Neural, Parallel, and Scientific Computations xx (200x) xx-xx

AIRLINE SPILL ANALYSIS USNIG GUMBEL AND MOYAL PROBABILITY DISTRIBUTIONS

GOKARNA R. ARYAL [1] AND CHRIS P. TSOKOS [2]

[1]Department of Mathematics CS and Statistics
Purdue University Calumet
Hammond, IN 46323 USA.
[2]Department of Mathematics and Statistics
University of South Florida
Tampa, FL 33620 USA.

ABSTRACT. The analysis of the passenger spill data is an integral part for the determination of optimal aircraft capacities in the airline fleet assignment process. Most research in this study has assumed that the nominal demand is normally distributed. But empirical evidences show that normal probability distribution is not suitable in many cases because of asymmetry nature of the data. In the present study, we derive the essential formulae for calculating the expected number of spilled passengers when the nominal demand is assumed to follow a Gumbel and a Moyal probability distributions. A comparison of the corresponding results with a normal, a logistic, a log-normal, and a gamma probability distributions is also presented.

Key Words Nominal demand, Spill Analysis, Gumbel Distribution, Moyal Distribution.

1. Introduction

The total number of potential passengers who cannot obtain a reservation and travel on a given flight due to insufficient capacity is called the airline spill. Since these passengers are rejected by the airline because the number of seats on the aircraft assigned to the flight is less than total potential demand the spill is also known as rejected demand. That is,

Spill= Total demand - Total load of a flight.

When demand is less than capacity, load is equal to demand, and spill is zero. When demand exceeds the capacity of an aircraft, load is equal to capacity and spill is equal to demand minus capacity (load). Spill occurs as the result of greater potential demand for a flight than the physical capacity of the aircraft.

The correct estimation of spill, or passenger demand turned away, is an integral part of the determination of optimal aircraft capacities in the airline fleet assignment process.

In practice it is difficult if not impossible to observe the actual “demand ” for a flight [2]. Since the airlines have no way of keeping track of how many requested for booking were turn away or rejected because there is no enough space the notion of “total demand” for a particular flight or set of flights operated over a period of times is a theoretical concept.

Received xxxx, 200x

The Spill models predict average lost sales when demand exceeds the flight capacity, and have been used by the airline industry since mid-1970s (Schlifer and Vardi, 1975). These models provide critical information for decisions in selecting aircraft size for a particular market, assigning an airline's existing fleet to various markets and schedules, and planning the aircraft type for future expansion. They are an integral part of modern yield management system. The basic idea behind a spill model is that demand for a group of flights can be represented by a probability distribution. The group of flights can be defined according to the wishes of the analyst or the airline. For example, it can involve just one flight segment, a small group of segments served by single or multiple aircraft types or all the segments served by a single fleet type. Airline passenger spill analysis has traditionally relied on the normal probability distribution for the nominal demand. But empirical evidences show that the normal probability distribution does not fit very well in many cases, especially for business and first class compartments (Swan, 1992). The subject data is not very symmetrical and is often skewed to the right of its mean. Li Michael et al. [1] give some alternatives introducing the possible models could be logistics or log-normal and the gamma probability distributions. Since empirical evidence shows that the positively skewed model will be appropriate it is very important to choose a right probability distribution among the many competing distributions. In this paper we will present Gumbel and Moyal probability distribution that offer a more flexibility of the probabilistic characterization of spill data.

2. **Definitions and Notations**

Here we would like to mention some fundamental definition essential in the subject study. Also see Li et al.(2000).

Let X be the nominal demand for a flight with capacity of C that could be the effective capacity. Then we proceed to identify the following:
1. Nominal load factor (NLF) is defined as E(X)/ C, E(X) being the expected value of the nominal demand.
2. Observed mean load is the expected value of the nominal demand X truncated at the capacity level C, that is, E(min(X, C)); and the mean observed load factor (OLF) is defined as E(min(X, C))/C.
3. The fill rate (FR) for the pth seat is defined as the probability that the demand is equal to or greater than p, that is, $P(X \geq p)$.

