

Cowles Foundation Paper 119b

AIRLINE DEMAND:
AN ANALYSIS OF SOME FREQUENCY DISTRIBUTIONS*

Martin J. Beckmann and F. Bobkoski[†]
Yale University

Following some trial fittings of Poisson and compounded Poisson distributions, Gamma distributions are fitted to histograms of manifested passengers, to total losses (late cancellations, no-shows, and misconnections) and to passengers added.

INTRODUCTION

The scheduling of departure times for flights on a given route is a complex problem involving many considerations. One of the simplest questions concerns the optimal over-all capacity of flights on a route for a given period. Obviously this problem requires a balancing of the revenue obtainable with a certain capacity against its cost. Now passenger demand is a random variable. The mean value of revenue as a function of capacity has the shape of a saturation curve, related in a simple manner to the distribution function of passenger demand: it is the expected value truncated at the capacity level. For a more sophisticated calculation of expected revenues, the probabilities of cancellations, of no-shows and misconnections, and of stand-by demand should be considered separately. It is thus of some value to know the actual passenger frequency distributions on typical flights. In this paper some empirical frequency distributions will be presented, together with some theoretical distribution functions that have been fitted to them.

Some obvious difficulties should be pointed out first. The "permanent factors" affecting demand rarely remain constant over significant periods, changes of schedule and of equipment occurring every few months. In the past, total demand for airline transportation has also been subject to a substantial upward trend. On flights with a heavy demand relative to capacity—the most interesting ones as regards the decision problem—only part of the demand distribution is observable, the frequency of demands for more seats than capacity appearing in the aggregate only.

The data used were the daily load reports of one airline covering the period from September 1954 through May 1955, listing for each flight the number of passengers manifested (i.e., on the passenger list a few hours before departure). Available for shorter periods were

*Research undertaken by the Cowles Commission for Research in Economics under contract Nonr-358(01), NR 047-006, with the Office of Naval Research.

We are indebted to the airline which kindly supplied us with the data for this analysis, and particularly to Messrs. McQuillen and Lets from whose advice we have greatly benefited.

[†]Manuscript received April 10, 1957.

data on the various losses (late cancellations, no-shows, and misconnections) and, for a few months, the passengers added.

1. THE DEMAND FOR RESERVATIONS

1.1 The Poisson Approximation

Flying to a particular destination on a particular day may be regarded a "rare event" for any given passenger. The probability distribution of the total demand for a flight may therefore be assumed related to the Poisson distribution for rare events. More precisely we may argue as follows: let the time period during which reservations may be made for a flight be divided into a number of intervals such that the mean number of requests for reservations received during any one interval is the same. Let time now be measured in terms of these interval units, one interval being one time unit, and consider the probability of receiving a call for a reservation during a small time interval. If the population of potential customers is large (consequently, if the probability of any one desiring to go on this flight is small) then the number of reservations already received does not affect the probability of further requests for a reservation. Suppose, to begin with, that reservation requests are made for one seat at a time. If the interval considered is sufficiently small, the probability of getting one reservation request is proportional to the length of the interval, and the probability of getting more than one request is zero. It is well known [Ref. 1, pp. 115-118] that on these assumptions the accumulated reservations over any fixed time interval of finite length must then be Poisson distributed. Let n be the number of reservations. Then the probability $f(n)$ of n reservations is

$$(1) \quad f(n) = \frac{\lambda^n e^{-\lambda}}{n!},$$

where $\lambda > 0$ is a constant, the mean of n . In particular, this must be true for the total reservation requests that have accumulated at flight-departure time if we count as requests only those that are not cancelled later.

Examination of data restricted to seasons and to such weekdays that have an approximately stationary demand shows that there is some tendency for the empirical distributions to approximate the Poisson distribution, but the fit is rather poor (Figures 1 and 2). The usefulness of this approximation is also limited because one would not wish to distinguish too many different cases of weekdays and seasons as would be required in order to arrive at a "pure" situation.

