

STATISTICAL DISTRIBUTIONS
FOR DECISION ANALYSIS:
A SUPPLEMENT

BY

J. M. GATES

DOCUMENT DE TRAVAIL

No. DRV/86 09 SDA-Nantes

Department Strategies

de Developpment

et Amenagement

IFREMER, Paris.

NOVEMBER, 1985

107

TABLE OF CONTENTS

I. Introduction

II. The Triangular Distribution

III. The Uniform-Triangular Distribution

IV. The V Distribution

V References

INTRODUCTION

This working paper presents simple statistical distributions useful for decision analysis. For a concise summary of some twenty-eight commonly used distributions and their associated functions, see Hastings and Peacock (1974). This reference does not include the distributions described below; hence the *raison d'être* of this working paper. For a brief discussion of decision analysis, see Budnick, et al (1977).

The distributions presented are (1) the triangular distribution, (2) an extension of the triangular distribution (the rocket distribution) which allows "fat tail(s)" and (3) the V distribution.

THE TRIANGULAR DISTRIBUTION

The triangular density is well described by its name. It requires three parameters; a , b , and c . The parameter a is the minimum value the random variate X is allowed to take. The parameter b is the maximum value X may take on. Thus the interval (a,b) defines the range of X . The parameter c defines the modal frequency of the density function $f(X)$ defined over the interval (a,b) and zero elsewhere. This density has a triangular shape, hence the name. By definition of a and b , the triangle cannot be obtuse. A vertical line from the apex of the triangle (i.e. the modal frequency) intersects the line interval ab on the X axis at point c ; $a \leq c \leq b$. The apex point $(c, f(c))$ defines the modal point of the frequency. Limiting cases of $c=a$ or $c=b$ correspond to right triangles. By choosing the value of c , it is possible to approximate more complex distributions; some of which are more data hungry. The choice between, let us say, a lognormal and a skewed triangular distribution depends in part on the plausibility of the respective tails of the two distributions. The lognormal can be given no bounds other than $0 \leq X \leq \infty$. The triangular distribution can (must) be assigned upper and lower bounds; a fact which may be considered good or bad depending on circumstances of application.

Figure II.1
Triangular Distribution

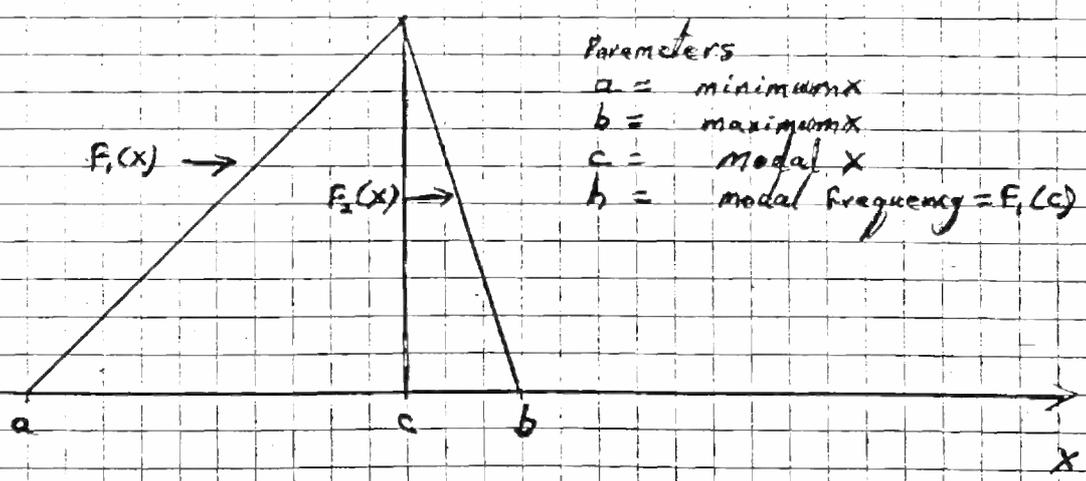
density

$f(x)$

$$\begin{aligned} \text{Density Function: } f(x) &= f_1(x) \\ &= f_2(x) \\ &= 0 \end{aligned}$$

$$\begin{aligned} a \leq x \leq c \\ c \leq x \leq b \\ \text{otherwise} \end{aligned}$$

h



Parameters

- $a =$ minimum x
- $b =$ maximum x
- $c =$ modal x
- $h =$ modal frequency $= F_1(c)$

