SOME IMPLICATIONS OF UNCERTAINTY IN PERIODIC-REVIEW INVENTORY THEORY

- o O o -

DAVID ARNOLD LESLIE WILSON

Thesis submitted for the Degree of Ph.D.

University of Aston
in Birmingham.

April 1971

SUMMARY

The first part of this thesis is concerned with finding optimal ordering policies for multistage inventory systems in which stocks are reviewed periodically. A new exact method of calculating critical stock levels is developed in order to avoid difficulties in the classical dynamic programming approach. The equations developed for the new method also provide a clearer picture of the way in which optimal policies for periodic-review processes of finite duration converge to those for processes of infinite duration. Models with leadtime are then considered in a way which leads to a new generalisation of earlier work on stochastic leadtimes. Consideration of adaptive inventory systems leads on to the second part of the thesis, where the emphasis is on Bayesian methods of treating imperfectly known demand distributions. Important original results are obtained when the adaptive behaviour of several well known distributions is considered.

INDEX

SUMMARY i

INDEX ii

ACKNOWLEDGEMENTS iv

CHAPTER 1. INTRODUCTION

1-1. Storage and inventory 1

1-2. Inventory models with periodic review 5

1-3. Plan of this thesis 7

CHAPTER 2. PERIODIC-REVIEW INVENTORY MODELS

2-1. The single-stage model 9

2-2. The multistage model 15

2-3. Other cost structures 25

2-4. Demand distributions and convexity 33

CHAPTER 3. MODELS WITH MORE UNCERTAINTY

3-1. Known leadtimes 35

3-2. Stochastic leadtimes 41

3-3. Demand distributions which vary with time 47

CHAPTER 4. THE BASIC DISTRIBUTION THEORY

4-1. Pólya functions and total positivity 50

4-2. The Exponential Family and the Range Family 55

4-3. Convolutions, mixtures and Bayesian methods 57

4-4. Conjugate distributions 62

CHAPTER 5. THE POISSON AND GAMMA DISTRIBUTIONS

5-1. Definitions and inventory applications 67

5-2. Convolutions and Bayesian treatment of Poisson distributions 71

5-3. Bayesian treatment of Gamma distributions 74

CHAPTER 6. THE NORMAL AND LOGNORMAL DISTRIBUTIONS

6-1. Definitions and inventory applications 82

6-2. The conjugate distributions 84

6-3. Bayesian theory for Normal distributions 87

CHAPTER 7. THE BINOMIAL AND BETA DISTRIBUTIONS

7-1. Definitions and inventory applications 90

7-2. Bayesian theory for Binomial distributions 93

7-3. Bayesian theory for Beta distributions 95

CHAPTER 8. THE UNIFORM DISTRIBUTIONS

8-1. Definitions and inventory applications 99

8-2. The Rectangular distribution 101

8-3. The Uniform discrete distribution 104

CHAPTER 9. SURVEY AND CONCLUSION

9-1. Selection from a family of conjugate distributions 106

9-2. Further work 109

9-3. Conclusion 111

APPENDIX 1. CONVERGENCE OF CRITICAL STOCK LEVELS 113

APPENDIX 2. THE EPSILON AND TRUNCATED ZETA FUNCTIONS 117

BIBLIOGRAPHY 122

ACKNOWLEDGEMENTS

Many of my colleagues at Aston have helped me build up my present view of inventory theory, in particular I would like to thank my supervisor, Dr.A.M.Rajput, for his encouragement during this long task. Lecturing on inventory theory has helped me to clarify many ideas on the subject and some interesting numerical work by one of my students, Mrs. Gillian Baynes, is referred to in an appendix at the end of this thesis. Finally, I would like to acknowledge my debt to Mr. William Lampkin for his help during my early work on inventory theory at Imperial College.

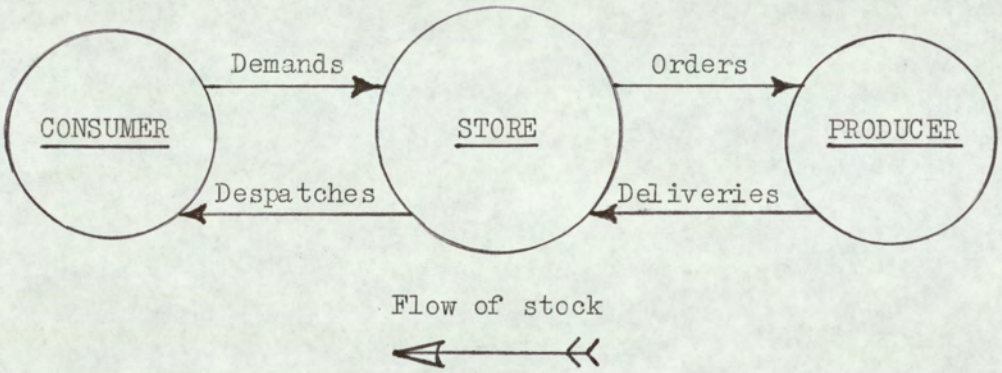
D.A.L.Wilson.

CHAPTER 1

INTRODUCTION

1-1. Storage and inventory

During the last sixty years many writers have worked on mathematical models to describe the operation of systems in which goods are held in store as an intermediate stage between production and consumption. We may picture the general storage model as follows:



The system consists of three levels - consumer, store and producer - with the store smoothing out differences in the rates of production and consumption. We look at the system from the point of view of the store and its operation. Requests for stock from consumer to store are called demands, those from store to producer are called orders. Shipments of stock from producer to store are called deliveries, those from store to consumer are despatches.

In most practical systems of this kind, there is a constraint that stock must flow only from producer to consumer. The implication of this one-way flow is that the quantities of stock involved in demands, despatches, orders and deliveries will all be non-negative.

This general class of models is usually sub-divided according to the relationship between the store and the producer. If deliveries are not controllable by orders from the store, we obtain a class of models typified by the storage of rainwater which is covered by the theory of dams - see Moran (1959) and Prabhu (1965). If deliveries are controlled, albeit imperfectly, we obtain the class of inventory models. Typical situations to which inventory models are applicable include the operation of an ordinary retail shop, a wholesale warehouse (where the consumer represents the aggregate of retailers), a spare-parts store in a workshop (where machines consume spare parts as they break down and get repaired) and so on - see Hadley and Whitin (1963).

A major part of the theory of inventory systems is concerned with finding optimal rules for ordering and the criterion of optimality is usually taken to be minimum cost or, equivalently, maximum profit. We develop this approach by examining a very simple deterministic model of the type proposed by Harris (1915).

Imagine that we are interested in stocking a single commodity for which demand occurs at a uniform known rate of λ units per unit time. Our store is empty initially but we can order z units of stock for immediate delivery. The cost of purchasing this stock

is $y_p > 0$ for any $z > 0$ plus $c_p > 0$ for each unit, making the total purchasing cost $y_p + c_p z$ for any non-negative z which satisfies the one-way flow condition. We also impose a holding cost of k_H per unit stock per unit time held in store and allow time to vary continuously. Assuming that only one initial order may be placed and that the process finishes when no stock remains, the duration of the process will be $T = z/\lambda$ and the total cost will be

$$(1\text{-}1.1) \qquad C(z) = y_p + c_p z + k_H \int_0^T (z - \lambda t)\,dt$$

$$= y_p + c_p z + k_H \frac{z^2}{2\lambda} .$$

A trivial solution for minimising C would be $z = 0$, but this would eliminate the need for a store and mean that no demand would be satisfied. Treating z as continuous and differentiating C to locate a minimum, we obtain the solution

$$(1\text{-}1.2) \qquad z = -\lambda c_p / k_H ,$$

which is not feasible since z must be non-negative.

However, a feasible solution can be found if we make our objective function not total cost $C(z)$ but total cost per unit time

$$(1\text{-}1.3) \qquad K(z) = \frac{C(z)}{T} = \frac{z}{\lambda} C(z).$$

Treating z as continuous and differentiating $K(z)$ we find the minimum to be at $z = z^*$ where

$$(1\text{-}1.4) \qquad z^* = \sqrt{\frac{2\gamma_p \lambda}{k_H}} \quad .$$

This is the famous square-root formula of early inventory theory - often known as the formula for "economic ordering quantity (EOQ)" or "the Wilson formula" after R.H.Wilson, whose work followed that of Harris.

If z is constrained to take integer values, differencing shows that z^*, the optimal value, is the least z for which

$$(1\text{-}1.5) \qquad z(z+1) \geqslant \frac{2\gamma_p \lambda}{k_H} \quad .$$

By allowing another order to be placed as soon as the stock level falls to zero we can build up a multistage model in which each stage is identical to the single stage of duration T which we have just considered. If we have N such stages, all complete, the total cost per unit time will still be K(z) and the optimal order quantity for each stage will be z^*: this will be true for finite or infinite N. If the last stage is incomplete, the contribution of this last stage to the total cost per unit time will become negligible as the total duration of the process tends to infinity, and in this case also the square-root formula - or (1-1.5) - will give the optimal order quantity for each stage.

1-2. Inventory models with periodic review

The simple deterministic model considered in the previous section provides a useful starting point for the development of more complicated models, but as soon as we consider the demand process to be stochastic rather than deterministic we are faced with a further sub-division of the class of models. On the one hand we have models in which the stock position may be reviewed at any time and orders may be placed when necessary: we call this a continuous-review (CR) system. On the other hand we have models in which the review and ordering decision can only be made at certain fixed times: this is a periodic-review (PR) system.

The CR system is applicable in situations involving high-value goods with a low demand rate, but such a system is costly to administer when many commodities are stocked. In practice, therefore, it is usually preferable to sacrifice the closer control afforded by CR to the greater convenience of PR with the period between reviews chosen to fit in with the planning and accounting periods used by the imdustry concerned. However, much the greatest part of the literature has been devoted to CR systems and in the United Kingdom very little attention has been paid to developments in PR theory - see Iglehart's comments in the discussion on a paper by Thatcher (1962) and the bibliography of Eilon and Lampkin (1968).

The general theory of PR systems stems from a model proposed by Arrow, Harris and Marshak (1951) - often called the AHM model. Most of the development of this basic model has been done by workers at Stanford University headed by Samuel Karlin and Herbert Scarf

and the classical exposition of the basic theory appeared in a collection of papers by Arrow, Karlin and Scarf (1958). Another important early work was a book by Whitin (1953), whose later researches are summarised in Hadley and Whitin (1963), which remains the best general textbook on inventory theory.

An important tool in the analysis of PR systems is the technique of dynamic programming developed by Bellman (1957,1961), who has proposed the following hierarchy of multistage decision processes.

LEVEL 0 : Deterministic processes.

LEVEL 1 : Stochastic processes with completely known probability distributions.

LEVEL 2 : Adaptive processes with imperfectly known probability distributions. The increasing amount of information gained by observing the process is incorporated into the description of the system at each stage.

Two further levels of uncertainty have since been proposed by White (1962), but it is Bellman's Levels 1 and 2 which are of most interest in PR inventory theory and it is through models at Level 2 that inventory theory ties in with the Bayesian methods of modern statistical decision theory. These methods are still slightly controversial but seem to have gained ground since the publication of an important book by Raiffa and Schlaifer (1961).

1-3. Plan of this thesis

In Chapter 2 we begin by examining a single-stage stochastic (Level 1) inventory model with a simple linear cost structure. This leads on to multistage PR models and we also show how some superficially different cost structures can be considered as equivalent to the linear structure.

We examine the practical computation of optimal stock levels and find that the dynamic programming scheme suggested by previous authors poses a number of problems which we resolve by developing a new scheme based on considering derivatives or differences. The equations used in the improved scheme provide a new insight into the connection between models with a finite number of stages and those with an infinite number. Since the infinite case is soluble without any recursive computation it is of interest to find how well it approximates the finite case. Some quantitative work on this topic is shown in Appendix 1.

In Chapter 3 we introduce the idea of leadtime, or delay between ordering and delivery, and then study the effect of deterministic and stochastic leadtimes upon ordering policy (the distributions of the stochastic leadtimes being treated as completely known). An important original contribution is the extension of the theory of stochastic leadtimes to the case where more than one order may be outstanding at any time. At the end of the chapter we touch on further generalisations of the model, including the consideration of imperfectly known demand distributions, which leads on to models at Bellman's Level 2 and to Bayesian theory.

Chapter 4 brings together the necessary basic ideas of distribution theory, many of which are not yet common currency. We give a brief introduction to the theory of totally positive kernels in order to provide some explanation of requirements for demand distributions to belong to certain classes. This is followed by a general description of the Bayesian methods used in the next four chapters.

Chapters 5,6,7 and 8 examine in detail the adaptive behaviour of some well known distributions which might be used to describe demand (and/or leadtime). We find that Bayesian methods for dealing with some of these distributions involve completely new distributions based upon functions which are here defined for the first time. Two of the new functions are tabulated in Appendix 2.

The final chapter surveys what has been achieved in the thesis, discusses some broader applications of the distribution theory and points the way for further work ahead.

CHAPTER 2

PERIODIC-REVIEW INVENTORY MODELS

2-1. The single-stage model

To develop the notation required for the more general treatment of multistage models which follows in later sections, we first consider a process consisting of one stage of known duration T. At the start of the process (t=o) we place a single order for immediate delivery of a suitable quantity of some commodity which becomes obsolete at time t=T. The model is therefore applicable to the stocking of perishables which can not be reordered: newspapers or Christmas trees are often quoted as typical commodities of this kind. We can relax the assumption of immediate delivery if the time origin (t=o) is taken to be the time of delivery rather than that of placing the order.

Total demand in time T is considered to be a non-negative random variable, R, which has a distribution function (d.f.), denoted by F, such that

(2-1.1) $F(r) = P[R \leq r]$, where P[A] means the probability of any event A.

The non-negativity constraint is imposed to ensure one-way flow through the inventory system.

We shall denote the mean demand during the stage by μ, which is constant with respect to time.

When R is discrete, it will have a probability mass function (p.m.f.) denoted by p, such that

(2-1.2) $$p(r) = P[R = r] \text{ and hence } F(y) = \sum_{r=o}^{y} p(r).$$

Unless stated otherwise, we shall assume that any discrete distributions which we use are defined on zero and the positive integers. For some discrete distributions it will be convenient to use the right-hand cumulative function, which we shall call P(.) with round brackets to avoid any confusion with P[.], the probability of some event.

$$(2\text{-}1.3) \qquad P(y) = \sum_{r=y}^{\infty} p(r) = 1 - F(y-1)$$

When R is continuous we shall denote the corresponding probability density function (p.d.f.) by f, such that

$$(2\text{-}1.4) \qquad f(r) = \frac{d}{dr} F(r) \quad \text{and hence} \quad F(y) = \int_{o}^{y} f(r)dr$$

We shall choose our order quantity to satisfy the criterion of minimum total expected cost. To set up the required objective function we need to impose a cost structure, and we shall consider the simple case with three positive linear unit costs:

c_p = purchasing cost per unit quantity,
so z units bought cost $c_p z$;

c_H = holding cost per unit positive stock remaining at time T,
so z units surplus cost $c_H z$;

c_R = runout cost per unit shortage or runout at time T,
so z units short cost $c_R z$.

If we have an <u>initial stock</u> of x units available before ordering, the problem of choosing an <u>order quantity</u>, which we shall usually call z, is equivalent to choosing a <u>target stock</u>, y, which we aim to have after delivery. Because of our one-way flow condition, the only feasible target stocks are those not less than the initial

stock so we shall be obliged to choose some $y \geq x$. Of course, in most practical cases $x = 0$ in a single-stage process of this kind, and we would consider possible target stocks $y > 0$.

The total expected cost of the process, say C_1, will depend on both x and y and may be written

$$C_1(x,y) = c_p(y-x) + L(y), \tag{2-1.5}$$

where the term $L(y)$ represents the expected combined cost of holding and runout (i.e. surplus and shortage). The function L will be called the loss function and we define it for $y > 0$ to be

$$L(y) = \begin{cases} c_H \sum_{r=0}^{y} (y-r)p(r) + c_R \sum_{r=y+1}^{\infty} (r-y)p(r) & \text{for discrete demand,} \\ c_H \int_0^y (y-r)f(r)dr + c_R \int_y^{\infty} (r-y)f(r)dr & \text{for continuous demand.} \end{cases} \tag{2-1.6}$$

If there is some finite maximum value for demand, say $r_{max} > y$, then the top limit on the runout terms in (2-1.6) will be r_{max} instead of ∞.

To simplify later calculations, it is convenient to rearrange the expressions on the right-hand side of (2-1.6) to give the following equivalent forms.

In the discrete case:

$$\begin{aligned} L(y) &= c_R(\mu-y) + (c_H+c_R) \sum_{r=0}^{y-1} F(r) \\ &= c_H(y-\mu) - (c_H+c_R) \sum_{r=y+1}^{\infty} P(r), \end{aligned} \tag{2-1.7}$$

where the top limit in the last sum may be r_{max} instead of ∞.

In the continuous case:

$$(2\text{-}1.8) \qquad L(y) = c_R(\mu - y) + (c_H + c_R)\int_0^y F(r)dr.$$

Although $C_1(x,y)$ depends on both x and y, the value of x is predetermined and the minimising value of y, in either the discrete or the continuous case, will be that which minimises

$$(2\text{-}1.9) \qquad c_p y + L(y) = \Lambda_1(y), \text{ say.}$$

In the discrete case we locate the minimum by differencing.

$$(2\text{-}1.10) \qquad \Delta\Lambda_1(y) = \Lambda_1(y+1) - \Lambda_1(y)$$
$$= c_p + \Delta L(y).$$

From (2-1.7) we can immediately obtain the differenced loss function in terms of F or P.

$$(2\text{-}1.11) \qquad \Delta L(y) = -c_R + (c_H + c_R)F(y)$$
$$= c_H - (c_H + c_R)P(y+1).$$

The minimum of $\Lambda_1(y)$ lies at y_1^*, the least value of y for which $\Delta\Lambda_1(y) \geq 0$, i.e. y_1^* is the least value for which

$$(2\text{-}1.12) \qquad F(y) \geq \frac{c_R - c_p}{c_H + c_R} \quad \text{or} \quad P(y+1) \leq \frac{c_H + c_p}{c_H + c_R}.$$

In the continuous case we locate the optimal value of y by considering the derivative of $\Lambda_1(y)$.

$$(2\text{-}1.13) \qquad \Lambda_1'(y) = c_p + L'(y).$$

where the prime on a function denotes the first derivative with respect to the argument - here y. From (2-1.7) we get

$$(2\text{-}1.14) \qquad L'(y) = -c_R + (c_H + c_R)F(y).$$

The second derivative will be $(c_{H+}c_R)f(y)$ in which all the terms are positive and so the minimum of $\Lambda_1(y)$ lies at $y = y_1^*$ such that

$$(2\text{-}1.15) \qquad F(y_1^*) = \frac{c_R - c_p}{c_H + c_R}$$

The optimal ordering policy with discrete or continuous demand will thus be:

$$\text{order } z_1^* = y_1^* - x \quad \text{if} \quad x < y_1^*$$

$$\text{and order nothing if} \quad x \geqslant y_1^* .$$

The total expected cost with this policy will therefore be

$$(2\text{-}1.16) \qquad C_1(x,y_1^*) = - c_p x + \Lambda_1(y_1^*) \text{ if } x < y_1^*,$$

$$(2\text{-}1.17) \qquad C_1(x,x) = - c_p x + \Lambda_1(x) \text{ if } x \geqslant y_1^*.$$

We can combine (2-1.16) and (2-1.17) by defining the function C_1^*, the total expected cost with optimal policy, which will be a function of a single variable, the initial stock.

$$(2\text{-}1.18) \qquad C_1^*(x) = \min_{y \geqslant x} \{-c_p x + \Lambda_1(y)\}$$

$$= -c_p x + \min_{y \geqslant x} \{\Lambda_1(y)\} .$$

Returning for a moment to (2-1.12) and (2-1.15), and remembering that since F is a d.f. we must have $0 \leqslant F(y) \leqslant 1$, it follows that there are mild constraints upon the unit costs. Firstly:

$$(2\text{-}1.19) \qquad c_R \geqslant c_p,$$

which is intuitively reasonable because we could not operate the inventory system profitably if the unit cost of shortage were less than the purchasing cost. Secondly, if we relax the condition that all the

unit costs are positive, when we set $c_p = 0$ we must have

(2-1.20) $\qquad c_H \geq 0$.

The net holding cost for a perishable commodity may be the difference between, on the one hand, a positive charge for holding which represents a loss of revenue from the capital tied up in the stock (Hadley and Whitin (1963) usually consider holding costs which are effectively of this type) plus a proportion of the expenses of running the store and, on the other hand, a resale value for the surplus stock after obsolescence. If the resale value exceeds the positive cost, then c_H will be negative. However, in any sensible system $c_H > -c_P$ and usually c_H is some small positive fraction of c_P, so that (2-1.12) and (2-1.15) give feasible results for y_1^*. Because c_H is usually small, the key to these results is the ratio of c_R to c_P.

