In this section of* Resonance, *we invite readers to pose questions likely to be raised in a classroom situation. We may suggest strategies for dealing with them, or invite responses, or both. "Classroom" is equally a forum for raising broader issues and sharing personal experiences and viewpoints on matters related to teaching and learning science.

T Krishnan
Systat Software Asia-Pacific (P) Ltd.
5th Floor, C Tower
Golden Enclave, Airport Road
Bangalore 560 017, India.
Email: krishnan@systat.com

Flight Overbooking Problem
Use of a Probability Model

Consider an airline that has a daily flight from Bangalore to Singapore. Suppose it flies an aircraft that has a capacity of 200. Suppose further that it has already given confirmed booking to 200 passengers. When the 201st passenger asks for a confirmed booking, should it oblige? For that matter, should it keep making confirmed bookings beyond 200 and if so for how many?

According to international air travel regulations, agreements and practices, airlines are allowed a certain amount of 'overbooking' (confirmed booking beyond the seating capacity for the flight); this is because cancellations are fairly frequent in international flights for many reasons, some of them due to no fault of the passengers, such as late arrival of flights connecting with this flight, non-arrival of visas, etc. And so there are no cancellation charges and so passengers can take another flight without extra payment.

Let us use some notations suitable for this context:

N: Number of passengers that can be booked
M: Number of passengers that were booked
X: Number of passengers showing up for check-in.

Keywords
Probability model, binomial distribution, expected value, optimisation.

How many passengers (N) should an airline overbook, if it has 200 seats?

If $M \leq 200$ passengers have been booked, there is no decision to be made. If $M > 200$ passengers have been booked and $X \leq 200$ show up for check-in, then also there is no problem. However, if $M > 200$ passengers have been booked and $X > 200$ passengers show up for check-in, there is a problem. Ideally, the airline would like to see exactly $X = 200$ passengers show up for check-in, out of the $M > 200$ booked.

Now, what happens when $X > 200$ show up for check-in out of the $M > 200$ who have been booked? According to international regulations, certain priority criteria are to be followed and those who cannot be accommodated have to be provided with certain facilities like hotel, food, transport, passage in the next available flight of any airline, etc. This results in a certain loss of goodwill and a certain amount of financial loss. Without going into details, let us say that every booked passenger not accommodated in the flight costs the airline Rs. $c = 25,000$. So the airline has to be careful as to the choice of N the number that can be booked. And this has to be determined beforehand and entered into the computer system. On the other hand, it cannot play safe by choosing N to be 200, thereby running the risk of seats going empty resulting in a loss of revenue. Let us say that this loss per seat is $d = 15,000$, the amount the airline would have gained had the seat been filled. How should it choose this N? This cannot be very large and $\gg 200$, resulting in a large chance of $X > 200$; and it cannot be just a little more than 200 either, resulting in a large chance of $X < 200$. So there must be an optimum choice (in respect of maximum gain) of N What does it depend on? Evidently on the probable values of X and on the relative values of c and d, say $r = c/d$. X is subject to chance and we regard it as a *random variable.* Clearly, if r is very large, then N should be close to 200 and if r is very small, we can have N to be large.

There must be an optimum choice of N, depending on relative losses and probability of a booked passenger not checking-in.

Under reasonable assumptions, the number of passengers reporting for a flight has a binomial distribution.

Thus to solve the optimisation problem, we need to model the distribution of X Let us make some reasonable assumptions and work out the probability distribution of X Let us assume that the chance of a booked passenger not showing up is $p = 0.05$. In practice, the airline will determine this p from past data on 'no show' for similar flights. Let us also assume that passengers show up or not show up independently – not completely correct since for instance families travelling together either all show up or none at all. So this is only an approximation. Under these assumptions, X has what is called the *binomial distribution* with 'parameters' M and p. However, since we are interested in the consequences of the choice of N and $M < N$ creates no additional problems, let us make M to be equal to N Thus we make use of the binomial distribution of X which has the probability distribution

$$P(X = x) = \binom{N}{x} p^x (1-p)^{N-x} \quad x = 0, 1, 2, \qquad N.$$

Intuitively, we feel that if we book N passengers, on an average $N \times 0.95$ will turn up and so if we solve the equation

$$0.95N = 200 \Longrightarrow N \approx 211.$$

But this solution does not take into account the relative costs of empty seats and cost of denial of a seat. Let us discuss the problem in greater detail.

Let us first make a table of the possible values of X and the consequences. This is *Table* 1.