Density Function:

$$\begin{aligned}f(X) &= f_1(X) = [h/(c-a)] * (X-a) = b_1 * (X-a) \quad a \leq X \leq c \\ &= f_2(X) = [h/(b-c)] * (b-X) = b_2 * (b-X) \quad c \leq X \leq b \\ &= 0 \text{ otherwise}\end{aligned}$$

where:

$h = [2/(b-a)]$ = constant of integration such that $F(b) = 1$; where $F(X)$ denotes the cumulative density or distribution function for $f(X)$.

Distribution Function:

$$\begin{aligned}F(X) &= F_1(X) = 0.5 * B_1 * ((X-a) \uparrow 2); \quad a \leq X \leq c \\ &= F_2(X) = F_1(c) + B_2 * [(-0.5 * X \uparrow 2) - bc + bX + (0.5 * c \uparrow 2)]; \\ &\quad \text{for } c \leq X \leq b \\ &= 0 \text{ otherwise.}\end{aligned}$$

In particular:

$$\begin{aligned}F(a) &= F_1(a) = 0 \text{ as required} \\ F(c) &= F_1(c) = F_2(c) = 0.5 * h * (c-a) \text{ (left tail)} \\ F(b) &= F_1(c) + F_2(b) = 1 \text{ as required.}\end{aligned}$$

The left tail of the probability mass is $F_1(c)$:

$$F_1(c) = [c-a] * [0.5 * (h_1 + h_2)]$$

The right tail of the probability mass is $F_2(b)$:

$$F_2(b) = [b-c] * [0.5 * (h_1 + h_3)]$$

Hence:

$$F(b) = F_1(c) + F_2(b) = 1; \text{ which implies:}$$

$$h_1 = 2 / [b-a + d_2 * (c-a) + d_3 * (b-c)]$$

$$h_2 = d_2 * h_1$$

$$h_3 = d_3 * h_1; \text{ by definition of } d_2 \text{ and } d_3.$$

These parameters are zero in the triangular density but they permit a generalization Part in III.

Inverse Distribution Function:

While the distribution function tells us the probability of $X \leq x_0$, it is not helpful in computer simulations where we wish to go the other way. That is, given an arbitrarily chosen probability, P_0 on a distribution function, how can we select values of X with the "correct" frequency? The inverse distribution

function: -1

$G(y) = F^{-1}(X) =$ inverse distribution function (IDF) answers this question.

Since negative densities are forbidden, $F(\cdot)$ is monotone non-decreasing and, in our case, positive monotone. Because of this, P and X have a "one-to-one" correspondence and $F(\cdot)$ is invertible. Except in the case where the mode and mean are equal (i.e. c is the midpoint of a, b), $F(X)$ has a "kink" at $X=c$. We must invert $F_1(X)$ and $F_2(X)$ separately.

$$\begin{aligned} G(y) = G_1(y) &= a + (1/B_1) * \text{SQRT}[(2 * B_1 * y)]; \quad 0 \leq y \leq F_1(c) \\ &= G_2(y) = b - \text{SQRT}[(b-c) * (b-a) * (1-y)]; \quad F_1(c) \leq y \leq 1 \end{aligned}$$

where $\text{SQRT}[\cdot]$ denotes the square root operator.

