CHAPTER 9.

CONCLUSIONS.

9.1 Discussion

The aim of this thesis has been to study characterizations of discrete distributions based on the conditional distribution of the random variable Y for a fixed X, where X > Y. The damage model which is a particular case of the general model, has also been examined. Before the introduction of the Rao-Rubin characterization of the Poisson distribution, together with the damage model, most of the characterizations of discrete distributions in this field were based on the assumption of independence between Y and Z (Z=X-Y) (e.g. Moran (1952), Chatterji (1963), Patil and Seshadri (1964)).

Under this assumption these authors characterized the distributions of Y and Z by means of the distribution of $Y|X$.

Rao and Rubin gave a new dimension to the problem. They characterized the distribution of X by assuming that the distribution of $Y|X$ has a given form and that it satisfies what is now known as the R-R property.

Shanbhag's 1976 results provided a general characterization; this contains many results (including that of Rao-Rubin) as special cases. The proof of Shanbhag's results did not require the complicated mathematics that was needed for the Rao-Rubin proof.

The Rao-Rubin condition and its variants along with some new methods are the basic ideas that we have used and have developed in order to acquire meaningful characterizations for many well-known discrete distributions; these have also enabled us to make more general other characterizations that had been introduced already. The area of characterizations of truncated discrete distributions (which had previously received little attention) have been examined in some detail. Most of the results obtained, have been extended to the bivariate and multivariate cases.

In Chapter 2 we have given the full version of the elementary proof given by Shanbhag (1974). It has been said already, that the extension provided by Theorem 3.1.1 yields the R-R result as a special case. However, Shanbhag's elementary proof given in Theorem 2.1.1 remains interesting because it provides a nice way of solving the functional equation $G(q+t) = c\, G(t)$ in the case where $G(t)$ is a p.g.f. with $0 < q < 1$. We should mention here that Aczél (1975) derives the general solution of this functional equation under the assumptions that $G(1)=1$ and $G(t) > 0$ as

$$G(t) \;=\; c^{t|q} P(t), \quad t \in (-\infty,+\infty)$$

with $P(t)$ an arbitrary periodic function of period q.

Aczél then goes on to determine the solution of the same functional equation under the assumption that $G(t)$ is a log-convex function with $G(1)=1$; he obtains the solution

$$G(t) \;=\; c^{(t-1)|q}, \quad t \in (-\infty,+\infty),$$

which reduces to the solution that Rao-Rubin and Shanbhag obtained for $G(t)$ under the assumption that $G(t)$ is a p.g.f.

Aczél maintains that his assumptions (i.e. that $G(1)=1$ and $G(t)$ is log-convex) are weaker than the conditions imposed by Shanbhag, which in

Aczel's opinion are

(a) $G(1)=1$

(b) $G(t) > 0$ for all real t

(c) $G(t)$ analytic on the real line

(d) $\frac{d}{dt}\left(\frac{G'(t)}{G(t)}\right) \geqslant 0.$

It is clear that Aczél has overlooked the fact that Shanbhag, as indeed Rao and Rubin, was dealing with the problem of solving this functional equation on the sole assumption that $G(t)$ is a p.g.f., in which case (a) is automatically satisfied. On the other hand, (b), (c) and (d) are not conditions that have to be imposed, but properties that $G(t)$ possesses because it is a p.g.f. satisfying the given functional equation. These properties have been proved in Section 2.1.

(Aczél has replaced the fact that $G(t)$ can be differentiated any number of times and has a Taylor power series expansion everywhere converging to $G(t)$ by assumption (c).)

It is evident that Aczél is dealing with the problem of determining the general solution of $G(q+t) = c\,G(t)$ with $G(1)=1$ and $G(t)$ log-convex. This is different from the problem dealt with by Shanbhag and stated in Theorem 2.1.1, because the class of log-convex functions does not include all p.g.f.'s. This can be seen from the fact that there exist p.g.f.'s which are not log-convex. For example for the Binomial distribution, $G(t) = (q+pt)^n$ and hence

$$\frac{d^2}{dt^2} \log G(t) = \frac{d}{dt}\left(\frac{np}{pt+q}\right) = -\frac{np^2}{(pt+q)^2} < 0$$

if $t \neq -\frac{q}{p}$. We may observe that in the case where, $t = -\frac{q}{p}$, $G\left(-\frac{q}{p}\right) = 0$

and hence we can say immediately that the $G(t)$ does not satisfy the functional equation.