If c_R is not known then our criterion of minimum total expected cost, resulting from purchasing, holding and runout, can not be used to find y_1^*; instead, a level of service (or minimum probability of meeting all demands from stock) may be specified - call this π. Then, in the discrete case, y_1^* is the least value of y for which

(2-1.21) $\qquad F(y) \geq \pi$.

Comparing this with (2-1.12), we see that this is equivalent to setting

(2-1.22) $\qquad \pi = \dfrac{c_R - c_P}{c_H + c_R} \quad \text{or} \quad c_R = \dfrac{c_P + \pi c_H}{1 - \pi}$.

The equivalence between other types of runout costs and differently defined levels of service is demonstrated by Hadley and Whitin (1963), who show that the runout cost can be considered as a Lagrange multiplier in the equation for the expected combined cost of purchasing and holding subject to the constraint of a suitably defined level of service; this effectively converts the combined cost equation into our original total cost

equation. We shall therefore, continue to pursue the total cost approach, but it can now be seen that our results will also be applicable to other problems because of the equivalence described here.

2-2. The multistage model

Continuing with the same linear cost structure as in (2-1), we now consider the case where stock remaining at the end of the first stage of a process is not obsolete but can serve as the initial stock for a second stage. The residual stock serves as initial stock for a third stage, and so on for a specified number of stages. At the end of the process the residual stock finally becomes obsolete, but until then no account is taken of the age of any unit of stock.

We shall assume that there are N stages (N a known positive integer) and that each stage is of length T. The stock position is reviewed at $t = o$ and then again at the beginning of each subsequent stage, $t = T, 2T, \ldots, (N-1)T$. If the review shows that the initial stock at any stage is too low then a suitable quantity is ordered for immediate delivery. The demand in each stage is distributed independently and identically with discrete or continuous d.f. F as in the single-stage model.

We consider first the case where $N = 2$. If we denote the total expected cost of the two-stage process by C_2, we see that this will depend on the initial stock (say x), the target stock when there are two uncompleted stages (say y_2) and the target stock when there is one uncompleted stage (say y_1). Using C_1 in the same sense as in the previous section, to represent the total expected cost of the uncompleted last stage alone, we can write the following equation for

the case of continuous demand.

$$(2\text{-}2.1) \qquad C_2(x,y_2,y_1) = c_p(y_2-x) + L(y_2) + \int_0^\infty C_1(y_2-r,\, y_1)f(r)dr.$$

The first two terms on the right-hand side represent the expected cost of the initial stage and the last term represents the expected cost of the rest of the process, i.e. the last stage alone, when it opens with the residual stock (y_2-r) from the first stage.

In considering the top limit for r in the integral to be ∞ (or $r_{max}>y_2$) we allow the initial stock for the last stage to be positive or negative. A negative initial stock implies that excess demands in the first stage have been held captive and form a backlog which is to be satisfied by the delivery of stock in the second stage before the second-stage demand commences.

If we assume that excess demands represent so much lost sales and cannot be backlogged, then (2-2.1) must be replaced by

$$(2\text{-}2.2) \quad C_2(x,y_2,y_1) = c_p(y_2-x) + L(y_2) + \int_0^{y_2} C_1(y_2-r,\, y_1)f(r)dr + C_1(0,y_1)\int_y^\infty f(r)dr.$$

Bellman (1957) and Karlin (1958b) examine the lost-sales case in detail, but we shall confine our attention to the backlog case and proceed on from (2-2.1).

We require to find values of y_1 and y_2 which minimise $C_2(x,y_2,y_1)$. Now y_1 only occurs in the last term as the second argument of C_1, and we have already seen that the optimal policy for a single stage results in C_1 becoming C_1^*. Thus if we use y_1^* given by (2-1.15) we can then write the total expected cost of the two-stage

process with optimal final stage as C_2^+, a function of two variables only.

$$(2\text{-}2.3) \qquad C_2^+(x,y_2) = c_p(y_2-x) + L(y) + \int_0^\infty C_1^*(y_2-r)f(r)dr.$$

Then proceeding as in (2-1.18), we can reduce the left-hand side to a function of one variable only, the initial stock x, by defining C_2^*, the total expected cost of a two-stage process with optimal policy in each stage, by the following equation.

$$(2\text{-}2.4) \qquad C_2^*(x) = \min_{y \geq x} \{C_2^+(x,y)\}$$

$$= -c_p x + \min_{y \geq x} \{\Lambda_2(y)\},$$

where Λ_2 is defined by

$$(2\text{-}2.5) \qquad \Lambda_2(y) = c_p y + L(y) + \int_0^\infty C_1^*(y-r)f(r)dr.$$

We then find the optimal value of y_2, say y_2^*, by locating the minimum of Λ_2. For the moment we shall assume that this is a unique minimum and return later to examine the conditions which ensure this.

As in the single-stage model, we order up to the optimal target stock if the initial stock is below this value, otherwise we order nothing, so that we can write (2-2.4) as

$$(2\text{-}2.6) \qquad C_2^*(x) = \begin{cases} -c_p x + \Lambda_2(y_2^*) & \text{if } x < y_2^*, \\ -c_p x + \Lambda_2(x) & \text{if } x \geq y_2^*. \end{cases}$$

Proceeding in the same way, if we define C_n^* as the total expected cost with optimal policy over n remaining stages, we shall get the following general recurrence relationship for $n = 2,3,\ldots,N$.

$$(2\text{-}2.7) \qquad C_n^*(x) = \min_{y \geq x} \{c_p(y-x) + L(y) + \int_0^\infty C_{n-1}^*(y-r)f(r)dr\}.$$

When n=2, as we have already seen, we can evaluate the last term by using (2-1.18).

If we define Λ_n for n=2,3,...,N by

$$(2\text{-}2.8)\quad \Lambda_n(y) = c_p y + L(y) + \int_0^\infty C^*_{n-1}(y-r)f(r)dr,$$

then we can find the optimal target stock when n stages remain, say y^*_n, and then rewrite (2-2.7) as

$$(2\text{-}2.9)\quad C^*_n(x) = -c_p x + \min_{y \geq x} \{\Lambda_n(y)\}$$

$$= \begin{cases} -c_p x + \Lambda_n(y^*_n) & \text{if } x < y^*_n, \\ -c_p x + \Lambda_n(x) & \text{if } x \geq y^*_n. \end{cases}$$

In the foregoing analysis we have given equal weight to present and future costs at any stage. Since money required to meet future costs can be usefully employed until settlement is required, it is sensible to modify Λ_n by introducing a <u>discount factor</u> α (taking some value between 0 and 1) by which we reduce costs due for settlement T time units hence. Equation (2-2.8) will then be replaced by:

$$(2\text{-}2.10)\quad \Lambda_n(y) = c_p y + L(y) + \alpha \int_0^\infty C^*_{n-1}(y-r)f(r)dr \text{ for } n = 2,3,\ldots,N.$$

Whether we use (2-2.8) or (2-2.10) we could obtain a sequence of optimal target stocks $\{y^*_1, y^*_2, \ldots, y^*_N\}$, which determines our optimal policy at each stage, as a result of the following scheme of calculation.

Scheme A

(1) Calculate y_1^* by equating $F(y_1^*)$ to the cost ratio in (2-1.18)

(2) Set $n = 2$

(3) Evaluate $C_{n-1}^*(x)$ at a sufficient number of positive and negative values of x to allow the integral in the last term of $\Lambda_n(y)$ to be evaluated by a suitable numerical method in Step (4).

(4) Keep n equal to the value used in Step (3) and calculate $\Lambda_n(y)$ at enough values of y to be able to locate the minimum by interpolation, and so find y_n^*.

(5) If $n = N$ then stop, otherwise let n become $n + 1$ and go to Step (3).

The main practical difficulty of this scheme of backwards dynamic programming lies in knowing what values provide suitable grids for evaluating $C_{n-1}^*(x)$ and $\Lambda_n(y)$ in Steps (3) and (4) - a difficulty which Bellman (1961) picturesquely calls "the menace of the expanding grid". Some help is provided by Karlin (1958b), who has shown that

$$y_1^* < y_2^* \cdots < y_n^*,$$

a result which we shall verify below.

Even with discrete demands, where (2-2.9) is still applicable if we replace (2-2.10) by the corresponding discrete function,

$$\Lambda_n(y) = c_p y + L(y) + \alpha \sum_{r=0}^{\infty} C_{n-1}^*(y-r)p(r), \qquad (2\text{-}2.11)$$

the problem of the grids is still considerable if very many discrete values of demand have an appreciable probability.

To arrive at a more efficient scheme, we first examine what happens when the duration of the process becomes infinite, i.e. when $N \to \infty$. In this case, whatever stage the process has reached there remains an infinite number of further stages. If we define C^* as the expected total discounted cost of an infinite number of uncompleted stages with optimal policy, we get a functional equation which corresponds to (2-2.7).

$$(2\text{-}2.12) \quad C^*(x) = - c_P x + \min_{y \geq x} \{c_P y + L(y) + \alpha \int_0^\infty C^*(y-r) f(r) dr\}.$$

We can find y^*, the optimal target stock for any stage of the infinite-duration process, by locating the minimum (assumed to be unique) of Λ, defined by:

$$(2\text{-}2.13) \quad \Lambda(y) = c_P y + L(y) + \alpha \int_0^\infty C^*(y-r) f(r) dr.$$

The total cost in (2-2.12) can be written as

$$(2\text{-}2.14) \quad C^*(x) = \begin{cases} -c_P x + \Lambda(y^*) & \text{if } x < y^*, \\ -c_P x + \Lambda(x) & \text{if } x \geq y^*. \end{cases}$$

Differentiating this with respect to the argument x, we get

$$(2\text{-}2.15) \quad \frac{d}{dx} C^*(x) = \begin{cases} -c_P + 0 & \text{if } x < y^*, \\ -c_P + \Lambda'(x) & \text{if } x \geq y^*. \end{cases}$$

Using (2-2.15) when we differentiate $\Lambda(y)$ to locate the minimum, we get

$$(2\text{-}2.16) \quad \Lambda'(y) = c_P(1-\alpha) + L'(y) + \alpha \int_0^{y-y^*} \Lambda'(y-r) f(r) dr.$$

When $y=y^*$ the final term becomes zero and, since $\Lambda(y^*)$ is a minimum, we have $\Lambda'(y^*) = 0$ and hence, using (2-1.18) we get

(2-2.17) $0 = c_P(1-\alpha) + L'(y^*) = c_P(1-\alpha) - c_R + (c_H+c_R)F(y^*).$

Thus y^* is found by equating $F(y^*)$ to a cost ratio similar to that used for the single-stage model.

(2-2.18) $$F(y^*) = \frac{c_R - (1-\alpha)\, c_P}{c_H + c_R} .$$

In the discrete case, by considering $\Delta\Lambda(y) \geqslant 0$, we find y^* to be the least value of y for which

(2-2.19) $$F(y) \geqslant \frac{c_R - (1-\alpha)\, c_P}{c_H + c_R} .$$

Now we return to the problem with a finite number of stages and consider the continuous case first. Differentiating (2-2.10) with respect to y and using (2-2.9) for $n = 2,3,\ldots N$, we get

(2-2.20) $$\Lambda'_n(y) = c_P + L'(y) + \alpha \int_0^\infty \frac{\partial}{\partial y} C^*_{n-1}(y-r)f(r)dr$$

$$= (1-\alpha)c_P + L'(y) + \alpha \int_0^{y-y^*_{n-1}} \Lambda'_{n-1}(y-r)f(r)dr$$

$$= -\alpha c_P + \Lambda'_1(y) + \alpha \int_0^{y-y^*_{n-1}} \Lambda'_{n-1}(y-r)f(r)dr.$$

When $n = 2$, this becomes

(2-2.21) $$\Lambda'_2(y) = -\alpha c_P + \Lambda'_1(y) + \alpha \int_0^{y-y_1^*} \Lambda'_1(y-r)f(r)dr.$$

If y_1^* gives the unique minimum of Λ_1, then the last two terms will only be positive (and hence $\Lambda'_2(y)$ can only be zero) if $y > y_1^*$, hence we must have $y_2^* > y_1^*$. Similarly, when $n = 3$:

$$(2\text{-}2.22)\quad \Lambda_3'(y) = -\alpha c_P + \Lambda_1'(y) + \alpha \int_0^{y-y_2^*} \Lambda_2'(y-r) f(r)\,dr.$$

Since $\Lambda_1'(y_2^*) < \alpha c_P$ for $\Lambda_2'(y_2^*)$ to be zero, this implies that $y_3^* > y_2^*$, and by induction we arrive at Karlin's result quoted above, viz.

$$(2\text{-}2.23)\quad y_1^* \leqslant y_2^* \leqslant \cdots \leqslant y^*,$$

the equalities only being applicable when $c_P = 0$.

Using (2-2.23) and (2-2.21), we can now construct another scheme for calculating our sequence of optimal target stocks.

Scheme B

(1) Find y_1^* from (2-1.15) and y^* from (2-2.18) Then calculate $\Lambda_1'(y)$ from (2-1.13) and (2-1.14) for suitably spaced values of y between y_1^* and y^*.

(2) Set n = 2

(3) Use (2-2.21) and (2-1.14) to evaluate $\Lambda_n'(y)$ at values of y between y_{n-1}^* and y^* and hence find y_n^* by interpolation.

(4) If n = N then stop, otherwise let n become n + 1 and go to Step (3).

This scheme does not involve tabulating $C_n^*(x)$, which accounts for much of the time required by Scheme A. If some values of $C_n^*(x)$ are required, these can be calculated separately after the y_n^* values have been found by Scheme B, so that an efficient grid can be used.

In the discrete case, our starting point is Equation (2-2.11). Differencing with respect to y, this yields

$$(2\text{-}2.24)\quad \Delta\Lambda_n(y) = -\alpha c_P + \Delta\Lambda_1(y) + \alpha \sum_{r=0}^{y-y_{n-1}^*} \Delta\Lambda_{n-1}(y-r)p(r).$$

Reasoning similar to that for the continuous case shows that (2-2.23) still applies and the values of y_n^* can be calculated by proceeding as in Scheme B, working out the corresponding discrete functions. There is no need for interpolation in the discrete case however.

Bellman (1957) points out that, when considering the recurrence relations of dynamic programming, the derivatives (and, by implication, the differences in discrete cases) are more basic quantities to consider and satisfy simpler equations. However, the implications of this for our present model do not seem to have been explored previously in the literature.

Many writers have noted that the objective functions used in inventory theory are very flat in the region of the minimum and this has important consequences when we study the rate at which y_n^* converges to y^* - in other words, when we see how well the infinite-duration model approximates a finite-duration one. This would seem to be a more practical way to view the relationship between the two types of model than to consider the finite-duration model as a truncated approximant for the infinite-duration one in the manner of Bellman or Karlin.

The intuitive argument for convergence is as follows. Looking at (2-2.20) and its discrete analogue (2-2.24), we see that provided Λ_{n-1} is still quite flat at y the integral or sum in the final term for Λ_n' or $\Delta\Lambda_n$ will be small in relation to the second term. When this is so, (2-2.20) becomes effectively

(2-2.25) $\Lambda_n'(y) = -\alpha c_p + \Lambda_1'(y)$,

and (2-2.24) becomes

(2-2.26) $\Delta\Lambda_n(y) = -\alpha c_p + \Delta\Lambda_1(y)$.

The right-hand side of (2-2.25) is the same as that for $\Lambda'(y)$ with a negligible last term and the right-hand side of (2-2.26) is the same as for $\Delta\Lambda(y)$ with a negligible last term. We know that the last term for the infinite case becomes negligible near $y = y^*$ and so it follows that if y^* lies in the flat region for Λ_1 then y_2^* will lie close to y^*.

Some examples of convergence are given in Appendix 1, where it will be noted that many discrete models of practical interest will converge at the second or third stage.

An important practical consequence of rapid convergence is that if the true number of stages in the process is not known with certainty, then the policy for an infinite-duration process will be only slightly sub-optimal and, as we have seen, an analytic solution in terms of the demand d.f. is available for the infinite-duration model.

2-3. Other cost structures

In this section we examine how the method of analysis is affected when we depart from the simple linear cost structure of Sections (2-1) and (2-2).

First we keep the three linear costs but add a set-up cost or fixed purchasing cost, γ_P, of the type we considered in connection with the simple deterministic model of Chapter 1.

If we order any quantity $z > 0$ we must now pay $\gamma_P + c_P z$. For a single-stage process, we shall therefore replace (2-1.5) by

$$C_1(x,y) = \gamma_P + c_P(y-x) + L(y) \tag{2-3.1}$$

$$= \gamma_P - c_P x + \Lambda_1(y).$$

Let us say that the minimum of Λ_1 lies at S_1, which we previously denoted by y_1^*. If no order is placed, the only costs incurred are the losses $L(x)$ resulting from starting the stage with initial stock x, whereas if we order up to S_1 we incur a purchasing cost of $\gamma_P + c_P (S_1-x)$. Thus it will be optimal to place no order so long as

$$L(x) \leqslant \gamma_P + c_P (S_1-x) + L(S_1). \tag{2-3.2}$$

Consider the break-even point and denote the stock level there by s_1. The relationship between s_1 and S_1 will then be

$$L(s_1) = \gamma_P + c_P (S_1-s_1) + L(S_1) \tag{2-3.3}$$

$$\text{or} \quad \Lambda_1(s_1) = \gamma_P + \Lambda_1(S_1).$$

The values S_1 and s_1 are known respectively as the upper and lower critical stocks. They characterise a policy which may be expressed as:

order up to S_1 if $x < s_1$

and order nothing if $x \geq s_1$.

Such a policy is often called an (S,s) policy, or an (S,s,T) policy when it is used in connection with a multistage model where the stock is reviewed periodically with period T.

From (2-3.3) we can see that $s_1 \to S_1$ as $\gamma_P \to 0$ and, since we have defined S_1 to be equal to y_1^* for the model with $\gamma_P = 0$, the optimal policy for Section (2-1) can be seen as a special case of the (S,s) policy where the two critical stocks are both equal to y_1^*. When the two critical stocks are equal at each stage, the (S,s,T) policy is said to become an (S,T) policy; our optimal policy in Section (2-2) is now seen to be of this type. If a set-up cost had been included in our multistage model then after calculating S_n $(= y_n^*)$ at each stage we would have to go on to calculate s_n using a relationship between values of Λ_n corresponding to (2-3.3) for Λ_1.

(2-3.4) $$\Lambda_n(s_n) = \gamma_P + \Lambda_n(S_n).$$

Two special cases of this model are of interest for their connection with the results for the simple deterministic model we examined in Chapter 1. Firstly, when $c_P = 0$ and the demand distribution is exponential with mean $\mu = \lambda T$ we find that putting $S_1 = y_1^*$ in (2-1.15) gives

$$(2\text{-}3.5) \qquad \exp(-S_1/\lambda T) = \frac{c_H}{c_H + c_R} \ .$$

Substituting this into (2-3.3) and rewriting this in terms of the difference between the critical stocks, we get

$$(2\text{-}3.6) \qquad \exp(Q_1/\lambda T) = \frac{\gamma_P}{\lambda T c_H} + 1 + \frac{Q_1}{\lambda T} \ ,$$

$$\text{where} \quad Q_1 = S_1 - s_1 .$$

If $Q_1/\lambda T$ is reasonably small then by considering just the first three terms of the Maclaurin series for the exponential we can get the following approximate solution for Q_1.

$$(2\text{-}3.7) \qquad Q_1 \doteq \sqrt{\frac{2\gamma_P \lambda T}{c_H}} \ .$$

The right-hand side of this is the same as the square-root formula for optimal order quantity with the simple continuous deterministic model if we consider c_H to be $k_H T$.

Secondly, we consider $c_P = 0$ and a geometric demand distribution with mean λT so that

$$(2\text{-}3.8) \qquad p(r) = \left(\frac{1}{1+\lambda T}\right)\left(1-\frac{1}{1+\lambda T}\right)^r \quad \text{for } r = 1,2,\ldots \ .$$

By considering the first three terms of a geometric series we get a result for the difference in critical stocks which is identical to that for optimal order quantity with the simple discrete deterministic model with $c_H = k_H T$; viz. Q_1 is the least value of Q for which

$$(2\text{-}3.9) \qquad Q(Q+1) \geqslant \frac{2\gamma_P \lambda T}{c_H} \ .$$

Scarf (1963) considers the problem of minimising long-run average cost per stage with exponential or geometric demand and he shows that the approximations in (2-3.7) and (2-3.9) give reasonably good results for the difference in critical stocks in that case. Because of this the classical square-root formula and its discrete analogue acquire a new importance.

The idea of equating the instantaneous cost c_H to a time-dependent cost k_H charged for a whole stage of length T suggests a way in which the results we have obtained for instantaneous linear holding and runout costs may be used for some other models with time-dependent costs.