4. The spilled passengers (SP) is the number of passengers turned away because the flight is fully booked, that is,

$$SP = E[(X - C)I_{X>C}],$$

where $I_{X>C}$ is the indicator function.
5. The spill rate (SR) is defined as the ratio of spilled passengers over the mean of the nominal demand, that is,

$$SR = \frac{SP}{E(X)} = \frac{E[(X - C)I_{X>C}]}{E(X)}.$$

The relationship between the nominal load factor (NLF), the observed load factor (OLF), and the spilled passengers (SP) is given by

$$OLF = NLF - SP/C$$

and consequently,

$$OLF = (1 - SR) \times NLF.$$

Li Michael et al.[1] provides the spill formula for the normal, logistic, log normal and gamma probability distributions. If the random variable X follows a normal distribution, $N(\mu, \sigma^2)$, with mean μ and variance σ^2 then the expected number of spilled passengers is given by

$$SP = \sigma[\phi(b) - b(1 - \Phi(b))]$$

where

$$b = (C - \mu)/\sigma.$$

$\phi(x)\&\Phi(x)$ denotes the probability density function (pdf) and the cumulative distribution function (cdf) of standard normal $N(0, 1)$ probability distribution.
If the random variable X follows a logistic distribution with parameters θ and β then the expected number of spilled passengers is given by

$$SP = \beta \ln\left(1 + \exp(-\frac{C - \theta}{\beta})\right).$$

If the random variable follows the lognormal probability distribution with parameters μ and σ then the expected number of spilled passengers is given by

$$SP = \exp(\mu + \sigma^2/2)\left(1 - \Phi(c - \sigma)\right) - C\left(1 - \Phi(c)\right),$$

where $c = \frac{\ln C - \mu}{\sigma}$
Also if the random variable follows the gamma probability distribution with parameters α and β then the expected number os spilled passengers is given by

$$SP = \alpha\beta[1 - G(C, \alpha + 1, \beta)] - C[1 - G(C, \alpha, \beta)]$$

where

$$G(x; \alpha, \beta) = \int_0^x \frac{1}{\Gamma(\alpha)\beta^\alpha} t^{\alpha-1} e^{-t/\beta} dt$$

is the cumulative probability function of the gamma probability distribution with parameters (α, β).

3. Spill formula for Gumbel Probability Distribution

The Gumbel probability density function(pdf) for a random variable X is given by

$$f(x) = \frac{1}{\sigma} \exp\left[-\frac{x - \mu}{\sigma} - \exp\left\{-\frac{x - \mu}{\sigma}\right\}\right] \qquad -\infty < x < \infty \tag{3.1}$$

where μ and σ are the location and scale parameters respectively. This is the first extreme value distribution also known as Type I extreme probability distribution. It has been used in engineering, environmental modeling, finance, fire protection and insurance problems and for the prediction of earthquake magnitude as well for modeling extreme temperatures and carbon dioxide in the atmosphere and others. Since its pdf is skewed to the right it is a good candidate for modeling the airline spill data. In this section we will develop the spill methods for Gumbel random variable.
We know that if X follows the Gumbel probability distribution then its expected value and variance are given by

$$\begin{aligned} E(X) &= \mu + \gamma \times \sigma \approx \mu + 0.57722\sigma \\ Var(X) &= 1.64493\sigma^2 \end{aligned}$$

respectively, where, $\gamma \approx 0.5772156649$ is the Euler's constant.

Therefore, the coefficient of variation (C.V.) is given by

$$C.V. = \frac{S.D.(X)}{E(X)} \approx \frac{1.28255\sigma}{\mu + 0.57722\sigma}.$$

Now, the expected value of the spill passengers is given by

$$\begin{aligned} SP &= E[(x-C)I_{X>C}] \\ &= \int_C^\infty (x-C)f(x)dx \\ &= \int_C^\infty \frac{(x-C)}{\sigma}\exp\left[-\frac{x-\mu}{\sigma} - \exp\left\{-\frac{x-\mu}{\sigma}\right\}\right]dx \\ &= \sigma\int_b^\infty y\exp[-y-\exp(-y)]dy - b\sigma\int_b^\infty \exp[-y-\exp(-y)]dy \end{aligned}$$

where $y = \frac{x-\mu}{\sigma}$ and $b = \frac{C-\mu}{\sigma}$.
Upon simplifying the above integrals, we have

$$SP = \sigma b[\exp(-z) - 1] - \sigma\int_0^z \exp(-t)\ln t dt \tag{3.2}$$

where $z = \exp(-b) = \exp(-\frac{C-\mu}{\sigma})$.