To give another example, where $G(t) > 0$ for all t, let us take $G(t) = p+qt^2$, $0 < p < 1$, $q = 1-p$. Clearly,

$$\frac{d}{dt} \log G(t) = \frac{2qt}{p+qt^2}$$

and

$$\frac{d^2}{dt^2} \log G(t) = \frac{2pq - 2q^2t^2}{(p+qt^2)^2} .$$

If we now consider these t's for which $t^2 > \frac{p}{q}$ we can see that $\frac{d^2}{dt^2} \log G(t) < 0$ and hence that G is not log-convex.

Wang (1975), also, in his attempt to present an alternative proof of Theorem 2.1.1 simpler than the one given by Rao and Rubin claims that because $\frac{G'(t)}{G(t)} = \frac{G'(t+kq)}{G(t+kq)}$ then $\frac{G'(t)}{G(t)} = k$. However, to claim this, he should have shown first (as we have done in Theorem 2.1.1) that $\frac{G'(t)}{G(t)}$ is a monotonic non-decreasing function; however Wang omitted to do this.

In Chapter 2, also, we started to examine the problem of characterizing the distribution of $Y|X$ when the distribution of X is truncated. Srivastava and Singh (1975) conjectured that if X is truncated Poisson then the R-R condition holds iff $P(Y=r|X=n)$ is "modified binomial". In Chapter 2 a counter-example was given in order to show that this conjecture was not justified. Srivastava and Singh were the first people to investigate characterizations of the conditional distribution of $Y|X$ (the "survival distribution" in terms of the damage model) based on the R-R condition, in the case where X follows a truncated distribution. Chapters 3, 4, 5 and 6 are devoted to this topic.

In Chapter 3 we reproduced on extension of the R-R characterization obtained by Shanbhag (1976). The importance of this extension lies in the fact that it characterizes a whole class of distributions by means of the R-R condition. This gave us the opportunity to obtain several characterizations of well-known distributions based on the R-R condition. It also helped in obtaining characterizations for the conditional distribution of $Y|X$ in relation to the R-R condition, when the form of the distribution of X is known. Again this is a new aspect of characterizations in this field. Up to now, only Srivastava and Srivastava (1970) have examined this situation, and then only for the particular case where X follows the Poisson distribution. To get the characterization in this case, they had to impose the restriction that the parameter λ of the Poisson is a variable. This restriction was lifted in Corollary 3.3.1 at the expense of another restriction concerning the form of the distribution of $Y|X$. In this area we made use of some well-known decomposition theorems.

Another interesting facet revealed by Theorem 3.1.1 is that the sole assumption that the distribution of $Y|X$ is of the form $a_r b_{n-r}/c_n$, the R-R condition implies that the random variables Y and X-Y are independent. In other words, it shows that independence of Y and X-Y over the set X-Y=0 implies complete independence for Y and X-Y. This, as it was pointed out in Section 3.2, makes the result of Theorem 3.1.1 more general than the one given by Patil and Seshadri (1964).

In Chapter 3 characterizations of truncated forms of the distribution of X through the R-R condition were also examined. The paper by Rao and Rubin (1964) seems to be the only previous work on characterizations from this point of view. As was pointed out in Chapter 2, they characterized the truncated Poisson distribution. Chapter 3 provides a general characterization for truncated forms of the distribution of X; several results are given as corollaries. Among them is the one for the truncated Poisson, the new proof of which is much simpler than the proof suggested by Rao and Rubin and given in Chapter 2. The question of extending Moran's (1952) result to the truncated case was also raised in Chapter 3, but as was shown this cannot be done. A counter-example was given showing that if the conditional distribution of $X|Y$ is truncated Binomial, then Y and X-Y can be of other forms, in addition to being Poisson and Truncated Poisson respectively.

Talwalker (1970) and Srivastava and Srivastava (1970) examined the R-R condition in the Bivariate case for the specific model in which the original is double Poisson and the survival is double Binomial. Chapter 4 of this thesis deals extensively with bivariate characterizations based on the R-R property. The importance of this Chapter is that it provides a characterization for a class of distributions, members of which are those characterized by Talwalker and Srivastava and Srivastava. It can also be observed that the theory by which the results of Chapter 4 were obtained is much simpler than that which has been used up to now and gives improved versions of known results.