1.2 Poisson Process with Clustered Demand

Histograms of passengers manifested seem to show that on certain flights demand occurs more frequently for an even number of seats than for an odd number (Figure 2). This fact may be explained by assuming that requests for reservations obey a Poisson distribution, as before, but that each request is for either one or two seats, with constant probabilities. If we assume that p is the probability that one seat is requested and $q = 1 - p$ the probability of a request for two seats, the probability of a total demand for n reservations can be shown to be

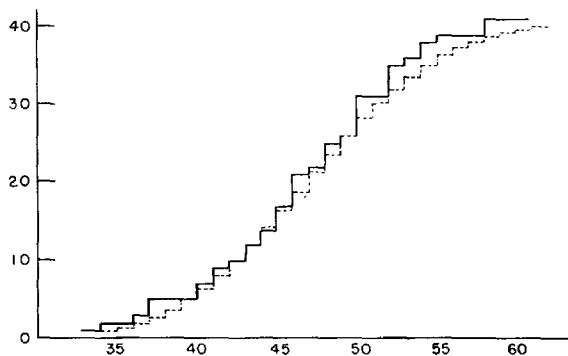


Figure 1 - Flight X, September 7 - November 18, 1954, Tuesday through Friday only. Dotted line: Poisson cumulative, $\lambda = 46.3$. Solid line: Observed cumulative.

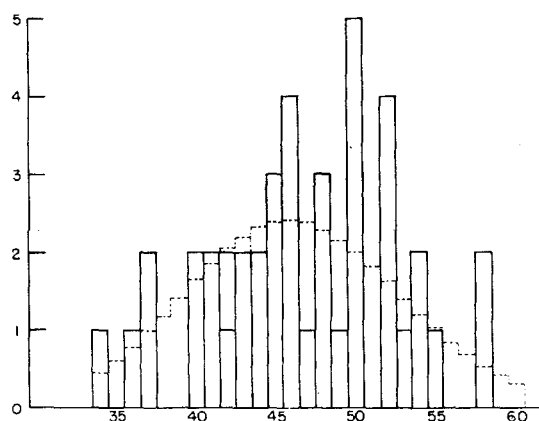


Figure 2 - Flight X, September 7 - November 18, 1954, Tuesday through Friday only. Dotted line: Poisson frequency distribution, $\lambda = 46.3$. Solid line: Observed frequencies.

$$(2) \quad p(n) = e^{-\lambda} \sum_{k=n-\lfloor n/2 \rfloor}^n \frac{\lambda^k q^{n-k} p^{2k-n}}{(n-k)! (2k-n)!}$$

where $[a]$ denotes the largest integer not greater than a . Therefore, the smallest integer greater than or equal to $[n/2]$ is $n - [n/2]$. The number of terms in the expression for $p(n)$ increases by one when n increases by one from an odd to an even number and does not change when n increases from an even to the next odd number. This indicates a tendency for even-numbered demands to have a larger probability.

Formula (2) is cumbersome. A more symmetric expression can be obtained on the assumption that demand occurs also for more than two seats at a time and that the clustering is subject to some simple probability distribution. Among plausible candidates, the geometric distribution is perhaps the simplest one. In order to achieve a rapid decrease of the probability of larger clusters, a rather considerable probability would be required for demands of one. It is natural now to associate the probabilities of the ordinary geometric distribution with demands that are one unit larger. Then the probabilities $p(k)$ of a demand for k seats are given by

$$(3) \quad \begin{aligned} p(0) &= 0 \\ p(k) &= (1 - p) p^{k-1} \quad k > 0, \end{aligned}$$

where p is a constant, $0 < p < 1$.