4. Spill formula for Moyal Probability Distribution

A random variable X is said to have Moyal probability distribution with parameters μ and σ if its probability density function (pdf) is given by

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left\{-\frac{1}{2}\left(\frac{x-\mu}{\sigma} + \exp(-\frac{x-\mu}{\sigma})\right)\right\} \tag{4.1}$$

where $-\infty < x < \infty$, $-\infty < \mu < \infty$ and $\sigma > 0$.
This distribution was introduced by J.E. Moyal as a good approximation to the Landau probability distribution. It has been shown that it remains applicable when we take into account quantasome resonance effects and details of atomic structure of the absorber. The Moyal probability distribution is used to characterize the energy loss by ionization for a fast charged particles and the number of ion pairs produced in this process. See [5] for details. Note that the Moyal probability distribution is skewed to the right and has mode μ but its mean and variance are given by

$$\begin{aligned} E(X) &= \mu + \sigma(\gamma + \ln 2) \approx \mu + 1.27\sigma \\ Var(X) &= \frac{\pi^2}{2}\sigma^2 \approx 4.9348\sigma^2 \end{aligned}$$

respectively, where, $\gamma \approx 0.5772156649$ is the Euler's constant.
Therefore, the coefficient of variation is given by

$$C.V. = \frac{S.D.(X)}{E(X)} \approx \frac{2.22\sigma}{\mu + 1.27\sigma}$$

Now,we can derive the analytical expression for the expected value of the spill passengers as

$$\begin{aligned} SP &= E[(x-C)I_{X>C}] \\ &= \int_C^\infty (x-C)f(x)dx \\ &= \frac{1}{\sqrt{2\pi}\sigma}\int_C^\infty (x-C)\exp\left\{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}+\exp(-\frac{x-\mu}{\sigma})\right)\right\} \\ &= \frac{\sigma}{\sqrt{2\pi}}\int_b^\infty (t-b)\exp\left(-\frac{1}{2}(t+e^{-t})\right)dt \end{aligned}$$

where $t=\frac{x-\mu}{\sigma}$ and $b=\frac{C-\mu}{\sigma}$.

We can further simplify the above integral and we have

$$\begin{aligned} SP &= \frac{\sigma}{\sqrt{2\pi}}\int_b^\infty t\exp\left(-\frac{1}{2}(t+e^{-t})\right)dt - \frac{b\sigma}{\sqrt{2\pi}}\int_b^\infty \exp\left(-\frac{1}{2}(t+e^{-t})\right)dt \\ &= \sigma b[1-2\Phi(z)]-4\sigma\Psi(z) \end{aligned}$$

where

$$z=exp(-b/2), \Phi(z)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^z e^{-\frac{y^2}{2}}dy, \Psi(z)=\frac{1}{\sqrt{2\pi}}\int_0^z e^{-\frac{y^2}{2}}\ln y dy$$

The researchers in the airline industry use the coefficient of variation and the expected demand to calculate the spill. Table 1 below provides the spill formulae for the Normal, Logistic, Log-normal, Gamma, Gumbel and Moyal probability distributions. Also in the table we expressed the parameters of each of the distribution in terms of mean μ_X and (or) coefficient of variation (CV).

Distribution	Reparametrization	Spill Formula(SP)
Normal	$\mu=\mu_X;\sigma=CV\times\mu_X$	$\sigma[\phi(b)-b(1-\Phi(b))], b=(C-\mu_X)/\sigma_X$
Logistic	$\theta=\mu_X;\beta=\frac{CV\times\mu_X}{\pi/\sqrt{3}}$	$\beta\ln(1+e^{-k}), k=\frac{C-\theta}{\beta}$
Log-Normal	$\mu=\ln\frac{\mu_X}{\sqrt{1+CV^2}};\sigma=\sqrt{\ln(1+CV^2)}$	$\mu_X(1-\Phi(c-\sigma))-C(1-\Phi(c)), c=\frac{\ln C-\mu}{\sigma}$
Gamma	$\alpha=\frac{1}{CV^2};\beta=CV^2\times\mu_X$	$\mu_X[1-G(C,\alpha+1,\beta)]-C[1-G(C,\alpha,\beta)]$
Gumbel	$\sigma=\frac{CV\times\mu_X}{1.28255},\mu=\mu_X-\gamma\sigma$	$\sigma b[e^{-z}-1]-\sigma\Omega(z), z=e^{(-b/2)}, b=\frac{C-\mu}{\sigma}$
Moyal	$\sigma=\frac{CV\times\mu_X}{2.22},\mu=\mu_X-1.27\sigma$	$\sigma b[1-2\Phi(z)]-4\sigma\Psi(z), z=e^{(-b/2)}, b=\frac{C-\mu}{\sigma}$