It should be pointed out here, that the model we used throughout Chapter 4 is simpler than the one based on the "damage" concept used by some of the previous workers in the field. This is so, because, using the notation adopted by Rao and Rubin, the R-R condition for the damaged case, i.e. $P(Y_1=r_1, Y_2=r_2|\text{damaged})$ can be viewed as any one of the following three probabilities.

$$P(Y_1=r_1, Y_2=r_2 | X_1=Y_1, X_2>Y_2), \quad P(Y_1=r_1, Y_2=r_2 | X_1>Y_1, X_2=Y_2),$$
$$P(Y_1=r_1, Y_2=r_2 | X_1>Y_1, X_2>Y_2).$$

Only one of these three probabilities is needed to obtain the required result. It is interesting to see that all the results obtained in Chapter 3 and 4 remain valid even when the R.H.S. or the L.H.S. of the R-R condition as given there, is replaced by the $P(Y=r|X>Y)$ in the univariate case and the $P(Y_1=r_1, Y_2=r_2|X_1>Y_1, X_2=Y_2)$ in the bivariate one.

The results obtained in Chapter 3 are useful in extending Moran's result when the parameter p of the binomial (n,r,p) is independent of n. As has been stated already Moran (1952) has shown that, if X=Y+Z and Y,Z are independent then $P(Y=r|X=n) = \binom{n}{r} p^r q^{n-r}$ iff Y,Z are Poisson distributed. Patil and Seshadri obtained the same result as a particular case of their more general theorem. The above result can be further extended as follows.

Suppose that X, Y, Z are three non-negative, integer-valued random variables with X=Y+Z, and that the conditional distribution of Y|X is independent of the parameter of the distribution of X. Then,

(i) If either Y or Z follow Poisson distributions, and $P(Y=r|X=Y) = \binom{n}{r} p^r q^{n-r}$ $0<p<1$, $p+q=1$, then Y,Z are independent Poisson r.v.'s.

(ii) If $P(Y=r) = P(Y=r|X=Y)$ and $P(Y=r|X=n) = \binom{n}{r} p^r q^{n-r}$, then Y,Z are independent Poisson r.v.'s.

(iii) If (Y,Z) follows a Bivariate Poisson with p.g.f.

$$G(t_1,t_2) = e^{\lambda_1(t_1-1)+\lambda_2(t_2-1)+\lambda_{12}(t_1t_2-1)}$$

and $P(Y=r|X=Y) = P(Y=r)$, then

$$P(Y=r|X=n) = \binom{n}{r} p^r q^{n-r}.$$

Proof

(i) Because $Y|X=n \sim$ Binomial we have the joint p.g.f. of Y and Z as $G_{Y,Z}(t_1,t_2) = G_X(pt_1+qt_2)$. Without loss of generality we may assume that $Y \sim$ Poisson. Then $G_Y(t_1) = e^{-\lambda+\lambda t_1}$. We also have

$G_Y(t_1) = G_X(pt_1+q)$,

i.e. $G_X(pt_1+q) = e^{-\lambda+\lambda t_1}$

Consequently

$$G_{Y,Z}(t_1,t_2) = G_X(pt_1+qt_2) = G_X\{(pt_1+qt_2-q)+q\}$$

$$= e^{-\lambda+\frac{\lambda}{p}(pt_1+qt_2-q)} = e^{-\frac{\lambda}{p}+\frac{\lambda}{p}(pt_1+qt_2)},$$

i.e.

$$G_{Y,Z}(t_1,t_2) = G_Y(t_1)\,G_Z(t_2),$$

which implies that Y and Z are independent Poisson r.v.'s.

(ii) We have seen already that if $Y|X \sim$ Binomial and the R-R condition holds, then X is a Poisson r.v. and Y and Z are independent. Hence each follows a Poisson distribution.