One such model has become widely known from Naddor's description of it in Churchman, Ackoff and Arnoff (1957). It assumes that demand in each stage occurs at a uniform rate with respect to time but that this rate is a random variable taking independent values in each stage. Holding and runout costs are time-dependent and are respectively k_H and k_R per unit quantity per unit time. When the demand rate is treated as a continuous random variable the loss function can be expressed in the form:

$$(2\text{-}3.10)\quad L(y) = T\Big[k_H \int_o^y (y-\tfrac{r}{2}) f(r)\,dr + k_H \int_y^\infty y^2 \frac{f(r)}{r}\,dr + k_R \int_y^\infty (r-y)^2 \frac{f(r)}{2r}\,dr\Big].$$

If we substitute c_H for $k_H T$ and c_H for $k_R T$ and following Naddor (1966) we define a new probability density

$$(2\text{-}3.11) \qquad h(r) = \int_r^\infty \frac{f(x)}{x} dx,$$

the loss function becomes

$$(2\text{-}3.12) \qquad L(y) = c_H \int_o^y (y-r)h(r)dr + c_R \int_y^\infty (r-y)h(dr),$$

which is in exactly the same form as the continuous version of (2-1.6) with h instead of f. Naddor (1966) shows that a similar equivalence exists in the discrete case.

A more important type of model with time-dependent costs is suggested by Hadley and Whitin (1963) who assume that demand is generated by a stochastic process at a known rate. By concentrating on the case of a Poisson process with rate λ they rather obscure the general approach to such models, which we can summarise as follows.

Assume that the d.f. for demand at any time t is F_t. If demand is discrete there is a corresponding p.m.f.p_t, if it is continuous then there is a p.d.f. f_t. Time may vary continuously or discretely but in either case the holding and runout costs are respectively k_H and k_R per unit quantity per unit time.

By considering $k_H T$ and $k_R T$ equivalent to c_H and c_R as in Naddor's model, the loss function for continuous demand can be written as (2-3.12) where:

$$(2\text{-}3.13)\qquad h(r)=\begin{cases}\frac{1}{T}\sum_{t=1}^{T} f_t(r) & \text{for discrete time,}\\ \frac{1}{T}\int_0^T f_t(r)\,dt & \text{for continuous time.}\end{cases}$$

For discrete demand, we define

$$(2\text{-}3.14)\qquad h(r)=\begin{cases}\frac{1}{T}\sum_{t=1}^{T} p_t(r) & \text{for discrete time,}\\ \frac{1}{T}\int_0^T p_t(r)\,dt & \text{for continuous time,}\end{cases}$$

and substitute this for $p(r)$ in the discrete version of (2-1.6). In the language of our Chapter 4, we obtain $h(r)$ as a <u>mixture</u> of the demand distribution at t with a uniform distribution for t.

Hadley and Whitin (1963) consider the Poisson process for which

$$(2\text{-}3.15) \qquad p_t(r) = p(r;\lambda t) = \frac{(\lambda t)^r e^{-\lambda t}}{r!} .$$

Mixing with a continuous uniform distribution on (0,T) gives

$$(2\text{-}3.16) \qquad h(r) = \frac{1}{\lambda T} P(r+1;\lambda T) = \frac{1}{\lambda T} \sum_{x=r+1}^{\infty} p(x;\lambda T).$$

Ignoring any purchasing costs, y_1^* is found by (2-1.12) to be the least value of y for which the d.f. corresponding to h satisfies the usual type of inequality. This may be expressed as:

$$(2\text{-}2.17) \qquad \frac{y+1}{\lambda T} P(y+2;\lambda T) - P(y+1;\lambda T) \geq \frac{c_H}{c_H+c_R} .$$

A useful discrete-time process which has been used in dam theory - see Prabhu (1965) - is based on the Gamma distribution and has

$$(2\text{-}3.18) \qquad f_t(r) = \frac{(1/\lambda)(r/\lambda)^{t-1} e^{-r/\lambda}}{(t-1)!} .$$

Mixing with a discrete uniform distribution on $\{1,\ldots,T\}$ gives a p.d.f. which can be expressed in terms of the Poisson right-hand cumulative function.

$$(2\text{-}3.19) \qquad h(r) = \frac{1}{\lambda T} [1 - P(t;r/\lambda)].$$

Using (2-1.15) and ignoring purchasing costs, we find y_1^* by equating the d.f. of the mixture to the usual cost ratio.

$$(2\text{-}3.20) \qquad \frac{1}{\lambda T} [y_1^* - y_1^* P(T;y_1^*/\lambda) + \lambda T P(T+1;y_1^*/\lambda)] = \frac{c_R}{c_H+c_R} .$$

Similarly, using a loss function based on h we can solve multistage problems with time-dependent costs. When instantaneous costs of our original type are also included the solutions become slightly more complicated but the method is still basically the same.

2-4. Demand distributions and convexity

When we were looking for the minimum of Λ_n and later for that of Λ, we assumed that each of these functions had a unique minimum. With linear costs of the type we have considered, this assumption will always be valid so long as F is strictly increasing. We can see this most clearly in the single-stage model, where we can rewrite (2-1.13) as

$$\Lambda_1'(y) = (c_P - c_R) + (c_H + c_R)F(y). \tag{2-4.1}$$

When $c_R > c_P \geq 0$ and $c_H > 0$, Λ_1' will only change sign once as F increases and so Λ_1 will have a unique minimum at y_1^* and the simple ordering policy based on this value, the limiting (S,s) policy when s = S, will be optimal.

Scarf (1960a) has shown that the condition of linearity of the cost structure can be relaxed without destroying the uniqueness of the minimum, provided that the loss function remains convex. His proof is based upon showing that, although Λ_n may have a number of maxima or minima, the convexity of the loss function will ensure that Λ_n possesses a weak form of convexity which is nevertheless sufficiently strong to ensure that any oscillations in Λ_n will be too small to cause a departure from an (S,s) policy.

The convexity of L can be used to establish some inequalities relating to the convergence of y_n^* to y^*, e.g. Scarf (1963) gives the result

$$\lim_{n\to\infty} \left| \frac{y^* - y_n^*}{\alpha^n} \right| \leq \frac{2c_P}{L''(y^*)} \quad \text{when } \gamma_P = 0. \tag{2-4.2}$$

This indicates that convergence is approximately geometric, a result which may be compared with our figures in Appendix 1.

Karlin (1958a,b) has shown that to make the loss function convex the remaining constraints on the individual costs become fairly mild and easy to determine when the demand p.d.f.(or p.m.f.) belongs to the class of Pólya frequency functions. He has since elaborated the theory of such functions, which we shall define and discuss more fully in Chapter 4 below. At this stage it is sufficient to say that the class contains most of the theoretical distributions which would usually be considered for the demand and that by using a distribution belonging to the class we can regulate the number of possible sign changes in the derivative (or difference) of Λ_n. If there is only one sign change and Λ_n has a unique minimum, then the optimal policy will be affected only slightly by small variations in parameter values and, if the form of the demand distribution is known except for the value of some parameter, each estimate of the unknown parameter can be converted directly into a corresponding optimal stock value. A method of tackling this estimation problem is one of the topics we shall discuss in the next chapter.

CHAPTER 3

MODELS WITH MORE UNCERTAINTY

3-1. Known leadtimes

Depending upon whether ordering or delivery is taken as datum, the time which elapses between placing an order and taking delivery of the same order is called the lag in delivery or the leadtime for ordering.

We now examine how the models in the previous chapter need to be modified when leadtimes equal to ℓ stage lengths (i.e. equal to ℓT) are considered. First we look at the model of Section (2-2) with a leadtime of T, i.e. with $\ell = 1$. This means that stock ordered in the current stage is delivered at the beginning of the next stage, if there is a next stage. Any stock ordered in the final stage will be delivered after the end of the process, so it will always be optimal to order nothing in the final stage. Therefore, using the notation $C^*_{\ell,n}$ in place of C^*_n for the optimal total expected cost of n stages with a leadtime of ℓ stages, we get:

$$(3\text{-}1.1) \qquad C^*_{1,1}(x) = L(x) \text{ for all } x,$$

where L is defined as before for positive argument and

$$(3\text{-}1.2) \qquad L(x) = c_R(\mu - x) \text{ for } x < 0.$$

Following Karlin and Scarf (1958), we may assume that goods are paid for when they are ordered: we shall call this convention cash on order. As an alternative, we may consider the more usual

convention of cash on delivery.

With a continuous demand distribution and cash on order, the recurrence relationship for n > 1 will be:

$$(3\text{-}1.3) \qquad C^*_{1,n}(x) = -c_P x + L(x) + \min_{y \geq x} \{c_P y + \alpha \int_0^\infty C^*_{1,n-1}(y-r) f(r)\,dr\}.$$

The function corresponding to Λ_n is $\Lambda_{1,n}$, which no longer contains the loss-function term since losses in the current stage depend upon available stock rather than upon target stock.

$$(3\text{-}1.4) \qquad \Lambda_{1,n}(y) = c_P y + \alpha \int_0^\infty C_{1,n-1}(y-r) f(r)\,dr.$$

Using (3-1.14) we can rewrite (3-1.3) as:

$$(3\text{-}1.5) \qquad C^*_{1,n}(x) = -c_P x + L(x) + \min_{y \geq x} \{\Lambda_{1,n}(y)\}$$

Differentiating to locate the minimum of $\Lambda_{1,2}$, we find y^*_2 to be the value of y for which

$$(3\text{-}1.6) \qquad \int_0^y F(y-r) f(r)\,dr = \frac{c_R - [1/\alpha]c_P}{c_H + c_R}$$

$$\text{subject to } \alpha c_R > c_P.$$

The term on the left can be written as $F^{(2)}(y)$, using $F^{(2)}$ in the usual sense of a two-fold convolution of F with itself.

The result shown in (3-1.6) can be explained by realising that the ordering decision in the penultimate stage must take account of demand in that stage plus the demand in the final stage and the total demand in two stages will have d.f. $F^{(2)}$.

With cash on delivery, the corresponding result is

$$(3\text{-}1.7) \qquad F^{(2)}(y_2^*) = \frac{c_R - c_P}{c_H + c_R}, \quad \text{subject to the original condition, } c_R > c_P.$$

The infinite-duration (or steady-state) solutions are also expressible in terms of $F^{(2)}$.

$$(3\text{-}1.8) \qquad F^{(2)}(y^*) = \begin{cases} \dfrac{c_R - [(1-\alpha)/\alpha]c_P}{c_H + c_R} & \text{for cash on order,} \\ \dfrac{c_R - (1-\alpha)\,c_P}{c_H + c_R} & \text{for cash on delivery.} \end{cases}$$

Comparing (3-1.8) with (2-2.18), we see that the cash-on-delivery solution is the same as that for a zero-leadtime model with $F^{(2)}$ instead of F. For cash on order, the purchasing cost is divided by α as well.

When $\ell > 1$, $C^*_{l,n}$ becomes a function of the available stock, x, and of all orders already placed but not yet delivered. Now dynamic programming with functions of more than two variables is usually impracticable, but Karlin and Scarf (1958), in their classic treatment of backlog models with leadtime ℓT, show that this problem can be reduced to one involving functions of single variables only. This is a very important result because it shows that computation for any member of stages with a known leadtime of

ℓT is essentially the same as for the corresponding problem with zero leadtime - provided that all excess demands are backlogged.

We present Karlin and Scarf's results as applied to our usual linear cost structure with $\gamma_P = 0$. We use our own notation, which is designed to clear up some ambiguities in the original paper. First we define two new types of function, L_j and $B_{\ell,n}$.

$$(3\text{-}1.9) \qquad L_j(y) = \begin{cases} 0 \text{ for } j = 0, \\ L(y) \text{ for } j = 1, \\ \alpha \int_0^\infty L_{j-1}(y-r)f(r)dr \text{ for } j = 2,3,\ldots\ell. \end{cases}$$

It can be shown that L_j is equivalent to L with c_H and c_R each multiplied by α^{j-1} and $f(r)$ replaced by its j-fold convolution with itself, viz. $f^{(j)}(r)$.

With the cash-on-order convention:

$$(3\text{-}1.10) \qquad B_{\ell,n}(x) = \begin{cases} -c_P x + L_\ell(x) \\ \qquad + \min_{y>x} \{c_P y + \alpha \int_0^\infty B_{\ell,n-1}(y-r)f(r)dr \\ \qquad \text{for } n > \ell \text{ and } \ell = 1,2,\ldots, \\ L_1(x) + L_2(x) + \ldots + L_n(x) \text{ for } n \leqslant \ell. \end{cases}$$

By defining

$$(3\text{-}1.11) \qquad \Lambda_{\ell,n}(y) = c_P y + \alpha \int_0^\infty B_{\ell,n-1}(y-r)f(r)dr,$$

we can rewrite (3-1.10) for $n > \ell$ as:

$$(3\text{-}1.12) \qquad B_{l,n}(x) = -c_p x + L_l(x) + \min_{y \geq x}\{\Lambda_{l,n}(y)\}.$$

Corresponding functions can be defined for discrete demand distributions.

If z_j denotes stock already ordered and due for delivery j stages from the present, $C^*_{l,n}$ will be a function of $x, z_1, z_2, \ldots, z_{l-1}$ which we can express as the sum of a number of functions of single variables.

$$(3\text{-}1.13) \qquad C^*_{l,n}(x, z_1, z_2, \ldots, z_{l-1})$$
$$= L_1(x) + L_2(x + z_1) + L_3(x+z_1+z_2) + \ldots$$
$$+ L_{l-1}(x+z_1+z_2+\ldots+z_{l-2}) + B_{l,n}(x+z_1+z_2+\ldots+z_{l-1}).$$

In the special case where $\ell = 1$, we get

$$(3\text{-}1.14) \qquad C^*_{1,n}(x) = B_{1,n}(x),$$

which is consistent with our treatment of 1-stage leadtime above.

For all values of ℓ we find optimal values of y by locating the minimum of $\Lambda_{l,n}$. The solutions will involve the d.f. $F^{(l+1)}$ instead of the F we used in the zero-leadtime models, e.g. the infinite-duration solutions are:

$$(3\text{-}1.15) \qquad F^{(l+1)}(y^*) = \begin{cases} \dfrac{c_R - [(1-\alpha)/\alpha^{\ell}]c_P}{c_H + c_R} & \text{for cash on order,} \\ \dfrac{c_R - (1-\alpha)c_P}{c_H + c_P} & \text{for cash on delivery.} \end{cases}$$

The typical effect of convolution upon a d.f. is to shift all percentage points (fractiles) to the right, so that we may infer from (3-1.15) that increasing the leadtime, which results in more convolutions, will increase y*.

The practical implication of these results is that calculations will be greatly simplified if F is chosen so that tables of the j-fold convolution, $F^{(j)}$, are available or can readily be calculated. The simplest case is where F and $F^{(j)}$ both belong to a family of distributions which is <u>closed under convolutions</u> in the sense of Feller (1966).

3-2. Stochastic leadtimes

The results of the previous section can be used when we consider the leadtimes to be random variables. In this section we shall assume that orders are dealt with on the FIFO principle (first in, first out) and that if more than one order has been outstanding then leadtime for the next order to be delivered is measured from the delivery of the previous order: in this way we ensure that orders are delivered in the same sequence as they are placed and successive leadtimes do not overlap.

Leadtimes are taken to be independent and identically distributed with p.m.f. $p(\ell)$ for = 1,2,...and a corresponding right-hand cumulative probability $P(\ell)$. Demand will be treated as a continuous random variable, its p.d.f. being f for one stage and $f^{(\ell)}$ for ℓ stages, as before.

We shall denote the maximum permissible number of outstanding orders by k. Scarf (1960b) has considered the case of k=1 and found that the analysis can be brought into a form which resembles that for a known leadtime with ℓ=1. We show that the case of k=2 resembles that of ℓ=2 and suggest that the stochastic model with given k will always resemble the deterministic one with ℓ=k. Following Scarf, we shall use the cash-on-order convention in this section.

First, consider the steady state when k=1 and let the leadtime be ℓ, a known value. If k=1, there may be 1 or 0 orders outstanding. Let $C(x,z,\ell|t)$ denote the total expected optimal cost when a stage begins with an available stock of x units and z units on order with a leadtime of ℓ stages of which t have already elapsed.

We now let $C(x)$ denote the corresponding cost with x units available and none on order. Using the results of (3-1) we can write

$$(3\text{-}2.1)\quad C(x,z;\ell|t) = L_1(x) + L_2(x) + \ldots + L_{\ell-t}(x) + \alpha^{\ell-t}\int_0^\infty C(x+z-r)f^{(\ell-t)}(r)dr$$

Now considering the leadtime to be stochastic, we take the expectation of (3-2.1) with respect to the conditional leadtime distribution. If t stages have already elapsed since placing the order, leadtime must be at least t + 1 and the probability of this is $P(t+1)$; we therefore obtain the conditional leadtime distribution by using $P(t+1)$ as a normalising constant for the probability masses at $\ell > t$. To obtain a compact notation we define three new functions: M_1, α_o and f_o.

$$(3\text{-}2.2)\quad M_1(x|t) = \frac{1}{P(t+1)}\sum_{j=1}^{\infty} L_j(x)P(t+j) \text{ for } t=0,1,\ldots$$

$$\text{which becomes } \sum_{j=1}^{\infty} L_j(x)P(j) \text{ for } t=0.$$

$$(3\text{-}2.3)\quad \alpha_o(t) = \frac{1}{P(t+1)}\sum_{j=1}^{\infty} \alpha^j\, p(t+j) \text{ for } t=0,1,\ldots$$

$$\text{which becomes } \sum_{j=1}^{\infty} \alpha^j p(j) \text{ for } t=0.$$

$$(3\text{-}2.4)\quad f_o(r|t) = \frac{1}{P(t+1)\alpha_o(t)} \sum_{j=1}^{\infty} \alpha^j f^{(j)}(r)p(t+j) \quad \text{for } t = 0,1,\ldots$$

which is a proper normalised p.d.f. and becomes

$$\frac{1}{\alpha_o(t)} \sum_{j=1}^{\infty} \alpha^j f^{(j)}(r)p(j) \quad \text{for } t = 0.$$

The conditional expectation of $C(x,z;\ell|t)$ can then be written

$$(3\text{-}2.5)\quad C(x,z|t) = M_1(x|t) + \alpha_o(t) \int_o^{\infty} C(x+z-r)f_o(r|t)dr.$$

We can use (3-2.5) to express the total cost with no outstanding orders in a form similar to (3-1.13) for a known lag of one stage.

$$(3\text{-}2.6)\qquad C(x) = \min_{z\geqslant o} \{c_p z + C(x,z|0)\}$$

$$= \min_{z\geqslant o} \{c_p z + M_1(x|0) + \alpha_o(0) \int_o^{\infty} C(x+z-r)f_o(r|0)dr\}$$

$$= -c_p x + M_1(x|0) + \min_{y\geqslant x} \{c_p y + \alpha_o(0) \int_o^{\infty} C(y-r)f_o(r|0)dr\}.$$

This summarises the result of Scarf (1960b). He pointed out that while the function which we have called M_1 behaves like the loss function L, the p.d.f. in the equivalent L would not be our f_o except when the leadtime distribution is geometric, say

$$(3\text{-}2.7)\qquad p(\ell) = (1-q)q^{\ell-1} \quad \text{for } \ell = 1,2,\ldots,$$

in which case (3-2.6) will be the same as (3-1.3) with a discount factor of $[\alpha(1-q)/(1-\alpha q)]$, a holding cost of $c_H/(1-\alpha q)$, a runout cost of $c_R/(1-\alpha q)$ and a demand density

$$(1 - \alpha q) \sum_{\ell=1}^{\infty} (\alpha q)^{\ell-1} f^{(\ell)}(r) .$$

Now let us examine the case of k = 2, where there may be 2,1 or 0 orders outstanding. Let z_1 denote the quantity on the first order to be delivered and z_2 that on the second order. Let t be the leadtime for z_1, which has already elapsed. The equation when no more orders may be placed, which corresponds to (3-2.5), will be

$$(3\text{-}2.8) \quad C(x, z_1, z_2 | t) = M_1(x|t) + \alpha_o(t) \int_o^{\infty} C(x+z_1-r, z_2|0) f_o(r|t) dr .$$

The equation when one order may be placed, corresponding to (3-2.6), will be

$$(3\text{-}2.9) \quad C(x, z_1|t) = \min_{z_2 \geq 0} \{c_P z_2 + C(x, z_1, z_2|t)\}$$

$$= \min_{z_2 \geq 0} \{c_P z_2 + M_1(x|t) + \alpha_o(t) \int_o^{\infty} C(x+z_1-r, z_2|0) f_o(r|t) dr\}.$$

We may use this functional equation to replace the last term by a function of $x + z_1$ using the argument developed by Karlin and Scarf (1958) in the case of a known leadtime of two stages. To complete the analogy, we define functions M_2 and B_2^o.

$$(3\text{-}2.10) \qquad M_2(x|t) = \alpha_o(t) \int_o^{\infty} M_1(x-r|t) f_o(r|t) dr.$$

$$(3\text{-}2.11)\qquad B_2^{\circ}(x|t) = -c_p x + M_2(x|t) + \min_{y \geq x}\{c_p y + \alpha_o(t)\int_o^{\infty} B_2^{\circ}(y-r|0) f_o(r|t)\,dr\}.$$

Substituting into (3-2.9), we get

$$(3\text{-}2.12)\qquad C(x, z_1|t) = M_1(x|t) + B_2^{\circ}(x+z_1|t).$$

This is in the same form as (3-1.13) with M_1 instead of L_1 and B_2° instead of $B_{2,n}$. However, each function is now dependent upon t, but by considering t = o in (3-2.11) we can evaluate $B_2^{\circ}(\cdot|0)$ and hence $B_2^{\circ}(\cdot|t)$ for any t.