TABLE 1. *Reparametrization and Spill formulae of the Normal, Log-normal, Gamma, Logistic, Gumbel and Moyal probability distributions*

where $\Phi(z)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^z e^{-\frac{y^2}{2}}dy, \Psi(z)=\frac{1}{\sqrt{2\pi}}\int_0^z e^{-\frac{y^2}{2}}\ln y dy, \Omega(z)=\int_0^z \exp(-t)\ln t dt$

5. Comparisons of spill values

In this section, we will present the spill values computed under the six alternative distributions numerically. We have computed the spill values for the Gumbel and Moyal distributions and presented with the other four distributions that have been given by Li Michael et al.[1]. We have calculated the spill for two capacity levels: C=30 and and C=150, and three CV values: 0.2, 0.5 and 0.8. Tables 2 and 3 report spill values at capacity levels 30 and 150 seats per aircraft, respectively. Each of these tables are arranged so as to make it easy to compare spill values for the six alternative probability distributions at different levels of mean and coefficient of variation of the nominal demand.

$\mu_X =$		115	120	125	130	135	140	145	150	155	160	165	170
	A	0.6	1.2	2.1	3.3	4.9	6.9	9.2	12.0	15.0	18.4	22.0	25.8
	B	0.8	1.3	2.1	3.2	4.6	6.5	8.8	11.5	14.5	17.9	21.6	25.5
CV=0.2	C	1.1	1.7	2.6	3.8	5.3	7.1	9.3	11.8	14.7	17.9	21.3	25.0
	D	0.9	1.6	2.5	3.7	5.2	7.1	9.3	11.9	14.9	18.1	21.6	25.3
	E	1.4	2.0	2.9	4.0	5.4	7.1	9.1	11.5	14.2	17.2	20.6	24.2
	F	1.5	2.3	3.1	4.1	5.5	7.1	9.1	11.3	13.8	16.7	20.0	23.4
	A	9.6	11.9	14.4	17.1	20.1	23.2	26.5	29.9	33.5	37.2	41.0	44.8
	B	9.1	11.2	13.6	16.2	19.0	22.1	25.3	28.7	32.2	35.9	39.6	43.5
CV=0.5	C	11.0	12.9	15.0	17.3	19.8	22.4	25.1	28.0	31.0	34.2	37.5	40.9
	D	11.0	13.1	15.4	17.9	20.5	23.3	26.2	29.3	32.5	35.8	39.3	42.8
	E	10.8	12.9	15.1	17.5	20.1	22.8	25.7	28.7	31.9	35.2	38.6	42.2
	F	11.0	13.0	15.2	17.7	20.0	22.5	25.2	28.1	31.2	34.3	37.6	41.0
	A	21.8	25.2	28.6	32.3	36.0	39.9	43.8	47.9	52.0	56.2	60.5	64.8
	B	20.6	23.8	27.1	30.6	34.2	38.0	41.9	45.9	49.9	54.1	58.3	62.6
CV=0.8	C	21.4	23.9	26.5	29.2	32.1	35.0	38.1	41.2	44.5	47.8	51.2	54.7
	D	23.3	26.2	29.1	32.2	35.3	38.6	42.0	45.4	48.9	52.5	56.2	59.9
	E	22.7	25.7	28.8	32.0	35.4	38.8	42.4	46.0	49.7	53.5	57.4	61.3
	F	22.6	25.6	28.6	32.1	34.9	38.1	41.5	45.0	48.5	52.2	56.0	59.7

TABLE 2. *Spill table with capacity (C) =150. A=Normal, B= Logistic, C=Lognormal, D= Gamma, E=Gumbel and F= Moyal*