(iii) Since (Y,Z) has a Bivariate Poisson distribution we find that

$$P(Y=y,Z=z) = e^{-(\lambda_1+\lambda_2+\lambda_{12})} \sum_{i=0}^{\min(y,z)} \frac{\lambda_1^{y-i}\lambda_2^{z-i}\lambda_{12}^{i}}{(y-i)!(z-i)!i!}.$$

Hence

$$P(Y=r|X=Y) = P(Y=r|Z=0) = \frac{P(Y=r,Z=0)}{P(Z=0)} = \frac{e^{-(\lambda_1+\lambda_2+\lambda_{12})}\frac{\lambda_1^r}{r!}}{e^{-(\lambda_1+\lambda_{12})}}$$

i.e. $P(Y=r|X=Y) = e^{-\lambda_1} \frac{\lambda_1^r}{r!}$.

However,

$$P(Y=r) = e^{-(\lambda_1+\lambda_{12})} \frac{(\lambda_1+\lambda_{12})^r}{r!}$$

and $P(Y=r) = P(Y=r|X=Y)$.

Hence $e^{\lambda_1(t-1)} = e^{(\lambda_1+\lambda_{12})(t-1)}$ which implies that

$\lambda_1=\lambda_1+\lambda_{12}$ i.e. $\lambda_{12}=0$.

As a result of this it is clear that Y, Z are independent Poisson r.v.'s, and so $Y|X \sim$ Binomial.

The Multivariate extension of Theorem 3.1.1 is given in Chapter 4; this is interesting for two reasons. Firstly, it examines in general the problem of characterizations based on a multivariate version of the R-R condition. Secondly, this condition is simpler than the corresponding one

$$(P(\underset{\sim}{Y}=\underset{\sim}{r}) = P(\underset{\sim}{Y}=\underset{\sim}{r}|\text{undamaged}) = P(\underset{\sim}{Y}=\underset{\sim}{r}|\text{damaged})$$

used by Talwalker to obtain a characterization of the Multiple Poisson distribution. Here also, with the exception of Talwalker, Chapter 4 makes the first attempt to study characterizations based on the R-R property in the Multivariate case. For the first time also characterization of the conditional distribution of $\underset{\sim}{Y}|\underset{\sim}{X}$ have been obtained.

The bivariate and multivariate extensions of Theorems 3.5.1, 3.6.1 and 3.6.2 (which in fact give 3.5.1 and 3.6.1 as special cases), proved particularly useful because they opened the way for characterizations of truncated bivariate and multivariate distributions and also for characterizations based on truncated forms of the conditional distribution of $\underset{\sim}{Y}|\underset{\sim}{X}$.

The characterization of the Negative Multinomial distribution given in Chapter 4 provides an extension of a result by Janardan (1974). He showed that if Y and Z are independent, then they have Negative Multinomial distributions iff the conditional distribution of $\underset{\sim}{Y}|\underset{\sim}{X}$ (where $\underset{\sim}{X}=\underset{\sim}{Y}+\underset{\sim}{Z}$) is Multivariate Inverse Hypergeometric. In fact the "if" part of this result can be derived by assuming that relation (4.5.8) holds. This assumption is clearly more relaxed than the assumption of independence of $\underset{\sim}{Y}$ and $\underset{\sim}{Z}$.

In the same paper Janardan derived the Multivariate extension of the Patil and Seshadri theorem. Comments similar to those relating to the univariate case (Chapter 3, Section 3.2) can be made concerning his extension and theorem 4.5.1.

In Chapter 5 we dealt with the problem of characterizations when the r.v. X is known to have a finite distribution. The main result provided by Theorem 5.1.1 is interesting, since it shows that, when the conditional distribution of Y|X has a particular form then the first ℓ probabilities of the distribution P_n, n=0,1,...,N are determined uniquely, the remaining N-m being arbitrary, provided that Y and X-Y are independent over the set X-Y=0,1,...,ℓ; ℓ=0,1,...,N-m, fixed. In the extreme case that ℓ=N-m, all the probabilities P_n, n=0,1,...,N are determined uniquely. On the other hand, the relation

$$P(Y=r|X=Y) = P(Y=r|X=Y+1) = \ldots = P(Y=r|X=Y+N-m) \qquad (9.1.1)$$

is equivalent to Y and X-Y being independent. This is so, because

$$P(Y=r|X=n) = \frac{a_r b_{n-r}}{c_n}$$

i.e.

$$P(Y=r,X-Y=n-r) = \begin{cases} \frac{p_n}{c_n} a_r b_{n-r} & r=0,1,\ldots,m,\ n-r=0,1,\ldots,N-m \\ 0 & r=m+1,\ldots,N,\ \ n-r=N-m+1,\ldots \end{cases}$$

Therefore,

$$P(Y \leqslant m, X-Y \leqslant N-m) = 1. \tag{9.1.2}$$

Relation (9.1.1) shows that (9.1.2) is equivalent to Y and X-Y being independent.