The optimal policy when there are no outstanding orders can be found by using the relationship

$$(3\text{-}2.13)\qquad C(x) = \min_{z_1, z_2 \geq 0}\{c_p(z_1+z_2) + C(x, z_1, z_2|0)\}$$

and then proceeding in much the same way as for $C(x, z_1|0)$.

We may infer from this consideration of k = 2 that, for any value of k, the equation when one more order may be placed will always resemble that for known leadtime ℓ = k. We suggest that a general proof by induction would show the general equation to be of the form:

$$(3\text{-}2.14)\qquad C(x,z_1,z_2,\ldots,z_{k-1}|t) = M_1(x|t) + M_2(x+z_1|t)+\ldots + M_{k-1}(x+z_1+z_2+\ldots+z_{k-2}|t) + B_k^{\circ}(x+z_1+z_2+\ldots+z_{k-1}|t),$$

where

$$(3\text{-}2.15)\qquad M_j(x|t) = \alpha_o(t)\int_o^{\infty} M_{j-1}(x-r|0)f_o(r|t) \text{ for } j = 2,3,\ldots$$

and

$$(3\text{-}2.16)\qquad B_k^{\circ}(x|t) = -c_P x + M_k(x|t) + \min_{y\geq x}\{c_P y + \alpha_o(t)\int_o^{\infty} B_k^{\circ}(y-r|0)f_o(r|t)dr\}.$$

Certainly when k = 1 or 2 (and also, we suggest, in the general case) it will be of interest to study the known-leadtime model with a demand distribution which is a convolution of the basic distribution mixed with a geometric distribution for the convolution index.

Hadley and Whitin (1963) have examined the problem of stochastic leadtimes when demand is governed by a Poisson process. However, their solution depends upon the very restrictive assumption that the range of possible leadtimes is less than the stage length, T. The model which we have just considered takes T as the minimum leadtime rather than the maximum and this assumption is often more realistic.

3-3. Demand distributions which vary with time

For simplicity, we have assumed up to now that the demands in successive stages are independent and identically distributed. It is well known, however, that most practical demand processes show appreciable autocorrelation, which is often combined with trend in the mean and perhaps cyclical variations. Any satisfactory general model for prediction must take account of these features - see Box and Jenkins (1968) - and they can in fact be incorporated into the type of model which we have been considering. The general method for incorporating them is given in a series of papers by Karlin (1960a,b) and Iglehart and Karlin (1962), which we briefly summarise as follows.

Demand distribution is varied by the operation of a Markov transition matrix which governs the selection of a particular demand state for each stage; with each state is associated a p.d.f. or p.m.f. which is to govern demand during the stage. Provided each of the possible p.d.f's (or p.m.f's) belongs to the class of Pólya frequency functions the optimal policies will be of the simple form characterised by a set of critical stock levels of the same type as our $\{y_n^*\}$. The actual calculation of the critical stocks is rather complicated but it is described in detail by Karlin (1960a), who gives a worked example (unfortunately this appears to contain a misprint on p.620 where the value of y_1 is stated to be 5.6 instead of 4.6). The method is applicable to models with or without leadtime and with backlog or lost sales.

Another type of temporal variation in the demand distribution may arise when we know its general form but have imperfect knowledge

about the value of some parameter. One way of dealing with this situation is to use observations of demand to produce a maximum-likelihood estimate of the unknown parameter and use this estimate in the known form of p.d.f. or p.m.f. However, such estimates tend to be rather erratic when the sample size is very small and in this case a Bayesian method may be preferable. Such methods are often criticised for combining opinions with observations but from a practical point of view they provide a good way of using intelligent guesswork when little or no solid information is available and then letting real data submerge the initial guesswork as observation proceeds. In this way we get a regular narrowing down of the demand distribution with increasing information. In the next chapter we shall explain Bayesian methods more fully and show that they involve an important special case of distribution mixing, a topic to which we have already referred.

The history of Bayesian methods in inventory theory seems to begin with Dvoretzky, Kiefer and Wolfowitz (1952), who gave the topic a very abstract treatment in a paper dealing with various methods of treating imperfectly known demand distributions. Then Scarf (1959 ,1960c) showed how these methods could be applied when the demand distribution belonged to the Exponential Family (see Section(4-2) below) and in particular to the class of Gamma distributions. For further results with Gamma distributions see Fukuda (1960), Scarf and VanderVeer (1961) and Wilson (1966). A more general version of Scarf(1959) has been provided by Iglehart (1964), while Hansel (1966) and Lampkin (1967) suggest using Bayesian methods with an imperfectly known Poisson distribution.

To understand the Bayesian methods and various requirements which we have imposed upon distributions mentioned so far, we shall now go on to examine some basic ideas of probability distribution theory.

CHAPTER 4

THE BASIC DISTRIBUTION THEORY

4-1. Pólya functions and total positivity

We have seen that in his classical exposition of the Arrow-Harris-Marschak inventory models, Karlin (1958a,b) was able to generalise the results of Bellman (1957) and other early authors by introducing the restriction that demand distributions should belong to a special class called after George Pólya - the class of Pólya frequency functions. This restriction was quite mild since the class contains nearly all the usual distributions of statistical theory. The generalisation resulted in greater freedom in choosing the form of the cost functions for holding and runout so that the analysis could be extended to cost structures other than the linear type which we have considered.

The choice of Pólya's name for the class of functions is due to Schoenberg (1950,1951), who was following up some earlier work by Pólya. A considerable corpus of theory for Pólya frequency functions and related topics has built up in the last twenty years and is now being brought together by Karlin (1968) in two volumes, the first of which has already appeared. We shall briefly summarise the ideas which find applications in inventory theory, but first it should be pointed out

that a "Pólya frequency function" is not the same as a "Pólya distribution", the term used by many authors to denote a mixed distribution which is a form of Negative Binomial first studied by Eggenburger and Pólya (1923) - see our Chapter 7 below. Another source of confusion is the "Pólya - Aeppli distribution" of Skellam (1952) and Moran (1968), which is another mixture.

In what follows, we take f to be a typical univariate probability density or mass function. We may define f by

$$(4\text{-}1.1) \qquad f(x) = A.K(x),$$

where A is a normalising constant and the random variable whose distribution is described by f takes values in a real set $\Omega(x)$ to which x in (4-1.1) belongs. Alternatively, we may introduce a parameter θ and write.

$$(4\text{-}1.2) \qquad f(x|\theta) = \alpha(x)\, K(x,\theta).$$

Considered as a function of θ, the expression in (4-1.2) is called the <u>likelihood</u> of a single observation of the r.v. described by f. Following Raiffa and Schlaifer (1961) we shall call $K(x,\theta)$ a <u>kernel of the likelihood</u>. For any f there may be several possible factorisations like (4-1.2) and hence several possible kernels.

We now consider $K(x,\theta)$ to be any function of two real variables where x belongs to a set $\Omega(x)$ and θ to a set $\Omega(\theta)$. We can then summarise the relevant ideas of Karlin (1968) in the following definitions, where we have altered the wording so as to present the ideas without recourse to measure theory.

DEFINITION A. If we select m values from $\Omega(x)$ and m values from $\Omega(\theta)$ and order them so that $x_1 < x_2 < \ldots < x_m$ and $\theta_1 < \theta_2 < \ldots < \theta_m$, the function $K(x,\theta)$ is said to be <u>totally positive of order k</u> (TP_k) if the following determinant is non-negative for $m = 1,2,\ldots,k$.

$$\begin{vmatrix} K(x_1,\theta_1) & K(x_1,\theta_2) & \ldots & K(x_1,\theta_m) \\ K(x_2,\theta_1) & K(x_2,\theta_2) & \ldots & K(x_2,\theta_m) \\ \vdots & \vdots & & \vdots \\ K(x_m,\theta_1) & K(x_m,\theta_2) & \ldots & K(x_m,\theta_m) \end{vmatrix}$$

DEFINITION B. If $K(x,\theta)$ is TP_k and is a kernel of the likelihood $f(x|\theta)$, then the distribution specified by $f(x|\theta)$ is said to be <u>Pólya type of order k</u> (PT_k).

DEFINITION C. If $\kappa(u)$ is a function of a single real argument u and $K(x,y) = \kappa(x-y)$ is TP_k when y belongs to a suitable set of ordered real values, then $\kappa(u)$ is a Pólya frequency function of order k (PF_k) and a probability distribution with density or mass function $f(x) = A.\kappa(x)$ is called a PF_k probability density or mass function.

From these definitions it can be seen that all distributions characterised by a parameter are PT_1 and that all those with a monotone likelihood ratio, i.e. those for which the likelihood ratio $f(x|\theta_1)/f(x|\theta_2)$ is a non-decreasing function of x for $\theta_1 < \theta_2$, are PT_2. Most of the univariate parametric distributions of statistical theory, with the notable exception of the Cauchy distribution, are PT_2 with respect to one or other of their parameters. The most important PT_k distributions belong to the Exponential Family or the Range Family, which are described in the next section.

The requirements for monotone likelihood ratio (or equivalently for PT_2) provided Scarf (1959) and Iglehart (1964) with methods of demonstrating the behaviour of the optimal stocks $\{y_n^*\}$ when Bayesian methods are used.

An important property of all TP_k functions, and hence of PF_k distributions in particular, is the sign variation diminishing

property. For a PF_k density f(x), this property may be stated in terms of the convolution transformation.

$$(4\text{-}1.3) \qquad h(x) = \int_0^\infty g(y)\, f(x-y)dy.$$

If g(y) changes sign r times (r<k) as y traverses the real axis, then h(x) changes sign at most r times. Moreover, if h(x) does actually change signs r times, then h(x) will have the same arrangement of signs as g(y) has when x and y take ascending values in their respective sets.

It was this property, a direct result of the determinant inequality required in Definition A, which was used by Karlin (1958a,b) for the inventory models and has since been applied in many other fields - see Barlow and Proschan (1965) and Aoki(1967).

The most important PF_k distributions are also members of the Exponential Family or Range Family when their parameters are considered; they include the Poisson, Gamma, Normal, Lognormal, Binomial, Negative Binomial, Beta, Rectangular and Uniform distributions which we consider in the following chapters. A simple characteristic property of PF_2 distributions is that they are unimodal.

4-2. The Exponential Family and the Range Family

The well-known Fisher-Neyman factorisation criterion for sufficiency states that if the likelihood can be factorised as follows

$$(4\text{-}2.1) \qquad f(x|\theta) = \alpha(x)\, K(t(x),\theta),$$

in other words x appears in the kernel only in the function t(x), then the statistic t(X), which is the corresponding function of the observable r.v. X, will be sufficient for θ when this parameter (which may be multidimensional) is unknown.

An important class of distributions, which have sufficient statistics whose dimensionality remains unchanged with sampling, is the Exponential Family. In the older literature, this family is often called after Koopman, Pitman or Darmois - see Ferguson (1967). Distributions of the family have a likelihood which can be factorised in the following way.

$$(4\text{-}2.2) \qquad f(x|\theta) = \alpha(x)\beta(\theta)\exp\Big[\sum_{j=1}^{k} Q_j(x)R_j(\theta)\Big].$$

Comparison with (4-2.1) shows that the sufficient statistic for θ will be

$$(4\text{-}2.3) \qquad t(X) = \{Q_1(X),\ Q_2(X),\ \ldots,\ Q_k(X)\}.$$

In most cases we are only interested in θ being one- or two-dimensional. In what follows, we shall restrict θ to denote a single unknown parameter or the first of a pair and use ϕ for a second

unknown parameter. In practice Equation (4-2.2) will then become

$$(4\text{-}2.4) \qquad f(x|\theta) = \alpha(x)\beta(\theta) \exp [Q_1(x)R_1(\theta)]$$

for one unknown parameter and

$$(4\text{-}2.5) \qquad f(x|\theta,\phi) = \alpha(x)\beta(\theta,\phi)\exp [Q_1(x)R_1(\theta) + Q_2(x)R_2(\phi)]$$

for two unknown parameters.

A list of the factors for the likelihood of several useful distributions of the Exponential Family has been given by Lindgren(1968).

An important feature of the Exponential Family is that the set of x-values for which $f(x|\theta)$ is a proper non-negative probability density or mass function does not depend upon the unknown parameter (or parameters). However, there are distributions which admit of a sufficient statistic of fixed dimension and for which the admissible set of x-values is determined by θ (and possibly ϕ). Such distributions may be said to belong to the Range Family, of which the simplest members are the Uniform and Rectangular distributions of our Chapter 8 below.

The general form of likelihood for the Range Family, corresponding to (4-2.2) can be written in one of the following forms.

$$(4\text{-}2.6) \qquad f(x|\theta,\phi) = \alpha(x)\ \beta(\theta,\phi) \quad \text{for } \omega_1(\theta) \leqslant x \leqslant \omega_2(\phi).$$

$$(4\text{-}2.7) \qquad f(x|\theta) = \alpha(x)\ \beta(\theta) \quad \text{for } \omega_1(\theta) \leqslant x \leqslant \omega_2(\theta).$$

In cases of practical interest the functions ω_1 and ω_2 will usually have a very simple form such as $\omega_1(\theta) = -\theta$ and $\omega_2(\theta) = \theta$ or $\omega_1(\theta) =$ a constant and $\omega_2(\theta) = \theta$.

When a sample of n values of x is drawn from a distribution of the Range Family, the sufficient statistics will be of the form $t(X) = \max \{X_j\}$ or $\min \{X_j\}$. For further details and a form of likelihood which generalises both the Exponential Family and the Range Family, see Ferguson (1967).

4-3. Convolutions, mixtures and Bayesian methods

To avoid considering discrete and continuous distributions separately we shall temporarily make use of Stieltjes integrals in the manner of Parzen (1960). We then define the convolution of the d.f. F with the d.f. G to be H such that

$$(4\text{-}3.1) \qquad H(x) = \int F(y-x)\, dG(x).$$

If F is the same as G, then H is called the two-fold convolution of F with itself and is written $F^{(2)}$. Similarly, k-fold

convolutions of F with itself may be defined and written $F^{(k)}$ - the notation we used above in Chapter 3.

When the convolution of two members of a class or family of distributions is again a member of that class, we say that the class is closed under convolutions.

An operation which is somewhat reminiscent of convolutions is what we shall call mixing. If we have a parametric d.f. $F(x|\theta)$ and the parameter has d.f. $G(\theta)$, we define the marginal or mixed d.f. to be $H(x)$ given by the following equation

$$(4\text{-}3.2) \qquad H(x) = \int F(x|\theta)\, dG(\theta).$$

The literature on mixing is bedevilled by the variety of terms using to describe the functions and operations involved. The most popular alternative names for our marginal distribution would seem to be "contagious distribution" or "compound distribution". We shall refer to F as the basic distribution and G as the mixer.

Karlin (1968) deals with the mixing of PF_k distributions and their closure under convolutions. Our interest is mainly directed towards those sub-families of the Exponential family which are closed under convolutions and a special type of mixing which results from the Bayesian approach to parameter estimation.

Briefly, the Bayesian approach is this. We know the form of the likelihood $f(x|\theta)$ but θ is unknown. We therefore assign a prior distribution , G_o, to the unknown parameter. By mixing f with G_o we obtain a prior marginal mass or density function, h_o.

$$(4\text{-}3.3) \qquad h_o(x) = \int f(x|\theta)\, dG_o(\theta).$$

We may imagine that our basic r.v. represents demand in one stage of an inventory process, in which case we would use h_o instead of the f or p in the equations we developed for known demand distributions. In this way our model would reflect the uncertainty about the true parameter value which is expressed in its prior distribution.

After a value of the basic r.v. has been observed, say it is x_1, we can find a <u>posterior</u> or updated distribution for the unknown parameter, say G_1 where the subscript denotes the number of actual observations. Using the theorem of Bayes, we find G_1 as follows:

$$(4\text{-}3.4) \qquad dG_1(\theta|x_1) = \frac{f(x_1|\theta)dG_o(\theta)}{h_o(x_1)} .$$

Writing g_o and g_1 for the mass or density functions corresponding to G_o and G_1 respectively, we can express this more simply as

$$(4\text{-}3.5) \qquad g_1(\theta|x_1) = \frac{f(x_1|\theta)g_o(\theta)}{h_o(x_1)} .$$

Proceeding in a similar fashion we can find the marginal mass or density function h_1 which will depend upon x_1, hence a posterior function g_2 after 2 independent observations and so on until after n independent observations we may write the posterior and marginal functions for $n > 1$ as follows.

$$(4\text{-}3.6) \qquad g_n(\theta|x_1,x_2,\ldots,x_n) = \frac{f(x_n|\theta)g_{n-1}(\theta|x_1,x_2,\ldots,x_{n-1})}{h_{n-1}(x_n|x_1,x_2\ldots,x_{n-1})} .$$

$$(4\text{-}3.7) \qquad h_n(x|x_1,x_2,\ldots,x_n) = \int f(x|\theta)dG_n(\theta|x_1,x_2,\ldots,x_n).$$

The notation shows the dependence of g_n and h_n upon all the x - values observed to date. In the case of basic distributions which admit of a sufficient statistic for the unknown parameter, we may summarise the sampling history in the statistic instead of explicitly stating the observed x-values. Thus if t_n is a sufficient statistic summarising n independent observations we may replace the last two equations by:

$$(4\text{-}3.8) \qquad g_n(\theta|t_n) = \frac{f(x_n|\theta)g_{n-1}(\theta|t_{n-1})}{h_{n-1}(x_n|t_{n-1})} ,$$

$$(4\text{-}3.9) \qquad h_n(x|t_n) = \int f(x|\theta)dG_n(\theta|t_n) \text{ using the Stieltjes integral}$$

$$= \begin{cases} \int f(x|\theta)g_n(\theta|t_n)d\theta \text{ using the ordinary (Riemann) integral for continuous } \theta, \\ \sum f(x|\theta)g_n(\theta|t_n) \quad \text{for discrete } \theta . \end{cases}$$

These equations show us that, when the dimensionality of t_n is not affected by increasing n, the functional form of g_n will be the same as that of g_{n-1} and the form of h_n the same as that of h_{n-1} for all $n > 1$. If we introduce a parameter s_o into g_o so that g_o has the same functional form as g_n for $n > 0$, i.e. it may be

written $g_o(\theta|s_o)$, then provided that s_o combines with the x - values in the same way that they combine with each other to form t_n we may summarise the sampling history and the initial choice of parameter in a "statistic" s_n of the same dimensionality as t_n. Equations (4-3.8) and (4-3.9) with t_n replaced by s_n will then be applicable for $n = 0,1,2,\ldots$.

When g_n and h_n have the same functional form for each value of n they are called <u>self-reproducing</u> and the set or system of functions (f, g_n, h_n) is said to be <u>closed under sampling</u>.

It can be shown that if the unknown parameter has some true value $\tilde{\theta}$ and this is included in the set on which the prior distribution is defined, then as $n \to \infty$ the posterior distribution will tend to a unit impulse or delta function at $\theta = \tilde{\theta}$ and $h_n(x|s_n) \to f(x|\tilde{\theta})$. However, Bayesian methods are most useful when n is small and little information is available from sampling. Initial uncertainty about the parameter value is reflected in a broad prior marginal distribution; as sampling proceeds, this narrows down in a regular fashion.

4-4. Conjugate distributions

Raiffa and Schlaifer (1961) brought together the ideas which we have considered in the earlier part of this chapter and showed that a system of distributions closed under sampling would result when f was the mass or density function of a distribution of the Exponential Family or Range Family and g_o was a mass or density function based on a kernel of f. More explicitly, if $f(x|\theta)$ has a kernel $K(x,\theta)$, then $g_o(\theta|s_o)$ should be of the form $\nu.K(s_o,\theta)$, where ν is some suitable normalising constant. They called a prior distribution of this kind a natural conjugate of $f(x|\theta)$.

Now we have already stated that there may be several possible kernels of $f(x|\theta)$ and Raiffa and Schlaifer (1961) showed that the class of natural conjugates may be further enriched by introducing other parameters into the prior distribution beside s_o. Spragins (1965) has since shown that a closed system of distributions will still be obtained if the prior distribution is of the form $\nu(\theta)K(s_o,\theta)$ where ν is any non-negative function of θ containing a suitable normalising constant; a natural conjugate is then seen to be a special case in which ν is independent of θ. We shall use the term conjugate distribution to describe the general type suggested by Spragins, but for simplicity we shall be mainly interested in natural conjugates when we examine Bayesian treatment of useful distributions in the chapters which follow.