$\mu_X =$		20	22	24	26	28	30	32	34	36	38	40	42
	A	0.0	0.1	0.2	0.7	1.4	2.4	3.7	5.2	6.8	8.6	10.4	12.3
	B	0.0	0.1	0.3	0.6	1.3	2.3	3.6	5.1	6.8	8.6	10.4	12.3
CV=0.2	C	0.0	0.1	0.3	0.8	1.4	2.4	3.6	5.0	6.6	8.4	10.2	12.1
	D	0.0	0.1	0.3	0.7	1.4	2.4	3.6	5.1	6.7	8.4	10.3	12.2
	E	0.1	0.2	0.4	0.8	1.4	2.3	3.4	4.8	6.5	8.2	10.1	12.0
	F	0.1	0.2	0.4	0.8	1.4	2.2	3.3	4.7	6.3	8.1	10.0	12.0
	A	0.8	1.5	2.4	3.4	4.6	6.0	7.4	9.0	10.6	12.2	14.0	15.7
	B	0.8	1.4	2.2	3.2	4.4	5.7	7.2	8.7	10.3	12.0	13.7	15.5
CV=0.5	C	1.2	1.8	2.6	3.5	4.5	5.6	6.8	8.2	9.6	11.1	12.7	14.3
	D	1.2	1.8	2.6	3.6	4.7	5.9	7.2	8.6	10.0	11.6	13.2	14.8
	E	1.2	1.8	2.6	3.5	4.6	5.7	7.0	8.4	9.9	11.4	13.0	14.7
	F	1.2	1.8	2.6	3.5	4.5	5.7	6.9	8.2	9.6	11.1	12.7	14.3
	A	2.6	3.7	5.0	6.5	8.0	9.6	11.2	13.0	14.7	16.5	18.4	20.3
	B	2.5	3.5	4.8	6.1	7.6	9.2	10.8	12.5	14.3	16.1	17.9	19.8
CV=0.8	C	2.9	3.8	4.8	5.8	7.0	8.2	9.6	10.9	12.4	13.8	15.4	17.0
	D	3.1	4.1	5.2	6.4	7.7	9.1	10.5	12.0	13.5	15.1	16.7	18.3
	E	2.9	4.0	5.1	6.4	7.7	9.2	10.7	12.3	13.9	15.5	17.2	18.9
	F	3.0	4.0	5.1	6.3	7.6	9.0	10.5	12.0	13.5	15.1	16.7	18.4

TABLE 3. *Spill table with capacity(C)=30. A=Normal, B= Logistic, C=Lognormal, D= Gamma, E=Gumbel, F= Moyal*

6. The Shape of the Demand Distributions

As it has been already mentioned the normal probability distribution in most cases does not fit the actual spill data because of its symmetrical behavior among other restrictions. As pointed out in Li Michael et al.[1] the shape of demand distribution is skewed to the right for small cabins. Thus we recommend to choose the right skewed Gumbel and Moyal probability distribution for estimating the demand. In this section we will provide a graphical comparisons of all Normal, Logistic, Lognormal, Gamma, Gumbel and Moyal probability distributions for the same values of mean and cv. Statistical software R [4] have been used to perform all necessary calculations and graphics.

Note that both the Gumbel and Moyal probability distribution offer more flexibility with respect to skewness and asymmetry for small demand and coefficient of variation. As the demand increases we can observe that there is reduction in skewness. Also the graph clearly display as the mean and cv increases the probabilistic characterization of the skewed pdf is significantly different.