The situation with Y and X-Y independent has also been examined by Patil and Seshadri (see Chapter 3, Section 3.2). However, our set-up allows us to look at their problem from a different point of view. For finite forms of the distribution of X their result can be stated as follows:

if X and X-Y are independent then

$$P(Y=r|X=n) = \frac{a_r b_{n-r}}{c_n} \quad r=0,1,\ldots,n \tag{9.1.3}$$

iff

$$\frac{p_n}{c_n} = \frac{p_0}{c_0}\theta^n \quad \text{for some } \theta > 0 \quad n=0,1,\ldots,N. \tag{9.1.4}$$

Because of our earlier remarks, our findings in Chapter 5 – in the extreme case that ℓ=N-m – provide the following extension of Patil and Seshadri's result.

Suppose that the conditional distribution of Y|X is of the form (9.1.3). Then the distribution of X satisfies (9.1.4) iff Y and X-Y are independent. We have also seen that if (9.1.4) holds, then knowledge of X and independence of Y and X-Y do not always imply that the conditional distribution of $Y|X$ is uniquely determined. For example, while this is true when X is Binomial (in which case the Hypergeometric is the only possible form for the distribution of $Y|X$) it is not true if X is Negative Binomial (then $Y|X \sim$ Negative Hypergeometric is a solution but it is not the only solution).

In Chapter 5, the R-R condition was also examined for the situation when the distribution of $Y|X$ is Hypergeometric. Patil and Ratnaparkhi showed that if X is Binomial then the R-R condition is satisfied. Our Corollary 5.2.1 shows that a relation giving more information than the R-R condition is valid. On the other hand, a counter-example was given, showing that the R-R condition does not necessarily imply that X is Binomial. The condition

$$P(Y=r|X=Y) = P(Y=r|X=Y+1) = \ldots = P(Y=r|X=Y+N-m)$$

was shown to be the minimum required for this purpose. All the previously mentioned results were also extended to the truncated case.

Chapter 6 provided the Bivariate and Multivariate extension of the results obtained in Chapter 5, concerning finite distributions. The result obtained by Janardan (1974) about the Multinomial distribution has been extended. Comments similar to those made about the Negative Multinomial in Chapter 5, were also given.

In Chapter 7 we examined the damage model when the parameter of the Binomial survival is itself a variable, and when the parameter of the Poisson as original is a variable. General forms for the p.g.f. of the resulting distribution in general and the resulting distribution when no damage has occurred, were obtained in both cases. Relations between these p.g.f.'s were also derived. Various special cases were considered, and the corresponding distributions were studied. The interesting thing here was the effect that mixing in the original (or in the survival distribution) has upon the model.

In addition to the various special cases that were examined as corollaries in Chapter 7, the form of the resulting distribution can be obtained for many other cases. This is so, because the results obtained were for a general form of the mixing distribution. (Discrete or continuous.)

Another interesting property of the damage model was revealed in Chapter 8, where the effect of the convolution was studied. As it was shown, the p.g.f. of the resulting r.v. $Y = \sum_{i=1}^{s} Y_i$ is the product of the p.g.f.'s of the r.v.'s Y_i, provided that the X_i's are independent Poisson variables with parameters λ_i which themselves are variables, and the damaging processes in each case are Binomials with the same probability p of surviving. The more general case where there is a different p_ℓ of having $Y_\ell = r_\ell$ survivals, was also examined.

The characterizations of the Poisson distribution as well as that of the mixed Binomial, which we also derived in Chapter 8 (Theorems 8.2.1, 8.3.1), were very general and included many other results as special cases. On the other hand, Theorems 8.2.2 and 8.3.2 revealed the way things change if one considers both the original and the survival as mixed distributions.

Finally, an extension was obtained of the R-R characterization of the Poisson distributions based on a relation of the f.m.g.f. of the resulting r.v., and the p.g.f. of the undamaged resulting r.v. This result is important, because it uses the mixed Binomial as the survival distribution; it is the first attempt to extend the R-R characterization to the case where the parameter of the survival Binomial distribution is a random variable with a given probability function.