By suitable parameterisation it will usually be possible to reflect a wide range of prior knowledge or opinions about the possible values of the unknown parameter by choosing prior distributions

from the conjugate family. Bearing in mind the advantages of closure under sampling when g_n and h_n come to be evaluated, it will be advantageous to choose a prior distribution from the conjugate family unless this completely fails to reflect fairly definite prior knowledge about the unknown parameter.

Most of the existing literature on Bayesian methods is concerned with observing x-values and hence adapting g_n, from which suitable estimates of the parameter value (or unobservable state of nature) can be made at each stage of sampling. Much attention has also been paid to the possible types of sampling process. In the inventory models, where we take the basic random variable to be demand (or possibly leadtime), the sampling process is already determined: we get one independent observation at each stage. Also we have seen that optimal inventory policies depend on the effective d.f. for demand and it is this, rather than a parameter value, which we usually want to estimate at each stage. Our interest then is concentrated on h_n rather than g_n.

In the next chapter we give the detailed derivation for Poisson and Gamma basic distributions and the following chapters summarise the corresponding results for other useful distributions. The general method can be seen by considering a basic distribution for which $f(x|\theta)$ factorises to $\alpha(x)K(x,\theta)$. Imagine we choose a conjugate prior distribution with

$$(4\text{-}4.1) \qquad g_o(\theta|s_o) = \nu_o K(s_o,\theta).$$

Let us now introduce the symbol * to denote the operation which combines a new observation with an old "statistic" to produce a new "statistic", so that

$$(4\text{-}4.2) \qquad s_1 = s_0 * x_1 .$$

We shall see that in practice the operation is often simple addition or multiplication. Assuming that the prior distribution is continuous, the corresponding marginal mass or density function is then obtained as follows.

$$(4\text{-}4.3) \qquad h_0(x|s_0) = \int f(x|\theta)\, g_0(\theta|s_0)\, d\theta$$

$$= \alpha(x)\, \nu_0 \int K(x,\theta) K(s_0,\theta)\, d\theta$$

$$= \alpha(x)\, \nu_0 \int K(s_0 * x,\theta)\, d\theta .$$

If $\nu_1(x)$ is a factor which will normalise $K(s_0 * x,\theta)$ to convert it to a proper p.d.f., we can then write

$$(4\text{-}4.4) \qquad h_0(x|s_0) = \alpha(x)\, \frac{\nu_0}{\nu_1(x)} ,$$

since the integral of $\nu_1(x)K(s_0 * x,\theta)$ will be unity. When x_1 is observed, $s_0 * x$ becomes s_1 by (4-4.2) and let us say that $\nu_1(x)$ becomes ν_1 so that $g_1(\theta|s_1)$ can be expressed as $\nu_1 K(s_1,\theta)$. We can then find h_1 by using (4-4.3) with suffix 0 replaced by suffix 1 throughout. The extension to finding h_n for $n > 1$ is then obvious.

The implication for our inventory equations, e.g. (2-2.7) is that if we replace $f(r)$ by $h_n(r|s_n)$ the total expected cost will depend on the current value of s_n, so that (2-2.7) would be replaced by an equation of the following form.

$$(4\text{-}4.5) \qquad C_m^*(x|s_n) = \min_{y \geq x} \{c_p(y-x) + L(y|s_n) + \alpha \int_0^\infty C_{m-1}^*(y-r)h_n(r|s_n)dr\}.$$

The suffix n now denotes the number of demand observations or elapsed stages of the inventory process while the suffix m denotes the number of stages till the end of the process (the n of Chapters 2 and 3).

In general, the recurrence relationship for dynamic programming now involves functions of two variables, but Scarf (1960c) has shown that the relationship can be reduced to one involving functions of a single variable when the demand distribution is Gamma with unknown scale parameter - see Chapter 5 below.

When we have two unknown parameters, two "statistics" will be introduced into the recurrence relationships, which will then involve functions of three variables. Dynamic programming may still be feasible at this level but the curse of dimensionality prevents consideration of more unknown parameters if Bayesian methods are used. With more unknown parameters there appear to be two possible course of action; either to use Bayesian methods to find h_n, solve the infinite-duration problem with this marginal distribution and rely on convergence of the critical stock levels to ensure that the infinite-duration solution will be nearly optimal for finite duration or to use the basic distribution

with maximum-likelihood estimates of the unknown parameters and use dynamic programming. When only a small number of observations have been made, the first course seems preferable and we would recommend Bayesian methods as being most suitable when few observations have been made, even for the case of only one unknown parameter.

CHAPTER 5

THE POISSON AND GAMMA DISTRIBUTIONS

5-1. Definitions and inventory applications

We shall denote the typical random variable for the Poisson distribution by $Po(\lambda)$, whose p.m.f. is

$$(5\text{-}1.1) \qquad f_{P_o}(x|\lambda) = \frac{\lambda^x e^{-\lambda}}{x!} \text{ for } x = 0,1,2,\ldots \text{ and } \lambda > 0.$$

The mean and variance of $P_o(\lambda)$ are both equal to λ.

Closely connected with Poisson distributions are those of the Gamma family. The typical Gamma r.v. is $Ga(a,b)$, which is continuous and has p.d.f.

$$(5\text{-}1.2) \qquad f_{Ga}(x|a,b) = \frac{b(bx)^{a-1}e^{-bx}}{\Gamma(a)} \quad \text{for } x > 0,$$

where $\Gamma(\cdot)$ is the complete gamma function.

The parameter a specifies the order of the Gamma distribution, while b is the scale parameter. The mean of $Ga(a,b)$ is a/b and the variance a/b^2.

In general, it is sufficient to impose the condition $a > 0$ and $b > 0$, but certain special cases are of particular importance. When $b = 1/2$ and $a = n/2$ for some positive integer n, then we have the distribution of $\chi^2(n)$, i.e. Chi-squared with n degrees of freedom.

When $a = 1$ we have the exponential distribution (sometimes called negative exponential) with parameter b. When a is some positive integer, so that $\Gamma(a)$ can be replaced by $(a-1)!$, we have an Erlang distribution.

All standard textbooks on stochastic processes, e.g. Parzen (1962), introduce the Poisson process, in which events happen at random intervals. When the length of the intervals is exponentially distributed, the total number of events in a given time will be Poisson distributed. We have already discussed an application of the Poisson process in Section (2-3), where we were interested in periodic-review inventory processes with time-dependent costs. Whitin (1953) and Hadley and Whitin (1963) mainly concentrate on the Poisson process as the mechanism for generating demands in continuous and periodic inventory models. Most other writers on continuous models do likewise, although Bather (1966) has suggested using the Wiener process, which is based on the Normal distribution.

As we noted in Section (2-3), using a slightly different notation, the relationship between the Poisson p.m.f. and the Gamma p.d.f. can be expressed as

$$(5\text{-}1.3) \qquad \int_0^t f_{Ga}(y|a,b)dy = \sum_{x=a}^{\infty} f_{Po}(x|bt).$$

Using the d.f. and right-hand cumulative function, we can write this more succinctly as

$$(5\text{-}1.4) \qquad F_{Ga}(t|a,b) = P_{Po}(a|bt).$$

It follows from (5-1.4) that tables of probabilities for $P_o(\lambda)$ can be used to calculate probabilities for Ga(a,b), and vice versa.

The most useful tables are those of Molina (1942) and Pearson (1922). Details of these and other tables are given by Raiffa and Schaifer (1961). Since they wrote, a more extensive tabulation of the Poisson distribution has been prepared by the General Electric Company of America - see Pelletier (1962).

Numerous publications about the Poisson distribution are summarised in the monograph by Haight (1967).

The Gamma distributions are useful in periodic-review inventory theory both to describe the length of leadtimes, as we noted in Section (3-2), and as distributions of demand, as we noted in Section (2-3). Also in cases where a Lognormal distribution is applicable, a Gamma distribution may be used, since the two parameters can be chosen so as to give a Gamma p.d.f. almost identical to that for the Lognormal.

Bearing in mind Karlin's requirements for Pólya functions in inventory theory, we show that Poisson and Gamma distributions belong to the Exponential Family and so satisfy the Pólya conditions. Consider first $P_o(\lambda)$ with unknown λ. Putting $\lambda = \theta$ we can factorise f_{Po} as in (4-2.2) if we substitute as follows:

$$(5\text{-}1.5) \qquad \alpha(x) = \frac{1}{x!}, \ \beta(\theta) = e^{-\theta},$$

$$Q_1(x) = x, \ R_1(\theta) = \log \theta.$$

Using the sufficiency theory of Chapter 4, we see that since $Q_1(x) = x$ the sum of observed x's will be sufficient for θ. In the same way, the distribution of Ga(a,b) with $\theta = a$ gives

$$(5\text{-}1.6) \qquad \alpha(x) = \frac{1}{x}\exp(-bx),\ \beta(\theta) = \frac{b^{\theta}}{\Gamma(\theta)},$$

$$Q_1(x) = \log x,\ R_1(\theta) = \theta.$$

It follows that the sum of logarithms, or simply the product, of observed x's will be sufficient for θ. When the scale parameter is unknown, we put $\theta = b$ and get

$$(5\text{-}1.7) \qquad \alpha(x) = x^{a-1},\ \beta(\theta) = \frac{\theta^{a}}{\Gamma(a)},$$

$$Q_1(x) = x,\ R_1(\theta) = -\theta.$$

As in the case of the Poisson, the sum of x's will be sufficient for θ. When both parameters are unknown, we put $\theta = a$ and $\phi = b$ to get

$$(5\text{-}1.8) \qquad \alpha(x) = \frac{1}{x},\ \beta(\theta,\phi) = \frac{\phi^{\theta}}{\Gamma(\theta)},$$

$$Q_1(x) = \log x,\ R_1(\theta) = \theta,$$

$$Q_2(x) = x,\ R_2(\phi) = -\theta.$$

Here the product of x's will be sufficient for θ and the sum of x's for ϕ.

These factorisations and the knowledge about sufficiency will help us to identify the conjugates for Bayesian analysis in the next section.

5-2. Convolutions and Bayesian treatment of Poisson distributions.

It follows from the fact that the m.g.f of $P_o(\lambda)$ is

$$(5\text{-}2.1) \qquad \psi_{P_o}(t|\lambda) = \exp\left[-\lambda(1-e^t)\right]$$

that the n-fold convolution of the distribution of $P_o(\lambda)$ results in an m.g.f. of the same form as that in (5-2.1) but with λ replaced by $n\lambda$. Similarly, since the m.g.f of Ga(a,b) is

$$(5\text{-}2.2) \qquad \psi_{Ga}(t|a,b) = \left(\frac{b}{b-t}\right)^a ,$$

an n-fold convolution replaces a by na. Hence both the Poisson and Gamma families are closed under convolutions, and we may conveniently select demand distributions from these families when we are dealing with stochastic leadtimes.

When we consider mixing a Poisson distribution, we find that the best-known mixing distribution is in fact the natural conjugate. Following Eggenburger and Pólya (1923), we mix the distribution of $P_o(\lambda)$ with a Gamma distribution for unknown λ, which means that we let

$$(5\text{-}2.3) \qquad \begin{aligned} &f(x|\theta) = f_{P_o}(x|\theta) \quad \text{and} \\ &g_o(\theta|s_o) = f_{Ga}(\theta|s_o,t), \text{ where } s_o \text{ is some positive integer.} \end{aligned}$$

The prior marginal p.m.f. is then found to be

$$(5\text{-}2.4) \qquad h_o(x|s_o) = \int_o^\infty f(x|\theta)\, g_o(\theta|s_o)d\theta$$

$$= \frac{(s_o+x-1)!}{x!(s_o-1)!} \frac{b^{s_o}}{(1-b)^{s_o+x}} \int_o^\infty \frac{(1+b)^{s_o+x}\, \theta^{s_o+x-1}\, e^{-(1+b)\theta}}{(s_o + x - 1)!} d\theta$$

$$= \binom{s_o + x - 1}{x} \left(\frac{b}{1+b}\right)^{s_o} \left(\frac{1}{1+b}\right)^{x}$$

$$= f_{NBi}(x|s_o,\ 1/(1 + b)\ ),$$

which is the p.m.f. of NBi $(s_o,1/(1+b)\ )$, a Negative Binomial r.v. as defined below in Chapter 7.

Defining s_1 to be $s_o + x_1$ and remembering that the sum of x's is sufficient for θ, we find that the first prior p.d.f. for θ will be

$$(5\text{-}2.5) \qquad g_1(\theta|s_1) = f(x_1|\theta)\, g_o(\theta|s_o) \,/\, h_o(x_1|s_o)$$

$$= \frac{(1+b)^{s_1}\, \theta^{s_1-1}\, e^{-(1+b)\theta}}{(s_1 - 1)!}$$

$$= f_{Ga}(\theta|s_1,\ 1+b)\ .$$

Defining s_n to be $s_{n-1} + x_n$ and proceeding in the same way as the sample size increases, we find that

$$(5\text{-}2.6) \qquad g_n(\theta|s_n) = f_{Ga}\ (\theta|s_n,\ n+b),$$

$$(5\text{-}2.7) \qquad h_n(x|s_n) = f_{NBi}(x|s_n,\ 1/(n+b)\ ).$$

Comparing these two equations with (5-2.3) and (5-2.4), we see that the Gamma-mixed Poisson system is closed under sampling. This comes about because the p.d.f. of $Ga(s_o,b)$, which we used as prior mixer, is in the form of a kernel of $f_{P_o}(x|\theta)$ with a suitable normalising constant, $(s_o-1)!$, and the scale parameter, b, is simply introduced to enrich the family of mixing densities.

The practical implication of this result is that if we have an inventory process where demands are Poisson distributed with unknown mean we may introduce a prior estimate of the mean (s_o/b) to which we attach a variance (s_o/b^2). Having thus implicitly assigned values to s_o and b, we substitute the most recent version of h_n, as given by (5-2.6), for the demand p.m.f. p in the various formulae in Chapters 2 and 3 above. The relevant probabilities can then be found in the tables of Williamson and Bretherton (1963).

5-3. Bayesian treatment of Gamma distributions

When dealing with the Poisson distribution in the previous section we were able to confine ourselves to well-known functions. However, this is not the case when we apply Bayesian methods to the Gamma distributions.

Let us first consider the case of an unknown order parameter, i.e. the case of $Ga(a,b)$ with $a = \theta$. From (5-1.6) we see that a kernel of the likelihood is

$$(5\text{-}3.1) \qquad \beta(\theta)\exp[Q_1(x)R_1(\theta)] = \frac{b^{\theta}\exp(\theta \log x)}{\Gamma(\theta)} = \frac{(bx)^{\theta}}{\Gamma(\theta)} .$$

We already know that $a > 0$, so, if we further constrain θ to take integer values only, a possible prior p.m.f. of the same form as (5-3.1) would be

$$(5\text{-}3.2) \qquad g_o(\theta|t_o) = \frac{t_o^{\theta-1}\exp(-t_o)}{(\theta-1)!} \quad \text{for } \theta = 1,2,\ldots \text{ and } t_o > 0.$$

In other words, we may assume initially that our unknown parameter is distributed as $1 + P_o(t_o)$.

It is tempting to suppose that sampling will result in a closed system in which the unknown parameter always has a displaced Poisson distribution, but this proves not to be the case.

To demonstrate quickly what the actual system is, we introduce a new function which we call the epsilon function. We shall denote the epsilon function of order n by ϵ_n, which we use

in the definition:

$$(5\text{-}3.3) \qquad \epsilon_n(x) = \sum_{j=1}^{\infty} \frac{x^{j-1}}{[(j-1)!]^{n+1}} \quad \text{or} \quad \sum_{i=0}^{\infty} \frac{x^i}{[i!]^{n+1}}$$

$$\text{for } n = 0,1,2,\ldots$$

$$\text{and } x > 0.$$

It can be seen from the definition in (5-3.3) that the epsilon function of order zero, $\epsilon_0(x)$, is identical to the exponential function, e^x or $\exp(x)$.

The epsilon function of order one, $\epsilon_1(x)$, is related to the modified Bessel function of order zero, $I_0(x)$, by

$$(5\text{-}3.4) \qquad \epsilon_1(x) = I_0(2x^{\frac{1}{2}}).$$

Using the epsilon function as a normalising constant, we can define a p.m.f. based on the individual terms of the series. We shall say that a r.v. is distributed as Eps(n,λ) if its p.m.f.is

$$(5\text{-}3.5) \qquad f_{\text{Eps}}(x|n,\lambda) = \frac{\lambda^{x-1}}{[(x-1)!]^{n+1}} \; \frac{1}{\epsilon_n(\lambda)}$$

$$\text{for } x = 1,2,\ldots$$

$$n = 0,1,2,\ldots \text{ and } \lambda > 0.$$

The distribution of Eps $(0,\lambda)$ is therefore the same as that of $1 + P_0(\lambda)$, and we shall now consider the prior p.m.f. in (5.3.2) as being that of Eps $(0,t_0)$.

$$(5\text{-}3.6) \qquad h_o(x|t_o) = \sum_{\theta=1}^{\infty} f(x|\theta)g_o(\theta|t_o)$$

$$= \frac{be^{-bx}}{\epsilon_o(t_o)} \sum_{\theta=1}^{\infty} \frac{(bt_ox)^{\theta-1}}{[(\theta-1)!]^2}$$

$$= be^{-bx} \frac{\epsilon_1(bt_ox)}{\epsilon_o(t_o)} .$$

Defining $t_n = x_n t_{n-1}$ for $n = 1,2,\ldots,$ we can then find the posterior p.m.f for the unknown parameter after one observation conditional upon the sufficient statistic t_1,

$$(5\text{-}3.7) \qquad g_1(\theta|t_1) = \frac{(bt_1)^{\theta-1}}{[(\theta-1)!]^2} \quad \frac{1}{\epsilon_1(bt_1)}$$

$$= f_{Eps}(\theta|1,bt_1) .$$

By observing the behaviour of the system of functions as sampling proceeds, we find that the posterior p.m.f. for the unknown parameter after n observations is

$$(5\text{-}3.8) \qquad g_n(\theta|t_n) = \frac{(b^n t_n)^{\theta-1}}{[(\theta-1)!]^{n+1}} \quad \frac{1}{\epsilon_n(b^n t_n)}$$

$$= f_{Eps}(\theta|n,b_n t_n),$$

and so we have closure under sampling.

The marginal p.d.f. always comes out in the form of an exponential p.d.f. multiplied by a ratio of epsilon functions.

$$(5\text{-}3.9) \qquad h_n(x|t_n) = be^{-bx} \frac{\epsilon_{n+1}(b^{n+1}t_n x)}{\epsilon_n(b^n t_n)} \,.$$

We shall call a distribution with this kind of p.d.f. an Epsilon-Gamma mixture. Members of this family of distributions will be specified by the three parameters (n,b,t_n) and we may write the typical r.v. for the family as EpsGa(n,b,t_n), whose p.d.f. is that shown in (5-3.9).

Feller (1966) suggested mixing a Gamma distribution with what he calls a Bessel distribution, which corresponds to our distribution of Eps $(1,\cdot)$, but as he was not then looking for closure under sampling he did not discover the closed system which we have described here. It is also of interest to note that Poisson mixing of a Chi-squared (hence Gamma) distribution results in the Noncentral Chi-squared distribution of statistical theory - see Lehmann (1959). By introducing another parameter, corresponding to the degrees of freedom for Noncentral χ^2, so that the order parameter of Ga(a,b) was expressed as $a = k + \theta$, we could build up a more generalised closed system.

We look next at Ga(a,b) with the scale parameter, b, unknown. From (5-1.7) we see that a kernel of the likelihood is

$$(5\text{-}3.10) \qquad \beta(\theta) \exp\left[Q_1(x)R_1(\theta)\right] = \frac{\theta^a e^{-\theta x}}{\Gamma(a)} \,,$$

which becomes the p.d.f. of Ga(a,1) if we also incorporate the factor $\alpha(x) = x^{a-1}$. We may enrich the family of possible distributions by incorporating a scale parameter c, which may take any positive value.

Remembering that the sum of x's will be sufficient for the unknown $b = \theta$, we can then form the prior p.d.f. using an initial value of the sum statistic, s_o, which we introduced in the previous section.

$$(5\text{-}3.11) \qquad g_o(\theta|s_o) = \frac{s_o^c \theta^{c-1} e^{-\theta s_o}}{\Gamma(c)} .$$

Proceeding as before we find

$$(5\text{-}3.12) \qquad h_o(x|s_o) = \int_o^\infty f(x|\theta)\, g_o(\theta|s_o)\, d\theta$$

$$= \frac{\Gamma(a+c)}{\Gamma(a)\Gamma(c)} \frac{s_o^c\, x^{a-1}}{(s_o+x)^{a+c}} \int_o^\infty \frac{(s_o+x)^{a+c}\, \theta^{a+c-1}\, e^{-\theta(x+s_o)}}{\Gamma(a+c)}\, d\theta$$

$$= \frac{1}{B(a,c)} \frac{s_o^c\, x^{a-1}}{(s_o+x)^{a+c}} ,$$ where $B(.,.)$ is the complete beta function.

