flights over the years into useful patterns. For instance, United Airlines has the "collect and analyse" approach to data to study consumer behaviour, forming mathematical models to simulate consumer behaviour (Noyes, 2014).

In the case of overbooking, big data allows the law of large numbers to be applied to augment the degree of simulation of the binomial model in relation to empirical data. The tendency for consumers to show up for flights can be determined and analysed with data mining. Specific details from nationality to occupation are used. This allows the "showing up" behaviour to be generalised as a consumer-type. Passengers who are likely to miss flights will form another distinct category, with their behaviour predictable as a large group. With distinct consumer types based on behaviour generalized using big data, successive probabilities will then be approximately constant.

However, even as big data provides "timely" information on the most probable outcomes based on historical data, new trends of consumer behaviour develop and evolve. Thus, airlines need to keep up with new data and tease out relevant the correlation to allow them to adjust their sample size and probabilities. An advanced era where real-time data can be simultaneously analysed with a high level of accuracy to model the situations such as overbooking with a probability distribution will be ushered in as the development of machine learning and data mining technologies continue to unfold.

Bibliography

Bailey, J. (n.d.). Retrieved from The New York Times: http://www.nytimes.com/2007/05/30/business/30bump.html?pagewanted=print&_r=0

Lee, T. B. (n.d.). Retrieved from http://izquotes.com/quote/110047

Noyes, K. (2014). Retrieved from https://www.google.com/url?q=http://fortune.com/2014/06/19/big-data-airline-industry/&sa=D&ust=1499355037521000&usg=AFQjCNGtV9N65buGYec-J56J4mppUw1K3Q

Simon, C. (2013). Retrieved from http://corysimon.github.io/articles/by-how-many-flights-should-an-airline-overbook/

TEAM, T. D. (2017, April). Retrieved from The Economist: https://www.economist.com/blogs/graphicdetail/2017/04/daily-chart-6