Now we are working in a situation where things are uncertain and we have formulated a probabilistic model for the phenomenon. We need a suitable optimisation criterion for this situation. Let us agree to choose N in such a way that the average value $R(N)$ (*expected value* as it is called in probability theory) of the random variable 'net income' is maximised as a function of N

$X = x$	Revenue	Loss	Net income y_x	Probability p_x
0	0	0	0	$(0.05)^N$
1	d	0	d	$N(0.95)(0.05)^{N-1}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
x	xd	0	xd	$\binom{N}{x}(0.95)^x(0.05)^{N-x}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
200	$200d$	0	$200d$	$\binom{N}{200}(0.95)^{200}(0.05)^{N-200}$
201	$200d$	c	$200d - c$	$\binom{N}{201}(0.95)^{201}(0.05)^{N-201}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
x	$200d$	$(x - 200)c$	$200(c + d) - cx$	$\binom{N}{x}(0.95)^x(0.05)^{N-x}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
N	$200d$	$(N - 200)c$	$200(c + d) - cN$	$(0.95)^N$

Table 1. Net income and its probability.

The average value $R(N)$ is the weighted average of the various possible values of net income, the weights being the corresponding probabilities. Thus it is the sum of products of the last two columns, that is

$$R(N) = \sum_{x=0}^{N} y_x p_x.$$

And this is equal to

$$R(N) = d[Np+200(1+r)P(X > 200)-(1+r) \sum_{x=201}^{N} xp_x].$$

For the range of r from $\frac{15}{25}$ to $\frac{20}{10}$, we surmise that the optimal value of N (the value that maximises $R(N)$) must be between 205 and 215 and hence we tabulated $R(N)$, rather $\frac{R(N)}{d}$ in this range of N Also, it is intuitively clear that $R(N)$ is unimodal and so it should increase for a while and then decrease (in the range 205 to 215 of N we were hoping), so that we can locate the optimal value of N *Table* 2 is the result of such a computation.

The empirical approach solved the practice problem satisfactorily.

N	$P(X > 200)$	$\sum_{x=201}^{N} xp_x$	Value of $\frac{R(N)}{d}$ for $r =$			
			$\frac{15}{25}$	$\frac{15}{15}$	$\frac{25}{15}$	$\frac{20}{10}$
205	0.0224	4.512	194.699	194.686	194.664	194.654
206	0.0522	10.523	195.567	195.534	195.478	195.451
207	0.1036	20.904	196.355	196.282	196.159	196.097
208	0.1794	36.238	197.027	196.883	196.644	196.525
209	0.2781	56.248	197.546	197.295	196.876	196.667
210	0.3926	79.526	197.890	197.488	196.817	196.481
211	0.5128	104.059	198.051	197.451	196.452	195.952
212	0.6282	127.742	198.037	197.196	195.795	195.094
213	0.7300	148.800	197.869	196.749	194.882	193.949
214	0.8135	166.278	197.576	196.145	193.760	192.567
215	0.8773	179.875	197.187	195.422	192.479	191.007
	Optimal $N \Longrightarrow$		211	210	209	209

Table 2. Calculation of R(N) for various N and r for p = 0.95.

We used the computer and a statistical software called Systat to carry out these computations.

When $r = 25/15$ (for $c = 25,000$ and $d = 15000$, for instance) the optimal value of N is 209 for $p = 0.95$.

Of course, the solution here was obtained 'empirically' A mathematical approach maybe to show that $\frac{R(N)}{R(N+1)}$ is < 1 for values of $N < N_0$ and is > 1 for $N > N_0$ thereby locating the optimal value N_0 of N where this change occurs (and also demonstrating unimodality). Alternatively, one may try to show that $R(N) - R(N+1)$ is > 0 up to a certain value of N and < 0 thereafter. We found this difficult to do, especially since $P(X > 200)$ and p_x are functions of N Maybe there are other approaches. Anyhow, an empirical approach seemed a nice way out. It did solve the practical problem satisfactorily. If you are a mathematician, maybe this method of solution is not entirely satisfactory to you. But then, by all means try a mathematically satisfactory solution!

Suggested Reading

[1] R L Karandikar, On randomness and probability: How to model uncertain events mathematically, in: Mohan Delampady, T Krishnan and S Ramasubramanian (Editors): *Probability and Statistics*. Hyderabad: Universities Press, 2001, from *Resonance*, Vol.1, No.2, pp.55-68, 1996.