This is the p.d.f. of an Inverted Beta distribution. The typical r.v. with this type of distribution is IBe (p,q,s), whose p.d.f. is

$$(5\text{-}3.13) \qquad f_{IBe}(x|p,q,s) = \frac{1}{B(p,q)} \frac{s^q\, x^{p-1}}{(s+x)^{p+q}} .$$

The special case when $p = m/2$, $q = n/2$ and $s = n/m$ is Snedecor's well-known F-distribution with m degrees of freedom in the numerator and n in the denominator,

Sampling leads to a closed system with the following posterior and marginal p.d.f's.

$$(5\text{-}3.14) \qquad g_n(\theta|s_n) = \frac{s_n(s_n\theta)^{na+c-1}e^{-\theta s_n}}{\Gamma(na+c)} = f_{Ga}(\theta|na+c, s_n).$$

$$(5\text{-}3.15) \qquad h_n(x|s_n) = \frac{1}{B(a,na+c)} \quad \frac{s_n^{na+c}\, x^{a-1}}{(s_n+x)^{na+a+c}}$$

$$= f_{IBe}(x|a,na+c,s_n).$$

Scarf (1960c) makes use of the fact that

$$(5\text{-}3.16) \qquad f_{IBe}(x|p,q,s) = \frac{1}{s} f_{IBe}\left(\frac{x}{s}|p,q,1\right),$$

which enables him to eliminate dependence upon s_n and so reduce the dimensionality of the functionalequations for the periodic-review inventory model when demand is distributed as Ga(a,b) with unknown b. Scarf's idea was later used by Fukuda (1960) and Wilson (1966).

Raiffa and Schlaifer (1961) give more details about the Inverted Beta distributions (which they call inverted-beta-2 distributions). For our present purposes it suffices to say that the d.f. of IBe(p,q,s) is related to that of Beta r.v., Be(p,q) defined in Chapter 7 below, by

$$(5\text{-}3.17) \qquad F_{IBe}(x|p,q,s) = F_{Be}\left(\frac{x}{s+x}|p,q\right)$$

and that tables of F_{Be} are available - see Pearson (1956) and Schlaifer (1969).

We now examine the Bayesian treatment of the Gamma

distribution with both parameters unknown. Here we obtain a closed system of distributions if we choose the joint prior distribution from a family based upon what we shall call the psi function. We shall denote the psi function of order n by ψ_n, which we use in the definition:

$$(5\text{-}3.18) \qquad \psi_n(a,b) = \sum_{j=1}^{\infty} \frac{(nj)!}{a^{nj}} \frac{b^{j-1}}{[(j-1)!]^{n+1}}$$

for n = o,1,2,...

and a,b > o.

Comparison with (5-3.3) shows that $\psi_o(a,b) = \epsilon_o(b)$ and so $\psi_o(a,b) = \exp(b)$ for any value of a.

We now use the psi function to define the Psi distribution, a bivariate distribution for which the joint p.m.f. is given by the following expression.

$$(5\text{-}3.19) \qquad f_{Psi}(x,y|n,s,t) = \frac{s}{\psi_n(s,t)} \frac{t^{x-1}y^{nx}e^{-ys}}{[(x-1)!]^{n+1}}$$

The likelihood for each observation of $Ga(\theta,\phi)$ is

$$(5\text{-}3.20) \qquad f(x|\theta,\phi) = \frac{\phi^{\theta} x^{\theta-1} e^{-\phi x}}{(\theta-1)!} .$$

If we choose the prior joint mass and density function for the unknown parameters to be $f_{Psi}(\theta,\phi|o,s_o,t_o)$, we find that we have a system which is closed under sampling with the following posterior and marginal distributions, in which s_n is the sum statistic and t_n the product statistic as before.

$$(5\text{-}3.21)\qquad g_n(\theta,\phi|s_n,t_n) = \frac{s_n}{\psi_n(s_n,t_n)}\ \frac{t_n{}^{\theta-1}\phi^{n\theta}e^{-\phi s_n}}{[(\theta-1)!]^{n+1}}$$

$$= f_{Psi}(\theta,\phi|n,s_n,t_n)\,.$$

$$(5\text{-}3.22)\qquad h_n(x|s_n,t_n) = \frac{s_n}{s_n+x}\ \frac{\psi_{n+1}(s_n+x,t_n x)}{\psi_n(s_n,t_n)}$$

$$= f_{PsiGa}(x|n,s_n,t_n),\ \text{say}\,.$$

CHAPTER 6

THE NORMAL AND LOGNORMAL DISTRIBUTIONS

6-1. Definitions and inventory applications

The Normal (or Gaussian) distribution is the distribution of $No(\mu,\sigma^2)$, which has the following well-known probability density.

$$(6\text{-}1.1) \qquad f_{N_o}(x|\mu,\sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\tfrac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right] \text{ for } -\infty < x < \infty.$$

The parameters μ and σ^2 represent the mean and variance respectively. In the Bayesian analysis which follows, it is convenient to replace the factor $1/\sqrt{2\pi}$ by the equivalent expression $(1/2)^{\frac{1}{2}} / \Gamma(1/2)$.

A random variable whose natural logarithms is distributed as $No(\mu,\sigma^2)$ is said to have a Lognormal distribution or be distributed as $LNo(\mu,\sigma^2)$. A simple change of variable shows that

$$(6\text{-}1.2) \qquad f_{LN_o}(x|\mu,\sigma^2) = \frac{1}{x} f_{N_o}(\log x|\mu,\sigma^2). \quad \text{for } x > o.$$

Thus, while the Normal distribution is defined on the whole real axis the Lognormal is defined on positive values only. For this reason the Lognormal appears a more natural distribution to choose for essentially non-negative quantities such as demand and leadtime: Brown (1959) has in fact shown that it gives a good description of demand in many cases. However, as we remarked in Chapter 5, the two-parameter Gamma family is rich enough to provide distributions which closely match most members of the two-parameter Lognormal family - and for many purposes it is simpler to use a Gamma distribution.

On the other hand, it might be argued that the Lognormal may be used as an approximation to the Gamma, and the Lognormal methods outlined in this chapter may be considered as alternative to those considered in Section (5-3). The Normal distribution itself does not give a good approximation to Gamma distributions of low or moderate order but it is, of course, often used to approximate Poisson or Binomial distributions under suitable conditions: the principal condition is that the mean be reasonably large.

In the absence of any other information, the Normal distribution may therefore be used to describe demand when this occurs at a fairly high rate. Tables for minimising the loss in a single stage of an inventory process with known Normal demand are given by Eilon (1962). The analysis for some more complicated Normal models is given by Hadley and Whitin (1963), and we have already mentioned the continuous-time model of Bather (1966).

The Bayesian treatment of univariate Normal distributions from the point of view of parameter estimation has been well covered by many authors; in particular we refer to Lindley (1965) and to Raiffa and Schlaifer (1961), who also cover the multivariate Normal. The theory has also been widely applied to the more general problem of estimating unknown states of a system in stochastic control theory - see Aoki (1967).

In the next section we introduce the distributions which are required as conjugates when the various Normal parameters are unknown. We then present the standard results for the self-reproducing distributions of the unknown parameters and add those for the marginal distributions which are of interest in our inventory applications.

6-2. The conjugate distributions

We start by introducing the Inverted Gamma distribution (the inverted - gamma - 1 of Raiffa and Schlaifer). If a random variable X is distributed as Ga(a,b), then its reciprocal will be distributed as IGa(a,b), whose p.d.f. is

$$(6\text{-}2.1) \qquad f_{IGa}(x|a,b) = \frac{b^a (1/x)^{a+1} e^{-b/x}}{\Gamma(a)} \qquad \text{for } x > 0.$$

IGa(a,b) has mean $b/(a-1)$ and variance $b^2/[(a-1)^2(a-2)]$.

Next, we introduce the generalised Student distribution with parameters (m,s^2,ν). The typical r.v. for this is $St(m,s^2,\nu)$ whose p.d.f. is as follows

$$(6\text{-}2.2) \qquad f_{St}(x|m,s^2,\nu) = \frac{\nu^{\nu/2} s^{-1}}{B(1/2,\nu/2)} \left[\nu + \left(\frac{x-m}{s}\right)^2\right]^{-(\nu+1)/2}$$

for $-\infty < x < \infty$,

where $s, \nu > 0$.

The mean is m for $\nu > 1$ and the variance is $\nu s^2/(\nu-2)$ for $\mu > 2$.

If X is distributed as $St(m,s^2,\nu)$, we find that the r.v. T obtained by the transformation $T = (X-m)/s$ has the classical t-distribution with ν degrees of freedom, which is the distribution of our $St(0,1,\nu)$. If tables of the t-distribution are available, we can therefore find fractiles for $St(m,s^2,\nu)$ by using the relationship

$$(6\text{-}2.3) \qquad F_{St}(x|m,s^2,\nu) = F_{St}\left(\frac{x-m}{s}\,|\,0,1,\nu\right).$$

It is worth noting that the generalised Student distribution defined here is not the same as the Noncentral t-distribution of classical statistics, whose p.d.f. is

$$(6\text{-}2.4) \qquad f_{Nct}(x|\delta,\nu) = \int_0^\infty (y/\nu)^{\frac{1}{2}} f_{No}(x(y/\nu)^{\frac{1}{2}}|\delta,1) f_{Ga}(y|\nu/2,\tfrac{1}{2})dy.$$

When δ, the noncentrality parameter, is zero, this p.d.f. reduces to that of the ordinary t-distribution with ν degrees of freedom. However, the cumulative Noncentral t-distribution cannot be used to find fractiles for $St(m,s^2,\nu)$ when $\delta \neq 0$.

Raiffa and Schlaifer (1961) have shown that the generalised Student distribution arises as what we would call an Inverted-Gamma mixture of a Normal distribution.

$$(6\text{-}2.5) \qquad f_{St}(x|m,s^2,\nu) = \int_0^\infty f_{No}(x|m,ys^2)\, f_{IGa}(y|\nu/2,1/2)dy.$$

Because of this, we shall find that the generalised Student distribution plays an important role in the Bayesian treatment of Normal distributions with unknown variance.

The next distribution is the bivariate Conjugate Normal distribution, for which the joint p.d.f. is

$$(6\text{-}2.6) \qquad f_{CNo}(x,y|m,v,a,b)$$

$$= \frac{a^{\frac{1}{2}}(bv)^{b/2}}{\Gamma(1/2)\Gamma(b/2)} (\tfrac{1}{2})^{(b+1)/2} (\tfrac{1}{y})^{(b+3)/2} \exp\{-\tfrac{1}{2}[bv+a(x-m)^2]/y\}.$$

We shall need this when both parameters of the Normal distribution are unknown.

Finally, when dealing with the Lognormal distribution we shall need the LogStudent distribution, whose p.d.f. is

$$(6\text{-}2.7) \qquad f_{LSt}(x|m,s^2,\nu) = \frac{1}{x} f_{St}(\log x|m,s^2,\nu).$$

6-3. Bayesian theory for Normal distributions

We start by considering $No(\mu,\sigma^2)$ when μ is unknown. The conjugate prior p.d.f. is then

$$(6\text{-}3.1) \qquad g_o(\theta|m_o) = f_{N_o}(\theta|m_o,\sigma^2/a_o).$$

Multiplying $f(x|\theta)$ by $g_o(\theta|m_o)$, we find that the terms can be regrouped into the product of two Normal densities. Only one of these factors depends on θ and this becomes unity when integrated over all θ, so that the marginal p.d.f. is equal to the other factor.

$$(6\text{-}3.2) \qquad h_o(x|m_o) = f_{N_o}(x|m_o,(1+a_o)\sigma^2/a_o).$$

By defining a_n to be $1+a_{n-1}$, or simply $n+a_o$, and introducing a kind of sample mean,

$$(6\text{-}3.3) \qquad m_n = \frac{a_{n-1}m_{n-1} + x_n}{a_n} \qquad \text{for } n > 0,$$

we then find that our choice of prior distribution results in the following closed system.

$$(6\text{-}3.4) \qquad g_n(\theta|m_n) = f_{N_o}(\theta|m_n,\sigma^2/a_n),$$

$$(6\text{-}3.5) \qquad h_n(x|m_n) = f_{N_o}(x|m_n,a_{n+1}\sigma^2/a_n).$$

Next we consider $N_o(\mu,\sigma^2)$ with σ^2 unknown. In this case the conjugate distribution is an Inverted Gamma and we select

(6-3.6) $$g_o(\theta|w_o) = f_{IGa}(\theta|b_o/2, b_o w_o/2).$$

Remembering that $\pi^{-\frac{1}{2}}$ is the same as $\Gamma(1/2)$, we can factorise $f(x|\theta)\ g_o(\theta|w_o)$ into the product of an Inverted Gamma p.d.f. and a Student p.d.f., which gives us the marginal density.

(6-3.7) $$h_o(x|w_o) = f_{St}(x|\mu, w_o, b_o).$$

As sampling proceeds we get the following closed system:

(6-3.8) $$g_n(\theta|w_n) = f_{IGa}(\theta|b_n/2, b_n w_n/2),$$

(6-3.9) $$h_n(x|w_n) = f_{St}(x|\mu, w_n, b_n),$$

where b_n is $1 + b_{n-1}$, or simply $n + b_o$, and

(6-3.10) $$w_n = \frac{b_{n-1} w_{n-1} + (x_n - \mu)^2}{b_n}.$$

To give an intuitive meaning to the statistics, we can think of w_n as the estimated variance after n observations and b_n as the number of degrees of freedom in this estimate.

When both parameters of $No(\mu, \sigma^2)$ are unknown, our prior distribution is Conjugate Normal.

(6-3.11) $$g_o(\theta, \phi|m_o, v_o) = f_{CNo}(\theta, \phi|m_o, v_o, a_o, b_o).$$

We define a_n, b_n and m_n as before and introduce another statistic which may be thought of as a kind of sample variance based on the sample mean,

$$(6\text{-}3.12) \qquad v_n = \frac{b_{n-1}\, v_{n-1} + a_{n-1}\, m_{n-1}^2 + x_n^2 - a_n m_n^2}{b_n} .$$

In the closed system which results under sampling, we then have

$$(6\text{-}3.13) \qquad g_n(\theta,\phi|m_n,v_n) = f_{CNo}(\theta,\phi|m_n,v_n,a_n,b_n),$$

$$(6\text{-}3.14) \qquad h_n(x|m_n,v_n) = f_{St}(x|m_n,a_{n+1}v_n/a_n,b_n).$$

The results for $LN\ddot{o}(\mu,\sigma^2)$ follow directly from the theory for $No(\mu,\sigma^2)$. When μ is unknown we get

$$(6\text{-}3.15) \qquad h_n(x|m_n') = f_{LNo}(x|m_n',a_{n+1}\sigma^2/a_n),$$

where m_n' is defined to be the right-hand side of (6-3.3) with x_n replaced by $\log x_n$. If we similarly define w_n', v_n' to be w_n, v_n with x_n replaced by $\log x_n$, we get

$$(6\text{-}3.16) \qquad h_n(x|w_n') = f_{LSt}(a|\mu,w_n',b_n)$$

when σ^2 is unknown and

$$(6\text{-}3.17) \qquad h_n(x|m_n',v_n') = f_{LSt}(x|m_n',a_{n+1}v_n'/a_n,b_n)$$

when both μ and σ^2 are unknown.

CHAPTER 7

THE BINOMIAL AND BETA DISTRIBUTIONS

7-1. Definitions and inventory applications

The Binomial distribution is that of Bi(n,p), whose p.m.f. is

$$(7\text{-}1.1) \qquad f_{Bi}(x|n,p) = \binom{n}{x} p^x (1-p)^{n-x}$$

$$\text{for } 0 < p < 1,$$

$$n = 1,2,\ldots,$$

$$\text{and } x = 0,1,\ldots,n.$$

The mean and variance of Bi(n,p) are np and np(1-p) respectively. Being defined on non-negative integers, this distribution is suitable for describing demand or leadtime. The Binomial family is closed under convolutions and so may be useful for providing demand distributions in models with leadtime.

A related family of distributions is the Negative Binomial, whose typical r.v. is NBi(r,p) which has p.m.f.

$$(7\text{-}1.2) \qquad f_{NBi}(x|r,p) = \binom{r+x-1}{x} (1-p)^r p^x$$

$$\text{for } 0 < p < 1,$$

$$r > 0$$

$$\text{and } x = 0,1,2,\ldots .$$

The mean of NBi(r,p) is $rp/(1-p)$ and the variance is $rp/(1-p)^2$.

We have already seen, in Chapter 5, how distributions of this family arise when a Poisson distribution is mixed by a Gamma, but we may consider Negative Binomial distributions as important in their own right since in many ways they provide discrete analogues of the Gamma distributions. By a suitable choice of the two parameters of NBi(r,p), we can get a good approximation to many practical demand distributions. The Negative Binomial family, like the Binomial, is closed under convolutions.

When p is unknown, the distributions of both Bi(n,p) and NBi(r,p) can be shown to belong to the Exponential Family, so we shall be able to find Bayesian distribution systems which are closed under sampling.

Connected with the two families just mentioned is the family of Beta distributions, referred to in Chapter 5. The typical r.v. here is Be(a,b) whose p.d.f. is

$$(7\text{-}1.3) \qquad f_{Be}(x|a,b) = \frac{1}{B(a,b)} x^{a-1}(1-x)^{b-1}$$

$$\text{for } 0 < x < 1 \text{ and } a,b > 0.$$

The mean is $a/(a+b)$ and the variance is $ab/[(a+b)^2(a+b+1)]$. Raiffa and Schlaifer (1961) show that the Beta p.d.f. varies greatly in shape as the parameters are changed.

It is well known that Beta distributions for activity time are used in the PERT technique for critical path analysis. Although this use has been questioned, e.g. by MacCrimmon and Ryavec (1964), it seems that a suitably rescaled Beta distribution might give a useful description of demand or leadtime when a maximum value is known.

Beta distributions belong to the Exponential Family when either or both of the parameters are unknown, but we shall see that the Bayesian treatment requires us to define some new functions.

7-2. Bayesian theory for Binomial distributions

Positive Binomial distributions have probably been more widely studied from a Bayesian point of view than those of any other family except the Normal. Numerous applications have been suggested, e.g. by Barnard (1954) in adaptive quality control and by Sutton and Tomlinson (1965) in adaptive control of a Ward-Leonard system.

Briefly, the theory is that when Bi(a,p) has an unknown parameter $p = \theta$, we select a member of the Beta family as the prior distribution for this parameter. We have called the first parameter a here to avoid confusion with n used for sample size.

$$(7\text{-}2.1) \quad g_o(\theta|s_o) = f_{Be}(\theta|s_o, b-s_o)$$

$$= \frac{1}{B(s_o, b-s_o)} \theta^{s_o-1}(1-\theta)^{b-s_o-1} .$$

This leads to a system which is closed under sampling with the following posterior density, which depends upon the sum statistic, s_n.

$$(7\text{-}2.2) \quad g_n(\theta|s_n) = \frac{1}{B(s_n, na+b-s_n)} \theta^{s_n-1}(1-\theta)^{na+b-s_n-1}$$

$$= f_{Be}(\theta|s_n, na+b-s_n).$$

The marginal p.m.f. is that of a Negative Hypergeometric or Beta Binomial distribution and may be written as follows:

$$(7\text{-}2.3) \qquad h_n(x|s_n) = \frac{\binom{s_n+x}{x}\binom{na+a+b-s_n-x-1}{a-x}}{\binom{na+a+b-1}{a}}$$

$$\text{or} \quad \frac{a}{x(a-x)} \; \frac{B(a,na+b)}{B(x,s_n)B(a-x,na+b-s_n)} \; .$$

Raiffa and Schlaifer (1961) devote a section to the computation of the d.f. of this distribution, for which tables are not yet available.

The theory for Negative Binomial distributions is very similar. We consider NBi(a,p) with $p = \theta$ and choose a Beta prior distribution, so that

$$(7\text{-}2.4) \qquad g_o(\theta|s_o) = \frac{1}{B(s_o,b)} \; \theta^{s_o-1}(1-\theta)^{b-1} \; .$$

This leads to a closed system with Beta posterior distributions:

$$(7\text{-}2.5) \qquad g_n(\theta|s_n) = \frac{1}{B(s_n,na+b)} \; \theta^{s_n-1}\,(1-\theta)^{na+b-1} \; .$$

The marginal distribution does not appear to have been named previously, but we shall call it a Beta Negative Binomial distribution. Its mass function is

$$(7\text{-}2.6) \qquad h_n(x|s_n) = \frac{1}{x} \; \frac{B(s_n+x,na+a+b)}{B(x,a)B(s_n,na+b)} \; .$$

7-3. Bayesian theory for Beta distributions

Apparently Bayesian methods have not been applied before to Beta distributions, but we shall see that the theory has various points of interest. As in Chapter 5, we need to define some new functions for a concise presentation.