In general, the utility of the IDF of any density function lies in the fact that if y has a uniform $(0,1)$ distribution, then y can be used with the IDF to generate values of X with the correct frequency. All computers and many calculators have a uniform $(0,1)$ random variate function. Although the IDF exists, it may not be easy to obtain an analytic solution for it as we have for the triangular.

A criticism of the Triangular distribution is that it assumes zero frequencies at the extremes of its range. It may be desirable to permit "fat tail(s)" in a given application. By combining uniform and triangular distributions, we can create a generalization which encompasses the Triangular and uniform as special cases and permits, but does not require, "fat tails". For lack of a better term, this generalization may be termed the "Rocket" or "Uniform-Triangular" distribution.

THE UNIFORM-TRIANGULAR DISTRIBUTION

This distribution combines a triangular distribution with one or more uniform distributions. A rationale for such a distribution is that it permits fat tails which the triangular distribution does not. Also, like the triangular distribution, it permits the user to exploit knowledge of range; knowledge which is often known quite well in at least a subjective sense (see Figure III.1).

When this distribution is symmetric, it has a rocket-like shape consisting of an equilateral triangle on top of a rectangle. Special cases include the triangular (when $h_2=h_3=0$) and the uniform (when $B_1=B_2=0$ and $h_1=h_2$).

Density Function:

$$\begin{aligned}
 f(X) &= f_1(X) = h_2 - B_1 * (X - a); & a \leq X \leq c \\
 &= f_2(X) = h_1 - B_2 * (X - c); & c \leq X \leq b \\
 &= 0 & \text{otherwise.}
 \end{aligned}$$

Where:

$$\begin{aligned}
 a &= \text{minimum } X \\
 b &= \text{maximum } X \\
 c &= \text{modal } x \\
 f(c) &= \text{modal frequency} \\
 h_1 &= 2 / [b - a + d_2 * (c - a) + d_3 * (b - c)] \\
 B_1 &= (h_1 - h_2) / (c - a) \\
 B_2 &= (h_1 - h_3) / (b - c) \\
 h_2 &= d_2 * h_1; & 0 \leq d_2 \leq 1 \\
 h_3 &= d_3 * h_1; & 0 \leq d_3 \leq 1
 \end{aligned}$$

The parameters d_2 , d_3 define the height of the uniform distribution(s) relative to the modal frequency, h_1 which must be solved as a constant of integration.

Distribution Function:

$$\begin{aligned}
 F(X) &= F_1(X) = (X - a) * [h_2 + 0.5 * B_1 * (X - a)]; & a \leq X \leq c \\
 &= F_2(X) = (X - c) * [h_1 - 0.5 * B_2 * (X - c)]; & c \leq X \leq b
 \end{aligned}$$