As a concluding remark, it can be said that the results obtained in this thesis provide a thorough investigation into characterization theory of discrete distributions truncated and untruncated in the Univariate and Multivariate cases. It can also be said that a gap which existed in the field has been filled by introducing characterizations of many well-known discrete distributions (e.g. Hypergeometric, Binomial, Negative Hypergeometric, Compound Poisson) which had received little attention up to now.

9.2 Statistical Importance of the Results.

9.2.1 Introduction.

In this Section we are going to examine the statistical significance of the results presented in this thesis. Emphasis will be placed on the connection with the R-R condition for the Poisson model. We will also discuss the contribution of Shanbhag's extension of the R-R characterization and will refer to the practical situations where the results concerning truncated distributions may be useful. The statistical applications of the results concerning bivariate and multivariate discrete distributions will be studied. Finally we will consider practical situations where the damage model has a compound form of original or survival distribution. Throughout this Section an attempt will be made to illustrate the contribution of the characterizations obtained to applied statistics.

9.2.2 The R-R Condition in Relation to the Poisson Model.

The Poisson distribution, on which Rao's (1963) and Rao and Rubin's (1964) papers are based, is, in the words of Fisher, of first importance among discontinuous distributions. It was first derived by S.D. Poisson (1781-1840) - see Haight (1967) - as the limit of a Binomial distribution.

Bortkiewicz - see again Haight (1967) - renewed interest in the Poisson distribution after it had been neglected for half a century and gave many elementary properties including difference and differential equations for the probabilities.

Since then it has been used to explain many physical phenomena. A comprehensive account on the history of the Poisson distribution and its applications can be found in Haight (1967).

The damage model introduced by Rao (1963) is one of the recent models in which the Poisson distribution plays an important role. This model is based on the assumption that an original observation produced by a natural process (e.g. number of eggs, number of accidents, etc.) may be partially destroyed or may be only partially ascertained. In such a case the original distribution will be distorted. Suppose that the model underlying the partial destruction of the original observation (i.e. the survival distribution) is known. Then we can derive the distribution appropriate to the observed values, knowing the original distribution.

The damage model theory can be employed to give further information in certain situations where the Poisson distribution has been used.

Consider, for example, the following situation studied by Feller (1957). Suppose that the number of eggs laid by an insect is Poisson distributed. Each egg has a probability p of developing. If we assume mutual independence of the eggs we find that the distribution of the survivors is Poisson with parameter λp (where λ is the parameter of the original Poisson distribution). The R-R condition tells us that the distribution of survivors is the same as the distribution of survivors coming from a sample of eggs all of which have developed; it is also the same as the distribution of survivors who come from a sample of eggs from which at least one has not developed. In addition, the R-R condition has another important consequence. It indicates that if any two of these three distributions are equal, then the only possible form of the distribution of the number of eggs laid by the insect is the Poisson form.

Another case where the damage model is applicable is that of chromosome breakages in cells produced by X-rays. Feller (1957) has observed that for a given dosage and time of exposure the number n of breakages in individual cells has a Poisson distribution with mean λ. Each breakage has a fixed probability q of healing, whereas with probability $p = 1 - q$, the cell dies. Here again the resulting distribution of the observable breakages is Poisson with mean λp. The R-R condition is applicable here; it shows that the distribution of the observable breakages, given that in all breakages the cell died, and the distribution of observable breakages, given that at least one breakage was healed, are also Poisson (λp).

The above two examples illustrate cases where the damage model theory is directly applied. However, there are other statistical problems which, although they do not seem, at first glance, to fall into this category, can also be studied in the light of the R-R condition and the related results. The following examples are indicative of this fact.

As pointed out by Haight (1967), Falechini (1949) applied the Poisson distribution to the number of University graduates. It is now reasonable to assume that the number of male graduates given the total number of graduates is Binomially distributed. Let p denote the probability of a male graduate. Then on the basis of the results examined we can safely infer that the distribution of male graduates and the distribution of female graduates are also Poisson with parameters λp and $\lambda(1-p)$ respectively (where λ is the parameter of the original Poisson distribution).