First we define the eta function. The eta function of order n is denoted by η_n and defined by

$$(7\text{-}3.1) \qquad \eta_n(a,b) = \sum_{j=1}^{\infty} a^{j-1}\left[\frac{(j+b-1)!}{(j-1)!}\right]^n .$$

The individual terms of the series may then be used to define the Eta distribution, i.e. the distribution of Eta(n,a,b) whose p.m.f is

$$(7\text{-}3.2) \qquad f_{\text{Eta}}(x|n,a,b) = \frac{a^{x-1}}{\eta_n(a,b)}\left[\frac{(x+b-1)!}{(x-1)!}\right]^n$$

for $0 < a < 1$ and b, $x = 1,2,\ldots$.

Now we consider Be(a,b) with the first parameter unknown, so that

$$(7\text{-}3.3) \qquad f(x|\theta) = \frac{1}{B(\theta,b)}\, x^{\theta-1}(1-x)^{b-1}$$

$$= \frac{(\theta+b-1)!}{(\theta-1)!(b-1)!}\, x^{\theta-1}(1-x)^{b-1} \quad \text{if } \theta \text{ and b are integers.}$$

By choosing the prior distribution for the unknown parameter to be that of Eta(o,t_o,b), where t_o is a prior value of the product statistic, we get

$$(7\text{-}3.4) \quad g_o(\theta|t_o) = \frac{t_o^{\theta-1}}{\eta_o(t_o,b)} = t_o^{\theta-1}(1-t_o),$$

since $\eta_o(t_o,b)$ is simply a geometric series in powers of t_o. Under sampling, this results in a closed system where the posterior and marginal distributions at each stage depend upon the product statistic t_n.

$$(7\text{-}3.5) \quad g_n(\theta|t_n) = \frac{t_n^{\theta-1}}{\eta_n(t_n,b)} \left[\frac{(\theta+b-1)!}{(\theta-1)!}\right]^n .$$

$$(7\text{-}3.6) \quad h_n(x|t_n) = \frac{(1-x)^{b-1}}{(b-1)!} \frac{\eta_{n+1}(t_n x,b)}{\eta_n(t_n,b)} .$$

We shall call this h_n the p.d.f. of EtaBe 1(n,t_n,b), the typical r.v. of an Eta Beta One distribution.

An interesting special case arises when the parameter b = 1. The posterior and marginal distributions in this case are related to the moments of the Poisson distribution. If $m_k(\lambda)$ denotes the (k) th moment about the origin of Po(λ), then by choosing

$$(7\text{-}3.7) \quad g_o(\theta|t_o) = \frac{t_o^{\theta-1} e^{-t_o}}{(\theta-1)!} \text{ for } \theta = 1,2,\ldots$$

we find that (7-3.5) and (7-3.6) are replaced by the following:

$$(7\text{-}3.8) \quad g_n(\theta|t_n) = \frac{\theta^n t_n^{\theta-1}}{(\theta-1)!} \frac{t_n \exp(-t_n)}{m_{n+1}(t_n)} ,$$

$$(7\text{-}3.9)\qquad h_n(x|t_n) = \frac{\exp[-t_n(1-x)]}{x}\ \frac{m_{n+2}(t_n x)}{m_{n+1}(t_n)}\ .$$

We shall say that distributions with p.m.f's in the form of $g_n(\theta)$ in (7-3.8) are Poisson Moment distributions of order n + 1 with parameter t_n. Random variables with such a distribution will be said to be distributed as PoMo$(n+1,t_n)$ for positive integer n.

It can be shown that if $M_k(n,t)$ denotes the (k)th moment about the origin of PoMo(n,t), then this is related to the moments of Po(t) by

$$(7\text{-}3.10)\qquad M_k(n,t) = \frac{m_{n+k}(t)}{m_n(t)}\ .$$

The first ten moments of Po(λ) are given by Haight (1967).

The analysis is much the same when it is the second parameter of Be(a,b) which is unknown. Given an initial value, τ_o, we define a new statistic

$$(7\text{-}3.11)\qquad \tau_n = \tau_{n-1}(1-x_n).$$

Choosing the distribution of Eta (o,τ_o,a) as prior for the unknown b = θ, we get a closed system with:

$$(7\text{-}3.12)\qquad g_n(\theta|\tau_n) = \frac{\tau_n^{\theta-1}}{\eta_n(\tau_n,a)}\ \left[\frac{(\theta+a-1)!}{(\theta-1)!}\right]^n\ ,$$

$$(7\text{-}3.13)\qquad h_n(x|\tau_n) = \frac{x^{a-1}}{(a-1)!}\ \frac{\eta_{n+1}(\tau_n(1-x),a)}{\eta_n(\tau_n,a)}\ .$$

We shall call this h_n the p.d.f. of EtaBe2(n,τ_n,a), which is the same as that of EtaBel(n,τ_n,a) but with 1-x in place of x.

To study Be(a,b) with both parameters unknown, say $a = \theta$ and $b = \phi$, we define another new function. The xi function of order n , denoted by ξ_n, is defined by

$$(7\text{-}3.14)\quad \xi_n(a,b) = \sum_{j=1}^{\infty}\sum_{k=1}^{\infty} a^{j-1}\, b^{k-1} \left[\frac{1}{B(j,k)}\right]^n$$

for $a,b > 0$.

Using the individual terms, we define the bivariate Xi distribution, for which the joint p.m.f. is

$$(7\text{-}3.15)\quad f_{Xi}(x,y|n,a,b) = \frac{a^{x-1}b^{y-1}}{\xi_n(a,b)} \left[\frac{1}{B(x,y)}\right]^n$$

for $x,y = 1,2,\ldots$.

By putting

$$(7\text{-}3.16)\quad g_o(\theta,\phi|t_o,\tau_o), = f_{Xi}(\theta,\phi|0,t_o,\tau_o),$$

we get the following closed system:

$$(7\text{-}3.17)\quad g_n(\theta,\phi|t_n,\tau_n) = \frac{t_n^{\theta-1}\tau_n^{\phi-1}}{\xi_n(t_n,\tau_n)} \left[\frac{1}{B(\theta,\phi)}\right]^n$$

$$= f_{Xi}(\theta,\phi|t_n,\tau_o) \quad ,$$

$$(7\text{-}3.18)\quad h_n(x|t_n,\tau_n) = \frac{\xi_{n+1}(t_n x,\tau_n(1-x))}{\xi_n(t_n,\tau_n)} \quad .$$

CHAPTER 8

THE UNIFORM DISTRIBUTIONS

8-1. Definitions and inventory applications

To illustrate the Bayesian treatment of distributions belonging to the Range Family, we close our survey of useful univariate distributions with a look at uniformly distributed continuous and discrete random variables. In what follows, we shall reserve the term Uniform distribution for the discrete case and use the name Rectangular for the continuous case.

We define the Rectangular distribution on (a,b) to be that of Ra(a,b), which has p.d.f.

$$(8\text{-}1.1) \qquad f_{Ra}(x|a,b) = \frac{1}{b-a} \quad \text{for } a < x < b.$$

The Uniform distribution on $\{a,a+1,\ldots,b\}$ is that of Un(a,b), which has p.m.f.

$$(8\text{-}1.2) \qquad f_{Un}(x|a,b) = \frac{1}{b-a+1} \quad \text{for } x = a,a+1,\ldots,b.$$

Referring back to Section (7-1), we see that the distribution of Ra(o,1) is the same as that of Be(1,1), but our main interest in this chapter is in cases where one or other of the parameters of Ra(a,b) or Un(a,b) is unknown and also where both are unknown. We shall call a the lower limit and b the upper limit for both distributions.

In the early days of Bayesian statistics, it was usual to choose a Rectangular distribution as the prior in order to reflect vague prior knowledge about an unknown parameter - see Lindley (1965). If the unknown parameter was p of Bi(n,p) and the prior distribution was that of Ra(o,1), which is also a Beta distribution and hence a conjugate, a closed system resulted under sampling. In other cases, Rectangular distributions will not be self-reproducing.

Apart from a brief mention by Spragins (1965) and another by Raiffa and Schlaifer (1961), there does not seem to be much literature about Bayesian methods applied to the Rectangular and Uniform distributions when their own parameters are unknown.

Because of the simple form of the p.d.f. of Ra(a,b) and the p.m.f. Un(a,b), these distributions find numerous applications in all branches of probability theory. We have already noted in Section (2-3) that taking a time average of time-dependent costs is equivalent to mixing the demand distribution with a Uniform or Rectangular distribution, and similar distributions can be used to describe demand - e.g. Wagner (1962) works out the stationary operating characteristics of (S,s) models with Uniform demand.

8-2. The Rectangular distribution

We start by considering Ra(a,b) with an unknown lower limit, so that $a = \theta$ and we get

$$(8\text{-}2.1) \qquad f(x|\theta) = \frac{1}{b-\theta} \quad \text{for } \theta < x < b .$$

By choosing a value A_o as the prior maximum value of the lower limit, we can then define the following prior p.d.f.

$$(8\text{-}2.2) \qquad g_o(\theta|A_o) = \frac{k(b-A_o)^k}{(b-\theta)^{k+1}} \quad \text{for } -\infty < \theta < A_o \text{ where } k = 1,2,\ldots .$$

This leads to the following closed system, in which $A_n = \min\{A_{n-1}, x_n\}$.

$$(8\text{-}2.3) \qquad g_n(\theta|A_n) = \frac{(k+n)(b-A_n)^{k+n}}{(b-\theta)^{k+n+1}} \quad \text{for } -\infty < \theta < A_n ,$$

$$(8\text{-}2.4) \qquad h_n(x|A_n) = \begin{cases} \dfrac{(k+n)}{(k+n+1)} \dfrac{(b-A_n)^{k+n}}{(b-x)^{k+n+1}} & \text{for } x < A_n , \\ \dfrac{(k+n)}{(k+n+1)} \dfrac{1}{(b-A_n)} & \text{for } A_n < x < b. \end{cases}$$

The marginal p.d.f. therefore increases with x until it reaches a maximum at A_o; then it remains constant until x = b. As n increases, the left-hand tail of h_n shrinks so that the distribution becomes effectively rectangular.

To obtain the simple forms of distribution above, we have assumed that the unknown limit may be negative, and hence our basic random variable may be negative. For inventory applications with

positive r.v's, the numerator on the right-hand side of (8-2.2) and (8-2.3) will have to be modified.

When the top limit is unknown, we select for it a prior minimum value B_o and the following p.d.f.

$$(8\text{-}2.5) \qquad g_o(\theta|B_o) = \frac{k(B_o-a)^k}{(\theta-a)^{k+1}} \qquad \text{for } \theta > B_o \text{ and } k = 1,2,\ldots .$$

Defining $B_n = \max\{B_{n-1}, x_n\}$ we obtain the following closed system.

$$(8\text{-}2.6) \qquad g_n(\theta|B_n) = \frac{(k+n)(B_n-a)^{k+n}}{(\theta-a)^{k+n+1}} \qquad \text{for } \theta > B_n .$$

$$(8\text{-}2.7) \qquad h_n(x|B_n) = \begin{cases} \dfrac{(k+n)}{(k+n+1)} \dfrac{1}{(B_n-a)} & \text{for } a < x < B_n, \\ \dfrac{(k+n)}{(k+n+1)} \dfrac{(B_n-a)^{k+n}}{(x-a)^{k+n+1}} & \text{for } x > B_n. \end{cases}$$

The marginal p.d.f. is now constant from a to B_n and then decreases with x. Here the right-hand tail shrinks as n increases.

When both limits are unknown we select A_o and B_o as before and choose a prior joint p.d.f. of the form:

$$(8\text{-}2.8) \qquad g_o(\theta,\phi|A_o,B_o) = \frac{k(k-1)(B_o-A_o)^{k-1}}{(\phi-\theta)^{k+1}}$$

$$\text{for } \theta < A_o,\ \phi > B_o \text{ and } k = 2,3,\ldots .$$

This leads to the following closed system.

$$(8\text{-}2.9)\qquad g_n(\theta,\phi|A_n,B_n) = \frac{(k+n)(k+n-1)(B_n-A_n)^{k+n}}{(\phi-\theta)^{k+n+1}}$$

$$\text{for } \theta < A_n,\ \phi > B_n.$$

$$(8\text{-}2.10)\qquad h_n(x|A_n,B_n) = \begin{cases} \dfrac{(k+n-1)}{(k+n+1)}\dfrac{(B_n-A_n)^{k+n+1}}{(B_n-x)^{k+n+1}} & \text{for } x < A_n, \\ \dfrac{(k+n-1)}{(k+n+1)}\dfrac{1}{B_n-A_n} & \text{for } A_n < x < B_n, \\ \dfrac{(k+n-1)}{(k+n+1)}\dfrac{(B_n-A_n)^{k+n-1}}{(x-A_n)^{k+n}} & \text{for } x > B_n. \end{cases}$$

Here h_n is in the form of a rectangle with tails on either side. As n increases, both tails will shrink until the p.d.f. becomes completely rectangular.

8-3. The Uniform discrete distribution

The Bayesian treatment of Un(a,b) is simply the discrete analogue of the treatment of Ra(a,b) which we described in Section (8-2). By a suitable choice of g_o and values of A_o and/or B_o we obtain a closed system in which the marginal distributions are uniform with appropriate tails. Again the tails shrink away as n increases.

An interesting feature of the analysis is that the posterior and marginal distributions are connected with the Riemann zeta function. This function with argument k, is defined by

$$(8\text{-}3.1) \qquad \zeta(k) = \sum_{x=1}^{\infty} x^{-k} \quad \text{for } k > 1.$$

A general account of the theory of the zeta function is given by Titchmarsh (1951), and tables have been prepared by Haselgrove and Miller (1963). In the collection of papers on discrete distributions edited by Patil (1965), there is a two-page account by P.R.Rider of a distribution based on the individual terms of the series in (8-3.1). Rider is dubious about practical applications of such a distribution, but we shall show that in a truncated form it provides conjugates for the Uniform distribution.

First let us introduce the truncated zeta function of order k truncated at b , which is defined by

$$(8\text{-}3.2) \qquad \zeta(k,b) = \sum_{x=b}^{\infty} x^{-k} \quad \text{for } b = 1,2,\ldots \text{ and real } k > 1.$$

We then define the Zeta distribution to be that of a random variable Ze(k,b) which has p.m.f.

$$(8\text{-}3.3) \qquad f_{Ze}(x|k,b) = \frac{x^{-k}}{\zeta(k,b)} \quad \text{for } x = b,b+1,\ldots .$$

Rider's Zeta distribution is then that of our Ze(k,1).

To illustrate the Bayesian application, consider the distribution of Un(1,b) where we consider the unknown b as equal to θ, so that

$$(8\text{-}3.4) \qquad f(x|\theta) = \frac{1}{\theta} \quad \text{for } x = 1,2,\ldots,\theta.$$

Select B_o as a prior minimum for θ and let

$$(8\text{-}3.5) \qquad g_o(\theta|B_o) = f_{Ze}(\theta|k,B_o) \text{ for } \theta \geqslant B_o.$$

Under sampling, we obtain a closed system with:

$$(8\text{-}3.6) \qquad g_n(\theta|B_n) = f_{Ze}(\theta|k+n,B_n) \text{ for } \theta \geqslant B_n.$$

$$(8\text{-}3.7) \qquad h_n(x|B_n) = \begin{cases} \dfrac{\zeta(k+n+1,B_n)}{\zeta(k+n,B_n)} & \text{for } x = 1,2,\ldots,B_n, \\ \dfrac{\zeta(k+n+1,x)}{\zeta(k+n,B_n)} & \text{for } x > B_n. \end{cases}$$

Bearing in mind the way in which increasing k accelerates the convergence of $\zeta(k)$, it can readily be seen that, after a moderate-sized sample has been taken, the marginal distribution will become effectively Uniform with $1/B_n$ at each mass point.

CHAPTER 9

SURVEY AND CONCLUSION

9-1. Selection from a family of conjugate distributions

In the last four chapters we have looked at several examples of how to construct families of conjugate prior distributions based on the kernel of the likelihood. In some cases, such as $Ga(\theta,b)$ in Section (5-3), we obtained the conjugate prior by simply normalising the kernel. In other cases we introduced another parameter to enrich the family of conjugates - for instance, this was the reason for introducing a general scale parameter b in $Ga(s_o,b)$ used as the prior for $Po(\theta)$ in Section (5-2) rather than just having a unit scale (b=1).

We now consider how to choose a particular prior distribution from the conjugate family in a way which will reflect the appropriate degree of vagueness or precision of our initial knowledge about the unknown parameter. Raiffa and Schlaifer (1961) have pointed out that vagueness must be expressed by selecting a prior distribution which can be "substantially modified by a small number of sample observations" rather than by considering the shape of the prior distribution as expressed by its moments. The expression of vagueness therefore reduces to the selection of prior parameters which will be substantially modified by sampling. Thus in Chapter 5 we should choose s_o and t_o (the prior values of the sum and product "statistics") as small as possible for extreme vagueness, and the same applies to the parameters b and c (in the priors for Poisson and Gamma distributions respectively) since these parameters simply have a known constant added to them at each stage of sampling. A sensible practical rule seems to be to choose the initial value to be 1 when the choice is restricted to positive

integers and 0.1 when positive fractions are allowed, though in the latter case it may often be more convenient to choose an integer rather than a fraction so that simpler tables may be used. Thus, in Section (7-2) we expressed the marginal p.m.f. for $Bi(a,\theta)$ and $NBi(a,\theta)$ in terms of beta functions but it may be more convenient to find factorials or combinatorials. Similarly, if we are dealing with $Ga(\theta,b)$ and wish to use the short tabulation of the epsilon function in Appendix 2 then we would choose $t_0 = 1$ rather than 0.1.

In dealing with the Rectangular and Uniform distributions we should set A_0 as high as possible and B_0 as low as possible to reflect extreme vagueness.

For any of the distributions which we have considered, any departure from the values which do reflect extreme vagueness can be considered as equivalent to using a number of fictitious observations to modify a vague prior distribution before sampling actually starts.

We have advocated using Bayesian methods when very little sample information is available. In this case observations would not be greatly affected by any slow drift in the value of an unknown parameter. With larger samples such drift can be dealt with by using a type of exponential smoothing to update the "statistics" and so submerge the prior value and early sample values more quickly than would be the case with the simple addition (or multiplication) operations for updating which we have described - see Fu(1968).

We close this section with a few words about the new functions which arise in conjugate distributions. The epsilon and truncated zeta functions involve only two variables and can therefore be tabulated in

a reasonably compact form. Since the series which define these functions converge quite rapidly, direct computation is quite feasible when the functions are required for Bayesian analysis but a good quick guide to their behaviour can be obtained by studying the tables in Appendix 2.

The most obvious characteristic of the epsilon function is the way in which $\epsilon_n(x)$ decreases very rapidly when n is increased and x held constant; this is a natural consequence of the definition with powers of factorials in the denominator. Similarly, increasing the power (k) of the reciprocal terms in $\zeta(k)$ causes rapid convergence so that, when we truncate the lower end of the series, $\zeta(k,b)$ rapidly becomes zero as k increases.

In Section (7-3) we have already seen that the eta function of order zero is simply a geometric series and it can also be shown that eta functions of order one can be expressed in terms of negative binomial series, e.g. $\eta_1(a,1) = (1-a)^{-2}$, $\eta_1(a,2) = 2(1-a)^{-3}$, $\eta_1(a,3) = 6(1-a)^{-4}$, so that these can be evaluated directly. The fact that $\eta_2(a,1) = (1+a)(1-a)^{-3}$ and $\eta_2(a,2) = 4(1-a)^{-3}[1+6a(1-a)^{-2}]$ suggests that similar expressions can be found for eta functions of higher orders.

9-2. Further work

The results which we have obtained for the Bayesian treatment of demand distributions with unknown parameters are applicable in other fields where we are interested in updating the distribution of the observed variable. Such a situation might arise, for example, in reliability work where we observe the lifetime of components subject to failure. Gamma distributions are often used for lifetimes and the theory of Chapter 5 would therefore be relevant. Lifetimes are also commonly described by Weibull distributions, which belong to the Exponential Family if one parameter is unknown and yield a closed system under sampling if a Gamma prior is used.

In Chapter 6 we mentioned the Bayesian theory for multivariate Normal distributions. Another useful topic in multivariate Bayesian theory has recently been studied by Martin (1967), who shows that Bayesian methods can be used for estimating Markov transition probabilities. Theoretically this technique could be used in conjunction with the Markov inventory model for correlated demands, to which we referred in Section (3-3), but the computations involved for this model are already quite formidable without further complications. However, the actual adaptive process, which involves the use of a Matrix Beta prior, is quite straightforward. Murphy (1965) has suggested that related Bayesian adaptive methods applied to the Multinomial distribution may have applications in econometrics.