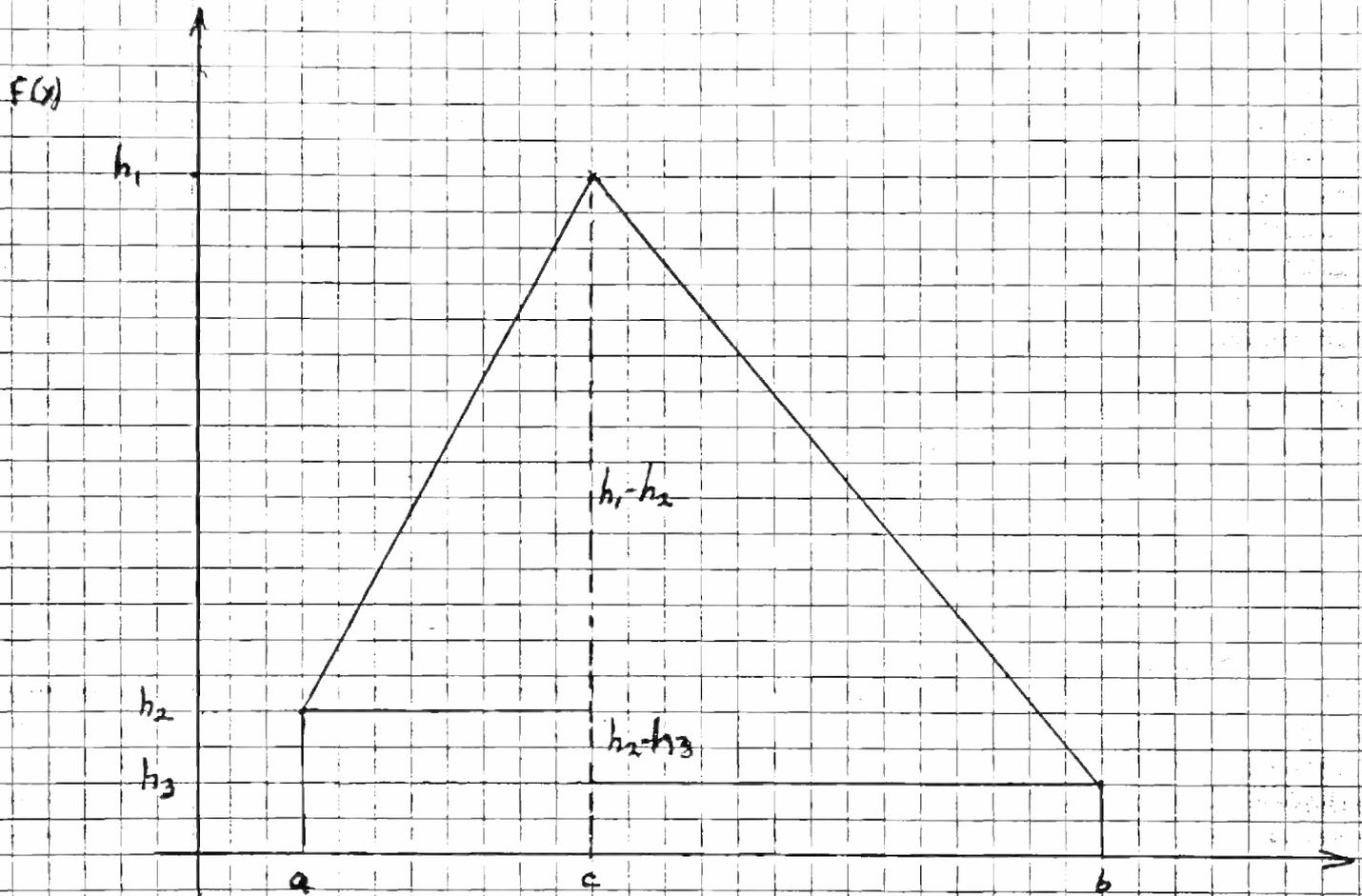
Inverse Distribution Function:

$$\begin{aligned}
 G(y) &= G_1(y) = (1/B_1) * [a * B_1 - h_2 + \text{SQRT}[h_2 * h_2 + 2 * B_1 * y]] \\
 &\quad \text{for } 0 \leq y \leq F_1(c) \\
 &= G_2(y) = (1/B_2) * [h_1 + B_2 * c - \text{SQRT}[h_1 * h_1 + ((c - a) / (b - c)) * \\
 &\quad (h_1 - h_3) * (h_1 + h_2) - 2 * B_2 * y]]; & F_1(c) \leq y \leq 1
 \end{aligned}$$

Examination of the equations for this distribution reveals how the triangular and uniform distributions emerge

Figure III.1: Uniform-Triangular

III.2



as special cases. If $d_2=d_3=1$, then $h_1=1/(b-a)$, $B_1=B_2=0$ and the distribution is uniform. However, the inverse distribution function as expressed would then contain a singularity and must be modified. Since uniform distributions are available in computers anyway, the correction seems pointless. If $d_2=d_3=0$, then the distribution is triangular and $h_1=2/(b-a)$.