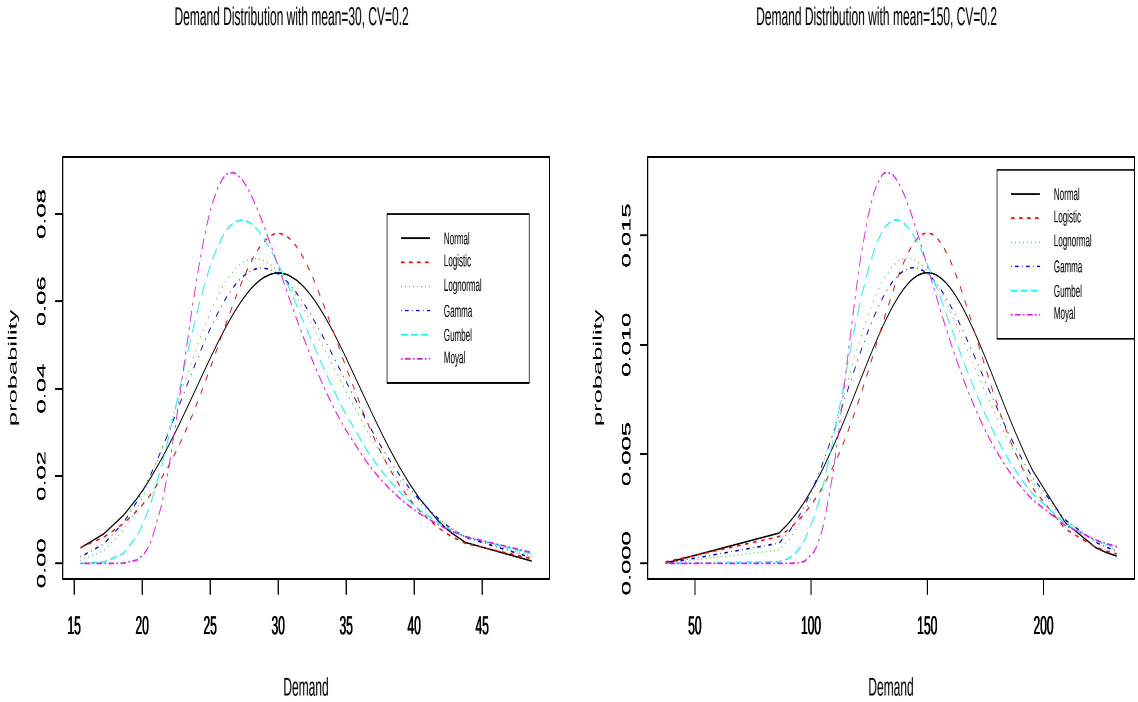

FIGURE 1. The demand distribution with cv=0.2

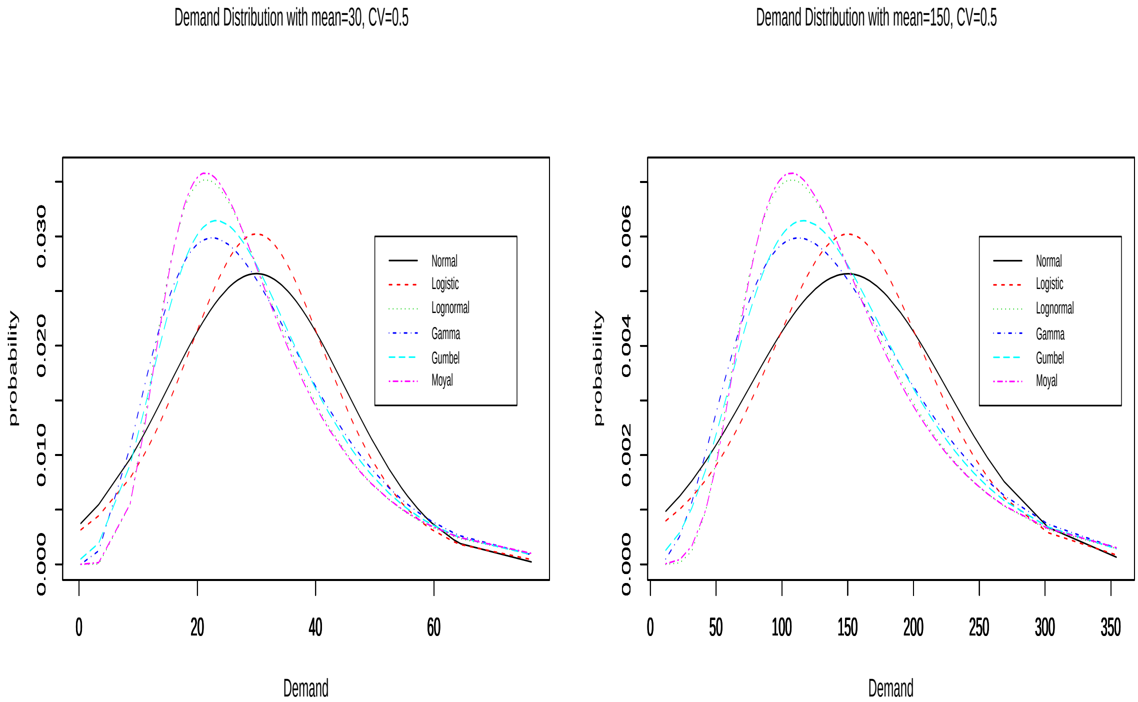

FIGURE 2. The demand distribution with cv=0.5

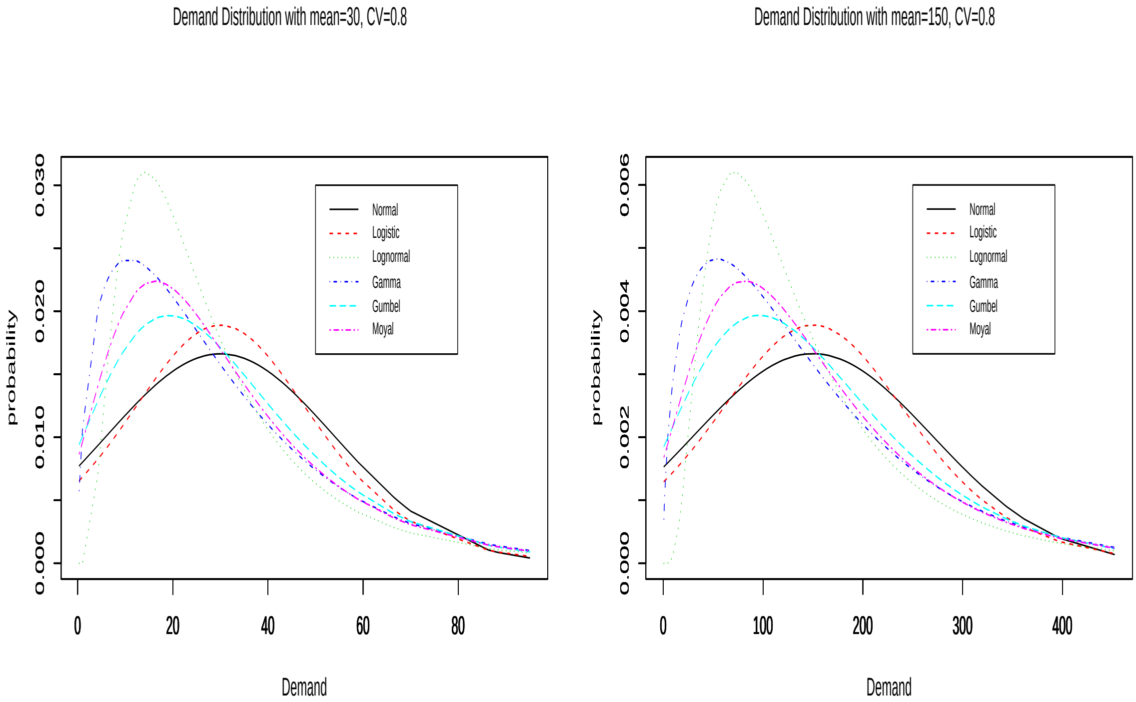

FIGURE 3. The demand distribution with cv=0.8

7. Concluding Remarks

We have introduced the Gumbel and Moyal probability distribution to characterize the probabilistic behavior of the airline spill data. In addition to the analytical development of all essential formulas to be applicable in the subject area, we have numerically and graphically compared the proposed pdf with other probability distributions that have been used to study the subject data. Finally, the proposed two probability distributions offer more flexibility to more accurately characterize the asymmetric behavior of spill data.

REFERENCES

[1] Michael Z.F. Li , Tae Hoon Oum (2000) Airline spill analysis beyond the normal demand, European Journal of Operational Research 125, 205-215.
[2] Peter P. Belobaba (2006) Airline demand Analysis and Spill Modeling.
[3] Johnson, N.L., Kutz, S., Balakrishnan, N., 1995. Continuous Univariate Distributions, vol. 2, 2nd Ed. Wiley, New York.
[4] R: A language for data analysis and graphics.
[5] Walck, Christian (2001). Hand-book on Statistical Distributions for experimentalists, university of Stockholm.