Returning to inventory theory, it does now seem reasonable to think of actually using Bayesian methods with the models which we considered in Chapters 2 and 3. It is well known that a great hindrance

to implementing modern systems of stock control is the inadequacy of existing demand records, which all too often omit lost sales and only show captive demands or despatches. However, as records become computerised, there would seem to be a case for the stock-control computer to have a Bayesian package which could build up demand distributions for important stock items.

Even with high-speed computers, it seems unlikely that on-line dynamic programming will ever be feasible for regular reordering of a large number of items in a store - even if our Scheme B (the differential scheme) is used. However, this scheme is very useful for looking at the way in which y_n^* converges to y^* and hence determining when an infinite-duration solution may reasonably be used for a finite-duration process. This is an area in which further numerical work would be very rewarding. From the computational point of view it is preferable to concentrate on discrete demand distributions but, as we have shown, the Poisson is not the only discrete distribution worth considering. Of course with discrete demands the critical stock levels will also be discrete and we can see that y_n^* converges to y^* precisely when $\Delta\Lambda_n(y^*-1)$ becomes negative, whereas for continuous demands y_n^* simply creeps closer and closer to y^* as n increases.

When we think of using discrete demand distributions to approximate continuous ones (or vice versa) it is perhaps natural to choose distributions with the same means and variances. However, when we consider that our critical stock levels are determined by the fractiles of the distribution and usually by the shape of its right-hand tail, it becomes clear that the skewness of the distributions must also be considered. Some preliminary calculations suggest that better approximations are obtained by considering equal means and third moments rather than means and variances.

9-3. Conclusion

We have come full circle, back to the calculation of optimal ordering policies, and the important results of this thesis can now be seen in perspective.

Firstly we developed a scheme for calculating optimal ordering policies which eliminated much of the redundant work involved in standard schemes. As a by-product of this, we obtained a clearer picture of the relationship between inventory processes with finite and infinite numbers of stages.

We then went on to develop the theory for stochastic leadtimes when more than one order could be outstanding and related this to the existing theory for deterministic leadtimes.

In the second part of the thesis, we examined the implications of unknown parameters in the demand distribution when Bayesian methods are used. The result of this was to reveal a number of new properties of well known basic distributions when these were examined from the point of view of their conjugates. To define the conjugate distributions we introduced a number of interesting new functions which should now be included in the theory of the various basic distributions.

The Bayesian inventory methods we have considered are recommended when information about demand is too vague to discern any trend or correlation. If these features become apparent as the inventory process unfolds, then more appropriate methods are available. However, if the demand distribution appears to be stationary and uncorrelated then our Bayesian methods can still be used.

In developing the idea of likelihood kernels, we touched on the subject of totally positive kernels, which Karlin had introduced into inventory theory to enable more general forms of loss function to be used. Although the theory of these general forms is interesting, the simple linear cost structure upon which we have concentrated will surely remain most important in practice.

Looking back over all the ideas which we have considered, it is clear that much more still remains to be done on periodic-review inventory theory and that developments in this subject are closely linked to advances in other fields of study.

APPENDIX 1

CONVERGENCE OF CRITICAL STOCK LEVELS

Tables A.1-1 and A.1-2 show examples of the way in which y_n^* converges to y^* for some discrete inventory models of the type described in Section (2-2). Using Scheme B, which we developed in that section, the key equations for calculating y_n^* are:

$$(1) \qquad \Delta\Lambda_1(y) = (c_P - c_R) + (c_H + c_R)F(y),$$

$$(2) \qquad \Delta\Lambda_n(y) = -\alpha c_P + \Delta\Lambda_1(y) + \alpha \sum_{r=0}^{y-y^*_{n-1}} \Delta\Lambda_{n-1}(y-r)p(r)$$

for $n > 1$.

For all the calculations, c_H was set at 2.0 and α at 0.9 while the mean (μ) and the purchasing and runout costs were varied as shown. The results for the Negative Binomial distribution in Table A.1-2 and many of those for the Poisson in Table A.1-1 are based on computations done by Miss Gillian M. Bennett (now Mrs. Baynes) as part of an undergraduate project under the supervision of the author.

The tables show that the number of stages to convergence does not vary uniformly with c_R, c_P or μ. For example, in Table A.1-1 we see that with demand distributed as Po(2) and $c_R = 50$ we get $y_n^* = y^*$ for $n = 2$ with $c_P = 10$ or 20 but we have to go to $n = 3$ with $c_P = 5$. Moving across to the distribution of Po(6), y_n^* converges at $n = 2$ for $c_P = 5$ or 10 but at $n = 3$ for $c_P = 20$.

However, two general tendencies can be observed with each of the distributions. First, if the costs are held constant then an increase in the mean will tend to accelerate convergence. Second, if μ and all the costs except c_P are held constant then convergence will slow down as c_P is increased. This second tendency may be explained by reference to the two equations quoted above. Consider the equation for $\Delta\Lambda_2(y)$ and substitute $\Delta\Lambda_1(y-r)$ for $\Delta\Lambda_{n-1}(y-r)$ in the final term, which we shall call the discounted partial convolution of $\Delta\Lambda_1$ with p. If c_P is much smaller than c_R then the first term of $\Delta\Lambda_1(y-r)$, viz. (c_P-c_R), will be very negative and the total value of the partial convolution will be small or even negative so that if $\Delta\Lambda_1(y)$ is considerably less than αc_P for $y < y^*$ we shall find that the effect of the partial convolution is insufficient to make $\Delta\Lambda_2(y)$ positive: hence y_2^* will equal y^* and the process will converge in 2 stages. As the value of c_P is increased towards c_R, the partial convolution becomes large enough to make $\Delta\Lambda_2(y)$ positive when $y < y^*$, and so convergence is delayed.

TABLE A.1-1

CONVERGENCE WITH POISSON DEMAND

Table shows values of c_P which give convergence at the stage indicated.

An entry such as 90-94 denotes similar behaviour for c_P = 90,91,92,93,94.

c_R (stage)	$\mu = 2$	$\mu = 6$	$\mu = 8$	$\mu = 10$
(2)	10,20	5,10,25	5,10,20,25	5,10,25,41
50 (3)	5,25,40	20		42-49
(4)	41-49			
(2)	5	12	12,30,55,56	12,50,51
60 (3)	10,12,30,50	30,55,56,59	59	30,52-59
(4)	51-59			
(2)	5	20	65	58,60
70 (3)	10,58,60	65,69	69	61-69
(4)	61-69			
(2)	10	20	75,79	70
80 (3)	70	75,79		71-79
(4)	71-79			
(2)	10	20	50,60,83-89	80
90 (3)	80	80	70,80-82	81-89
(4)	81-89			
(2)	10	10	95-99	87,88
100 (3)	87,88,90	20,90	90-94	90-99
(4)	91-99			

TABLE A.1-2

CONVERGENCE WITH NEGATIVE-BINOMIAL

AND UNIFORM DEMAND

Table shows nunber of stages to convergence with demand distributed as NBi($\mu,\frac{1}{2}$) and Un(0,2μ).

mean	c_R=60, c_P=55		c_R=50, c_P=25		c_R=50, c_P=15	
μ	NBi	Un	NBi	Un	NBi	Un
1.5	4	3	4	2	3	2
2.0	4		4		3	
2.5	4	3	4	2	3	2
3.0	4		4		3	
4.0	3	3	3	2	3	2
5.0	4		4		3	
7.0	3	3	3	2	3	2
10.0	3		2		2	

APPENDIX 2

THE EPSILON AND TRUNCATED ZETA FUNCTIONS

Table A.2-1 shows values of the epsilon function which we defined in (5-3.3), where we remarked that $\epsilon_0(x) = \exp(x)$ and $\epsilon_1(x) = I_0(2x^{\frac{1}{2}})$. Standard tables of the exponential and modified Bessel functions were accordingly used to check the calculated values in the first two columns of the table. Values were computed from the defining series, which was reckoned to have converged when the last term was less than 5×10^{-8} times the sum including that term. The number of terms required increased with x but decreased markedly as the order of the epsilon function was raised, so that no more than 5 terms were needed for any of the tabulated values of $\epsilon_5(x)$.

The truncated zeta function in Table A.2-2 was computed by subtracting the early terms from the sums of reciprocal powers tabulated by Abramowitz and Stegun (1964)*. Our table shows clearly how the marginal p.m.f. h_n defined in (8-3.7) will soon become effectively Uniform because the ratio $\zeta(k+1,b+1)/\zeta(k,b)$ can be seen to decrease as k and b increase and so the right-hand tail of h_n (for $x > B_n$) will shrink as n increases.

* ABRAMOWITZ M. and STEGUN I.A.(1964) Handbook of Mathematical Functions, National Bureau of Standards,Washington.D.C.

TABLE A.2-1 : EPSILON FUNCTION

x	$\epsilon_0(x)$	$\epsilon_1(x)$	$\epsilon_2(x)$
1	2.71828	2.27959	2.12970
2	7.38906	4.25235	3.53821
3	20.0855	7.15900	5.25600
4	54.5982	11.3019	7.31542
5	148.413	17.0578	9.75077
6	403.429	24.8921	12.5984
7	1096.63	35.3766	15.8967
8	2980.96	49.2086	19.6863
9	8103.08	67.2344	24.0102
10	22026.5	90.4760	28.9136
11	59874.1	120.161	34.4442
12	162755	157.760	40.6523
13	442413	205.022	47.5906
14	1202604	264.029	55.3149
15	3269017	337.242	63.8835
16	8886111	427.564	73.3576
17	2.42×10^7	538.410	83.8018
18	6.57×10^7	673.784	95.2833
19	1.78×10^8	838.369	107.873
20	4.85×10^8	1037.63	121.645

TABLE A.2-1 : EPSILON FUNCTION (continued)

x	$\epsilon_3(x)$	$\epsilon_4(x)$	$\epsilon_5(x)$
1	2.06327	2.03138	2.01565
2	3.25622	3.12603	3.06267
3	4.58358	4.28473	4.14120
4	6.05016	5.50826	5.25137
5	7.66085	6.79740	6.39331
6	9.42061	8.15294	7.56714
7	11.3345	9.57566	8.77299
8	13.4076	11.0664	10.0110
9	15.6451	12.6258	11.2813
10	18.0522	14.2549	12.5840
11	20.6344	15.9543	13.9192
12	23.3970	17.7248	15.2871
13	26.3456	19.5674	16.6879
14	29.4857	21.4827	18.1215
15	32.8230	23.4717	19.5882
16	36.3631	25.5350	21.0881
17	40.1121	27.6736	22.6214
18	44.0756	29.8883	24.1881
19	48.2599	32.1798	25.7883
20	52.6708	34.5490	27.4223

TABLE A.2-2 : TRUNCATED ZETA FUNCTION

b	$\zeta(2,b)$	$\zeta(3,b)$	$\zeta(4,b)$
1	1.644934	1.202057	1.082323
2	.644934	.202057	.082323
3	.394934	.077057	.019823
4	.283823	.040020	.007478
5	.221323	.024395	.003571
6	.181323	.016395	.001971
7	.153545	.011765	.001200
8	.133137	.008850	.000783
9	.117512	.006897	.000539
10	.105166	.005525	.000387
11	.095166	.004525	.000287
12	.086902	.003774	.000218
13	.079957	.003195	.000170
14	.074040	.002740	.000135
15	.068938	.002375	.000109
16	.064494	.002079	.000089
17	.060588	.001835	.000074
18	.057127	.001631	.000062
19	.054041	.001460	.000053
20	.051271	.001314	.000045

TABLE A.2-2 : TRUNCATED ZETA FUNCTION (continued)

b	$\zeta(5,b)$	$\zeta(6,b)$	$\zeta(7,b)$
1	1.036928	1.017343	1.008349
2	.036928	.017343	.008349
3	.005678	.001718	.000537
4	.001563	.000346	.000080
5	.000586	.000102	.000018
6	.000266	.000038	.000006
7	.000137	.000017	.000002
8	.000078	.000008	.000001
9	.000047	.000004	
10	.000030	.000003	
11	.000020	.000002	
12	.000014	.000001	
13	.000010	.000001	
14	.000007		
15	.000006		
16	.000004		
17	.000003		
18	.000003		
19	.000002		
20	.000002		

BIBLIOGRAPHY

AOKI M. (1967) Optimization of stochastic systems
Academic Press: New York.

ARROW K.J., HARRIS T.E. and MARSCHAK J. (1951)
"Optimal inventory policy". Econometrica 19,pp.250-272.

ARROW K.J., KARLIN S. and SCARF H.E. (eds) (1958)
Studies in the mathematical theory of inventory and production. Stanford Univ. Press; Stanford, Calif.

BARLOW R.E and PROSCHAN F. (1965) Mathematical theory of reliability. Wiley; New York.

BARNARD G.A. (1954) "Sampling inspection and statistical decisions". Jl R. Statist. Soc. (Series B), 16,pp.151-166.

BATHER J.A. (1966) "A continuous time inventory model".
J. Appl. Prob. 3, pp.538-549.

BELLMAN R. (1957) Dynamic programming. Princeton Univ. Press; Princeton, N.J.

BELLMAN R. (1961) Adaptive control processes: a guided tour.
Princeton Univ. Press; Princeton, N.J.

BOX G.E.P. and JENKINS G.M. (1968) "Some recent advances in forecasting and control, Part I".
Applied Statistics, 17, pp.91-109.

BROWN R.G. (1959) Statistical forecasting for inventory control.
McGraw-Hill; New York.

CHURCHMAN C.W., ACKOFF R.L. and ARNOFF E.L. (eds) (1957)
Introduction to operations research. Wiley; New York.

DVORETZKY A., KIEFER J. and WOLFOWITZ J. (1952)
"The inventory problem, II. Case of unknown distribution of demand". Econometrica, 20, pp450-466.

EGGENBURGER F. and PÓLYA G. (1923) "Ueber die Statistik verketteter Vorgaenge". Z.angew.Math.Mech.,3, pp.279-289.

EILON S. (1962) Industrial engineering tables.
Van Nostrand; London.

EILON S. and LAMPKIN W. (1968) Inventory control abstracts (1953-1965). Oliver and Boyd; Edinburgh.

FELLER W. (1966) An introduction to probability theory and its applications, Volume 2. Wiley; New York.

FERGUSON T. S. (1967) Mathematical Statistics: a decision theoretic approach. Academic Press; New York.

FU K.S. (1968) Sequential methods in pattern recognition and machine learning. Academic Press; New York.

FUKUDA Y. (1960) "Bayes and maximum likelihood policies for a multi-echelon inventory problem". PRC R-161. Planning Research Corpn; Los Angeles, Calif.

HADLEY G. and WHITIN T.M. (1963) Analysis of inventory systems. Prentice-Hall; Englewood Cliffs, N.J.

HAIGHT F.A. (1967) Handbook of the Poisson distribution. Wiley; New York.

HANSEL G. (1966) "Application d'une méthode bayésienne au calcul d'un volant de rechanges pour des pièces a faible consommation". Paper delivered at European Meeting of Statisticians; London.

HARRIS F. (1915) Operations and cost. Shaw; Chicago.

HASELGROVE C.B. and MILLER J.C.P. (1963) Tables of the Riemann zeta function. Cambridge Univ. Press; Cambridge.

IGLEHART D.L. (1964) "The dynamic inventory problem with unknown demand distributions". Management Science, 10, pp.429-440.

IGLEHART D.L. and KARLIN S. (1962) "Optimal policy for dynamic inventory process with nonstationary stochastic demands". Chapter 8 of K.J. ARROW et al. Studies in applied probability and management science. Stanford Univ. Press; Stanford, Calif.

KARLIN S. (1958a) "One stage inventory models with uncertainty". Chapter 8 of ARROW, KARLIN and SCARF (1958).

KARLIN S. (1958b) "Optimal inventory policy for the Arrow-Harris-Marschak dynamic model". Chapter 9 of ARROW, KARLIN and SCARF (1958).

KARLIN S. (1960a) "Optimal policy for dynamic inventory process with stochastic demands subject to seasonal variations". J.Soc. Indust. Appl.Math., 8, pp.611-629.

KARLIN S. (1960b) "Dynamic inventory policy with varying stochastic demands". Management Science, 6, pp.231-258.

KARLIN S. (1968) Total positivity, Volume 1. Stanford Univ. Press; Stanford, Calif.

KARLIN S. and SCARF H.E. (1958) "Inventory models of the Arrow-Harris-Marschak type with time lag". Chapter 10 of ARROW, KARLIN and SCARF (1958).

LAMPKIN W. (1967) "A review of inventory control theory". The Production Engineer, 46, pp.57-66.

LEHMANN E.L. (1959) Testing statistical hypotheses. Wiley; New York.

LINDGREN B.W. (1968) Statistical theory. (2nd edition) Macmillan; New York.

LINDLEY D.V. (1965) Introduction to probability and statistics from a Bayesian viewpoint. Part 2: inference. Cambridge Univ. Press; Cambridge.

MACCRIMMON K.R. and RYAVEC C.A. (1964) "An analytic study of the PERT assumptions". Operations Research, 12,pp.16-37.

MARTIN J.J. (1967) Bayesian decision problems and Markov chains. Wiley; New York.

MOLINA E.C. (1942) Poisson's exponential binomial limit. Van Nostrand; Princeton, N.J.

MORAN P.A.P. (1959) The theory of storage. Methuen; London.

MORAN P.A.P. (1968) An introduction to probability theory. Clarendon Press; Oxford.

MURPHY R.E. (1965) Adaptive processes in economic systems. Academic Press; New York.

NADDOR E. (1966) Inventory systems. Wiley; New York.

PARZEN E. (1960) Modern probability theory and its applications. Wiley; New York.

PARZEN E. (1962) Stochastic processes. Holden-Day; San Franciso.

PATIL G.P. (1965) Classical and contagious discrete distributions. Statistical Publ.Soc.; Calcutta.

PEARSON K. (1922) Tables of the incomplete Γ-function. H.M.S.O.; London.

PEARSON K. (1956) Tables of the incomplete beta-function. Cambridge Univ. Press; Cambridge.

PELLETIER R. (1962) Tables of the individual and cumulative terms of the Poisson distribution. Van Nostrand; Princeton, N.J.

PRABHU N.U. (1965) Queues and inventories. Wiley; New York.

RAIFFA H. and SCHLAIFER R. (1961) Applied statistical decision theory. M.I.T.Press; Cambridge, Mass.

SCARF H.E. (1959) "Bayes solutions of the statistical inventory problem". Annals of Mathematical Statistics, 30, pp.490-508.

SCARF H.E. (1960a) "The optimality of (S,s) policies in the dynamic inventory problem". Chapter 13 of K.J.ARROW et al. (eds) Mathematical methods in the social sciences, 1959. Stanford Univ. Press; Stanford, Calif.

SCARF H.E. (1960b) "Optimal policies for the inventory problem with stochastic lead time." PRC R-181. Planning Research Corpn; Los Angeles, Calif.

SCARF H.E. (1960c) "Some remarks on Bayes solutions to the inventory problem." Naval Research Logistics Quarterly, 7, pp.591-596.

SCARF H.E. (1963) "A survey of analytic techniques in inventory theory." Chapter 7 of H.E.SCARF et al. (eds). Multistage inventory models and techniques. Stanford Univ.Press; Stanford, Calif.

SCARF H.E. and VANDERVEER J. (1961) "A comparative analysis of the Bayes inventory policy". PRC R-201. Planning Research Corpn; Los Angeles, Calif.

SCHLAIFER R. (1969) Analysis of decisions under uncertainty McGraw-Hill; New York.

SCHOENBERG I.J. (1950) "On Pólya frequency functions, II". Acta Sci. Math. Szeged., 12, pp97-106.

SCHOENBERG I.J. (1951) "On Pólya frequency functions, I." Journal d'Analyse Mathematique, 1, pp.331-374.

SKELLAM J.G. (1952) "Studies in statistical ecology,I.Spatial pattern". Biometrika, 39, pp. 346-362.

SPRAGINS J. (1965) "A note on the iterative application of Bayes' rule". IEEE Transactions on Information Theory, IT-11, pp544-549.

SUTTON R.W. and TOMLINSON N.R. (1965) "Adaptive computer control of a simulated second-order system." Paper 18, Advances in Automatic Control (I.Mech.E./U.K.A.C.meeting); Nottingham.

THATCHER A.R. (1962) Some results on inventory problems. Jl R.Statist.Soc. (Series B), 24 pp.1-32. (+ Discussion pp.32-45).

TITCHMARSH E.C. (1951) The theory of the Riemann zeta-function. Clarendon Press; Oxford.

WAGNER H.M. (1962) Statistical management of inventory systems. Wiley; New York.

WHITE D.J. (1962) Studies in dynamic programming. (Ph.D.thesis) Dept. of Eng.Production, Univ.of Birmingham; Birmingham.

WHITIN T.M. (1953) The theory of inventory management (2nd edition). Princeton Univ.Press; Princeton, N.J.

WILLIAMSON E. and BRETHERTON M.H. (1963) Tables of the negative binomial probability distribution. Wiley; London.

WILSON D.A.L. (1966) The effect of parameter uncertainty upon inventory policy. (D.I.C.thesis) Dept.of Mech.Eng., Imperial College; London.