THE V DISTRIBUTION

This distribution has the following parameters:

a= minimum X
b= maximum X
c= notch point (minimum frequency)
d=h₂/h₁ = modal frequency ratio at a vs. b
 at a: f₁(a)=h₁
 b: f₂(b)=h₂
h₃=f₁(c)=f₂(c)= notch point frequency.

Due to the bimodal nature of this distribution, one must specify one of the three parameters h₁, h₂ and d; I have chosen to specify d as the more useful specification since it is a relative concept; h₁ and h₂ must then satisfy integration condition for probability. The notch point frequency f₁(c)=f₂(c)=h₃→0 allows for the possibility that the minimum frequency point may have a non-zero frequency. If h₃>0, the distribution appears geometrically like a "V" on top of a uniform (a,b) distribution with the notch located on a vertical line intersecting c. If c is the mid point of a and b and if d=1 then the notch or "V" is symmetric about c.

Density Function:

$$\begin{aligned} f(X) &= f_1(X) = h_1 - B_1 * (X - a); & a \leq X < c \\ &= f_2(X) = h_3 + B_2 * (X - c); & c \leq X \leq b \\ &= 0 & \text{otherwise} \end{aligned}$$

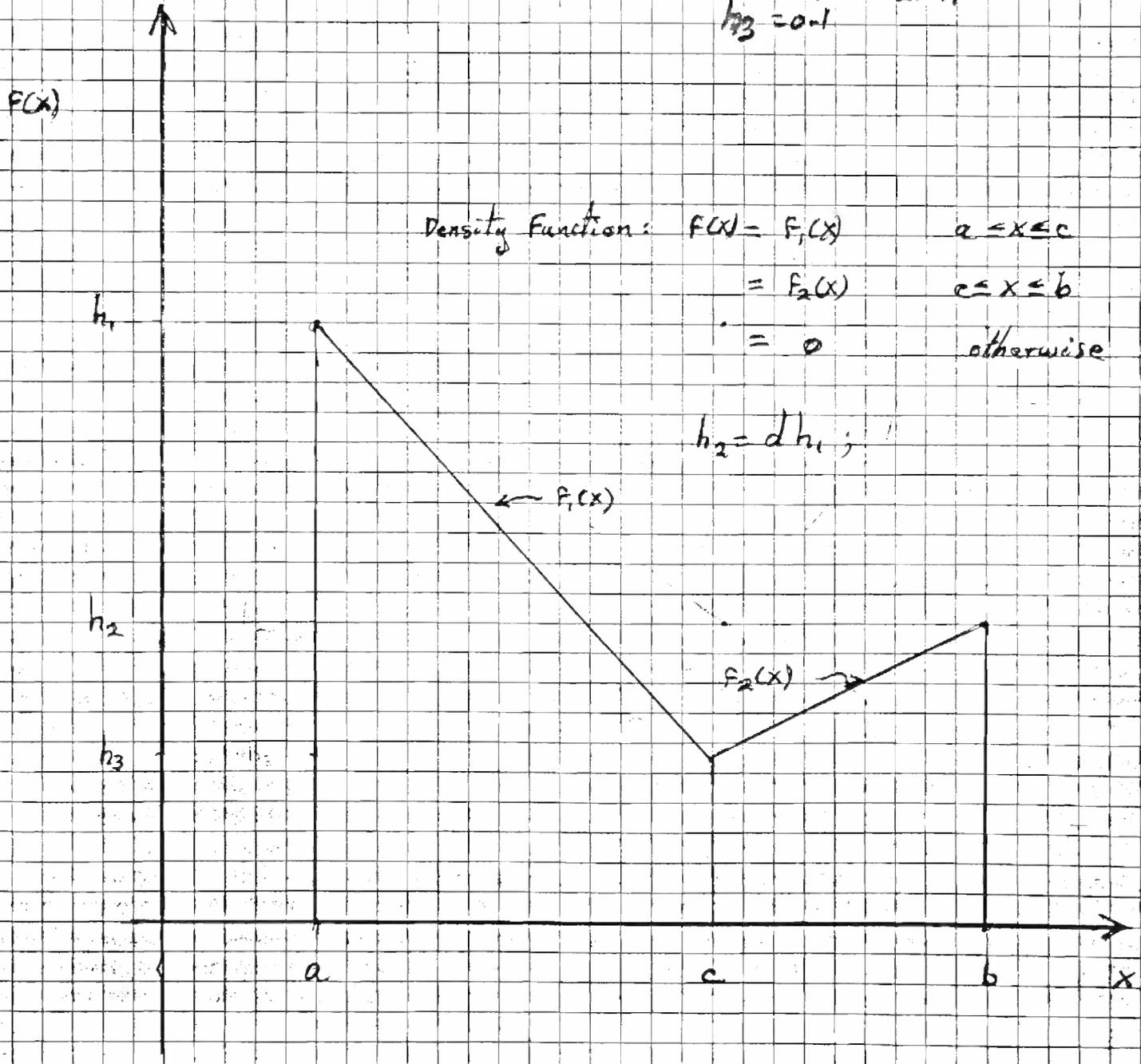
Where:

$$B_1 = (h_1 - h_3) / (c - a); \quad B_2 = (d * h_1 - h_3) / (b - c) = (h_2 - h_3) / (b - c)$$

The parameters c and d jointly partition the probability mass between the two tails of this bimodal distribution (see Figure IV.1). This is an unusual distribution which, in the limit, mimics a world with two unimodal, mutually exclusive triangular distributions. The limiting case of mutual exclusion is reached when h₃=0; i.e. if the "notch point" touches the X axis at point c. A slightly more general case would be that of a notch whose nadir is an interval rather than a point on the X axis. Of course, once one admits one "hole" or empty interval, the stage is set for generalization to multiple holes. Generalization to (for example) a "saw-tooth" distribution is easy but requires proportionately more parameters and hence more a priori or objective knowledge of the application.

Figure IV.1 : The V Distribution

Example Parameters : $a = 5$
 $b = 10$
 $c = 8$
 $h_1 = 0.375 = \text{constant of integration}$
 $d = 0.5 = h_2/h_1$
 $h_2 = 0.1$



Density Function: $f(x) = f_1(x) \quad a \leq x \leq c$
 $= f_2(x) \quad c \leq x \leq b$
 $= 0 \quad \text{otherwise}$

$$h_2 = d h_1;$$

$f_1(x)$

$f_2(x)$

Distribution Function:

$$\begin{aligned} F(X) = F_1(X) &= (X-a) * [h_1 - 0.5 * B_1 * (X-a)] ; \quad a \leq X \leq c \\ &= F_2(X) = (X-c) * [h_3 + 0.5 * B_2 * (X-c)] ; \quad c \leq X \leq b \\ &= 0 \text{ otherwise.} \end{aligned}$$

Inverse Distribution Function:

$$\begin{aligned} G(y) = G_1(y) &= (1/B_1) * [h_1 + aB_1 - \text{SQRT}[h_1 * h_1 - 2 * B_1 * y]] ; \\ &\quad \text{for } 0 \leq y \leq F_1(c) \\ &= G_1(F_1(c)) + (1/B_2) * [B_2 * c - h_3 + \text{SQRT}[h_3 * h_3 + 2 * B_2 * y]] ; \\ &\quad \text{for } F_1(c) \leq y \leq 1 \end{aligned}$$

REFERENCES:

Budinick, Frank S., Richard Mojena and Thomas E. Vollman. 1977. Principles of Operations Research for Management (Homewood, Illinois:Richard D. Irwin, Inc.) 756 p.

Hastings, N.A.J. and J.B. Peacock. 1974. Statistical Distributions (New York:John Wiley and Sons) 130 p.