The probability of a total demand for n seats may now be derived as follows. The generating function of the Poisson distribution is $e^{-\lambda + \lambda s}$, that of the ordinary geometric distribution is $\frac{1-p}{1-ps}$, and that of the geometric distribution as used here is $\frac{(1-p)s}{1-ps}$. We are interested in sums of geometrically distributed variables for which the number of terms obeys a Poisson distribution. The generating function of the distribution of these sums is obtained by substituting in $e^{-\lambda + \lambda s}$ the expression $\frac{(1-p)s}{1-ps}$ for s [Ref. 1, p. 223]. Thus

$$\Phi(s) = e^{-\lambda} + \lambda \frac{(1-p)s}{1-ps}$$

is the desired generating function. This may be expanded as follows.

$$\Phi(s) = e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^k s^k}{k!} \left(\frac{1-ps}{1-p} \right)^{-k}.$$

Developing $(1 - ps)^{-k}$ into a binomial series, we have

$$(1 - ps)^{-k} = \sum_{i=0}^{\infty} \binom{-k}{i} (-ps)^i.$$

Now

$$\begin{aligned} \binom{-k}{i} &= (-1)^i \frac{k(k+1)\dots(k+i-1)}{i!} \\ &= (-1)^i \binom{k+i-1}{i}. \end{aligned}$$

Thus

$$(1 - ps)^{-k} = \sum_{i=0}^{\infty} \binom{i+k-1}{i} (-p)^i s^i,$$

and, finally,

$$\Phi(s) = e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} \sum_{i=0}^{\infty} \binom{i+k-1}{i} p^i s^{i+k}.$$

Now let $i + k = n$

$$\Phi(s) = \sum_{n=0}^{\infty} \left[e^{-\lambda} \sum_{k=1}^n \frac{1}{k!} \binom{n-1}{k-1} \lambda^k p^{n-k} \right] s^n.$$

This shows the probability of a demand n to be

$$(4) \quad f(n) = e^{-\lambda} \sum_{k=1}^n \binom{n-1}{n-k} \frac{1}{k!} \lambda^k p^{n-k}.$$

The mean and variance of this distribution are

$$\mu_1 = \frac{\lambda}{1-p},$$

$$\mu_2 = \frac{\lambda(1+p)}{(1-p)^2}.$$

These two moments of the distribution were fitted to the histogram of one flight* with a sufficiently low load factor in order that all demands should fall within the capacity limits of the flights, and a maximum number of observations became available (Flight Y, Figure 3). A chi-square test did not reject the distribution at the 5-percent significance level. But an unexpectedly large value of p was obtained, namely p = .50, so that it remains doubtful that the distribution (4) really explains the observed frequencies. This p-value would predict probabilities $\left(\frac{1}{2}\right)^i$ for a demand cluster of size i.

1.3 The Gamma Distribution

The observed frequency distributions of manifested reservations tend to be more dispersed than Poisson distributions. This condition is probably due not only to demand clustering but also to variations in the mean demand that are caused by random events such as weather, holidays, or conventions. This would mean that the Poisson parameter λ should itself be regarded a random variable. If, for instance, λ obeys a Gamma distribution, then it is well known [Ref. 2, p. 125] that the variable n follows a negative binomial distribution

$$(5) \quad f(n) = \binom{-r}{n} (1-p)^r (-p)^n,$$

where p is a constant, $0 < p < 1$, and r a positive integer.

Compounding this distribution for the number of reservation requests, with different distributions for the number of seats per request, we can obtain various distributions for the total number of reservations but involving rather complex expressions. Rather than introducing the Γ -distribution in this way by the back door, so to speak, we shall use it for a direct approach to the distribution of total reservations.

For the cumulative distributions we use the notation

$$(6) \quad \int_0^x \frac{\alpha^k u^{k-1}}{\Gamma(k)} e^{-\alpha u} du = \int_0^{\alpha x} \frac{u^{k-1} e^{-u}}{\Gamma(k)} du = \Gamma_{\alpha x}(k).$$

The mean and variance of the Γ -distribution are, respectively,

$$\mu_1 = \frac{k}{\alpha}, \quad \mu_2 = \frac{k}{\alpha^2}.$$

*A change of schedule did not permit continued observation of flight X.

The mean of the hypothetical Γ -distribution was fitted to the observed means and then a maximum likelihood estimate of the exponent k was obtained for two flights (Figures 3 and 4).

In view of the fact that no observations were left out for the entire 10-month period, which included the winter season, the agreement is good enough to accept the Γ -distribution. More work would be needed to establish its universal applicability.

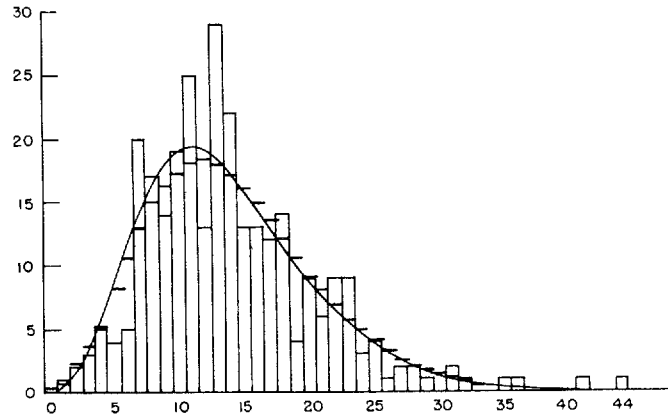


Figure 3 - Flight-Y manifests.
Curved line: Maximum likelihood fit of $(x + 1/2)$ to

$$\Gamma\text{-density; } \frac{\alpha^k}{\Gamma(k)} x^{k-1} e^{-\alpha x}; \hat{k} = 4.94991, \\ \hat{\alpha} = 0.148541.$$

Heavy short line: Method of moments fit to compound Poisson with geometric. Parameter in geometric, 0.506; parameter in Poisson, 6.903.

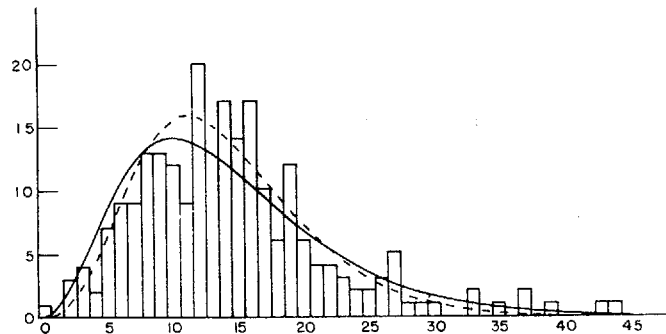


Figure 4 - Flight-Z manifests; Γ -distribution.
Dotted curve: Maximum likelihood estimate; $-\hat{\alpha} = 0.34, \hat{k} = 4.96.$
Solid curve: Moments estimate; $-\hat{\alpha} = 0.25, k = 3.69.$

2. LOSSES

Histograms of late cancellations, of misconnections, and of no-shows are presented without comment (Figure 5). The number of days for which data were available in useful form was limited and not sufficient to permit conclusions about the nature of the distributions involved.

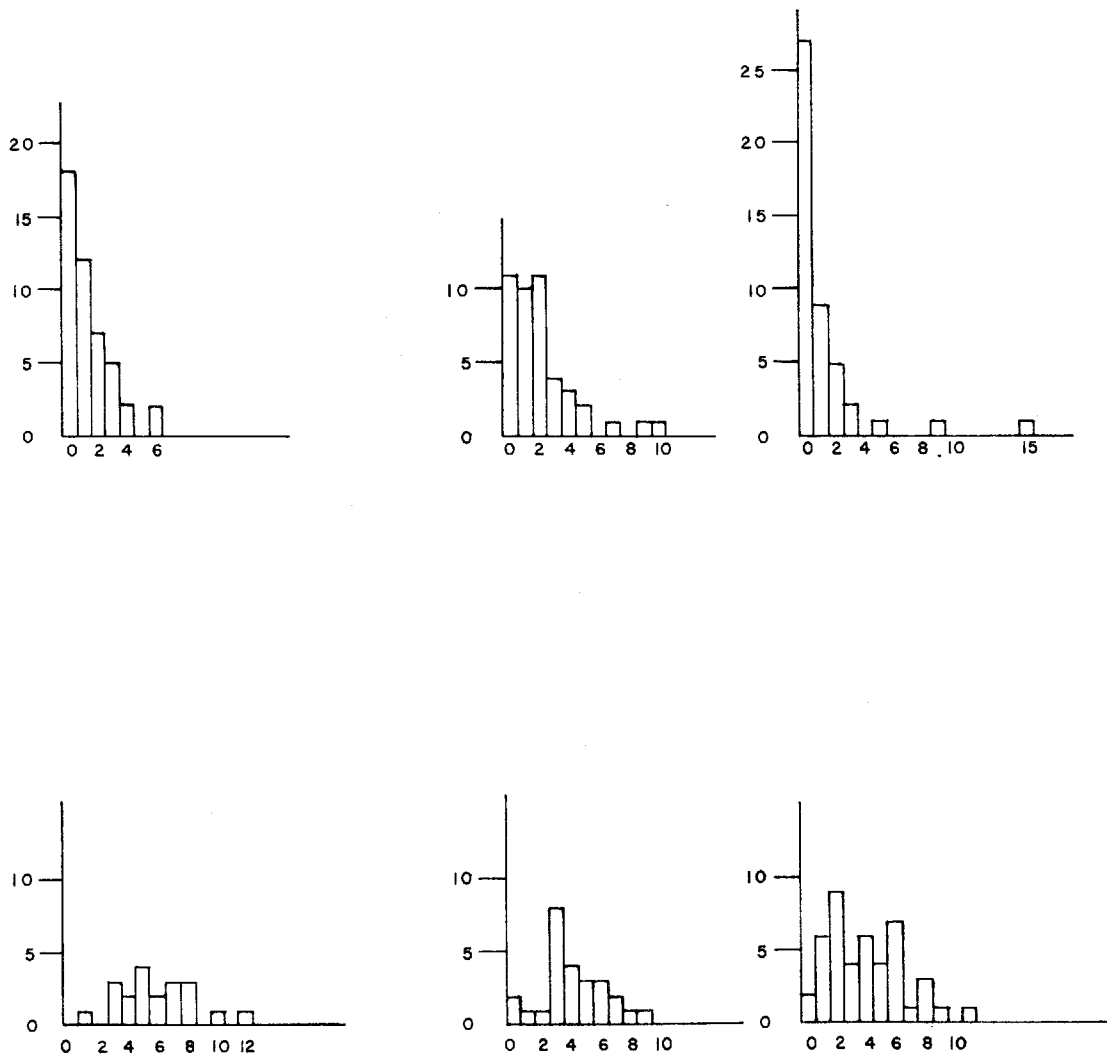


Figure 5 - Histograms of late cancellations, misconnections, and no-shows

Since late cancellations and misconnections are usually small compared with no-shows, the distribution of total losses may be regarded as a modified distribution of no-shows. Since it is the total losses that matter in the decisions mentioned later (the sale goal and the stop-sales problem) their distribution has been studied more carefully (Figures 6 and 7). The approximation by Γ -distributions seems satisfactory, and the exponents turn out close to each other and to those of the demand distributions.

3. DEMAND OF STAND-BY PASSENGERS

The population of stand-by passengers (adds) consists of through passengers who missed connections or changed plans and of local passengers who either did not bother about making a reservation or took a chance upon being unable to secure a reservation in time. The superposition of these demand elements may produce complex-shaped distributions. The two histograms (Figure 8) show unimodal distributions to which Γ -distributions have been fitted.

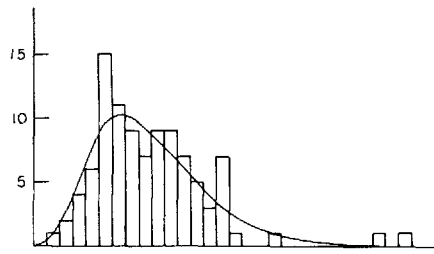


Figure 6 - Flight-Q total losses for manifests of 50 and over. Γ -distribution using $(x - 1/2)$; $\hat{\alpha} = 0.478$, $\hat{k} = 4.257$.

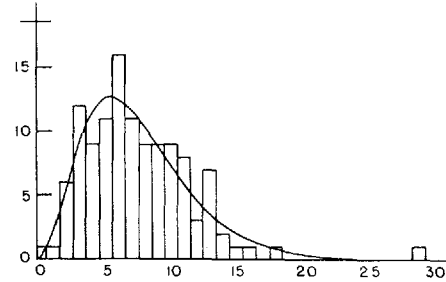


Figure 7 - Flight-P total losses for manifests of 50 and over. Γ -distribution; $\hat{\alpha} = 0.469$, $\hat{k} = 3.733$.

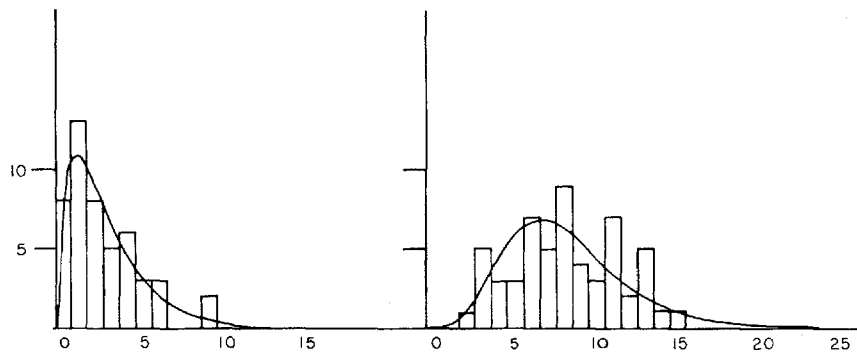


Figure 8 - Histograms of passengers added to which Γ distributions have been added

The observations cover a period of only two months and of only such days for which demand was low enough to permit the total demand of stand-bys to be accommodated. Unfortunately, this excludes from observation the demand of passengers "unable to secure a reservation"; for this demand would show up mainly on days when not all available adds can be accommodated.

CONCLUSIONS

Some further possibilities of applying demand distributions may be pointed out in conclusion. In choosing between types of equipment one must compare costs and expected revenues. As we have seen, the latter are proportional to a truncated expected value dependent on the demand distribution.

In setting an optimal price of a noncompetitive flight, the demand distribution should be considered for its effect on both revenue and on the passenger's average chance of obtaining a reservation. This has been discussed elsewhere [3]. Finally, some delicate questions of reservations' policy cannot be approached without information as to the distributions of reservations' demand, late cancellations, no-shows, and stand-by demand. These questions which revolve around the problem of keeping oversales below a certain limit, while simultaneously reducing lost revenue from no-shows to a minimum, have been investigated at length in another paper [4].

REFERENCES

- [1] W. Feller, An Introduction to Probability Theory and Its Application, Wiley, New York (1950).
- [2] R. G. Kendall, The Advanced Theory of Statistics. Giffin and Co., London (1943), Vol. I.
- [3] M. Beckmann, "The Pricing of Fixed Services Subject to a Random Demand," Economica Internazionale, Vol. IX No. 2, pp. 205-209, Mar. 1956.
- [4] M. Beckmann, "Decision and Team Problems in Airline Reservations," Econometrica Vol. 26, No. 1 (Jan 1958)