The R-R condition also indicates that the distribution of male graduates is the same as the distribution of graduates given that all the graduates are male. Moreover, the R-R characterization proves that if these last two distributions are identical, the Poisson is the only appropriate distribution for the number of graduates.

Kendall (1961) used the Poisson model to describe the number of strikes begun in a given week in the United Kingdom. The damage model theory can be utilized here for making inferences in the following way. Let us define as "short strikes" the strikes which end in the same week they started, and also define as "long strikes" those which go on in the following week(s). In the damage model set-up consider as original distribution the distribution of the number of strikes begun in a given week. Define as resulting distribution the distribution of short strikes. Assume that the distribution of the number of short strikes given the total number of strikes begun in a particular week is Binomial (p is the probability that a strike started in that week will be short). In such a situation we can deduce from the R-R condition that the distribution of the number of short strikes is the same as the distribution of the number of short strikes given that all strikes were short; it is also the same as the distribution of the number of short strikes given that there was at least one long strike.

Feller (1957) applied the Poisson distribution to the number of misprints. One can consider as original distribution the distribution of the number of misprints and as resulting distribution the distribution of the number of misprints spotted with a Binomial survival law.

In this case the R-R condition tells us that the distribution of the number of spotted misprints is the same as the distribution of the number of misprints in a sample where all the misprints have been spotted; it is also the same as the distribution of the number of misprints in a sample where we know that there are some misprints which have been missed.

Moreover, in the above two examples the R-R condition implies something more important. In fact it implies that if any two of the distributions mentioned are equal, then the original distribution can only be of a Poisson form.

Haight (1967) refers to another case where the Poisson distribution has been found to give a satisfactory fit. He indicates that Friedman (1956) has given a model in which the number of bidders for a contract is assumed to be Poisson. If one is interested in a more detailed study of the above model one mayconsider the distribution of the number of bidders for the contract as the resulting distribution. Then the distribution of the number of people who have originally shown interest in the contract will play the role of the original distribution.

Accident theory is an important field of statistics where the model related to the R-R condition may be useful. One can consider X to be the number of accidents in a given location and Y to be the number of fatal accidents. Alternatively, X can be assumed to be the number of accidents and Y the number of reported accidents. This latter model could be interesting in actuarial studies. Of course the R-R condition here suggests that the insurer will not be able to obtain any more information about the original distribution if he selects in his sample individuals who have reported all their accidents from the information he will obtain if he samples from people who had accidents but they have not necessarily reported all of them.

Another broad area of applied probability where our results may be useful, is the area of stochastic processes. The Poisson process is among the most important of these processes. Here again there are cases where the damage model theory may be adopted. The following examples are characteristic.

Consider the particles arriving at a Geiger counter (see Parzen (1962 Suppose that they arrive in accordance with a Poisson process $X(t)$ at a mean rate of λ per unit time. The Geiger counter activates a recording mechanism containing a defective relay which operates correctly with probability p. Consequently, each particle arriving at the counter has probability p of being recorded. For $t \geq 0$ let $Y(t)$ be the number of particles recorded in the interval 0 to t. As we have seen, $Y(t)$ is a Poisson counting process with mean rate λp. The R-R condition implies that the distribution of $Y(t)|(X(t) = Y(t))$ is the same as the distribution of $Y(t)|(X(t) > Y(t))$ and is also the same as the distribution of $Y(t)$. This is a property satisfied only if the process $X(t)$ is Poisson.

In a similar manner we can examine the problem in which customers pass by a shop in accordance with a Poisson process at a mean rate λ and each customer has probability p of entering the shop. Analogously we can study the problem where seeds are distributed over an area in accordance with a Poisson process where each seed has only probability p of germination.

The Poisson process has been widely used also in the theory relating to telephone calls. Haight (1967) points out that the number of calls placed in a given time interval can be assumed to be Poisson distributed. In practice, one may assume that there is a number of calls which, although placed, are never made. It is then reasonable to consider the distribution of the number of calls made, as the resulting distribution, with the distribution of the number of calls placed, representing the original distribution.

The Poisson distribution and the Poisson process also have applications in a number of other fields like reliability, inventory control, medicine etc. A comprehensive list of these applications is given in Chapter 7 of Haight (1967).

A careful study of the above examples reveals an important feature of the damage model.

It is evident that when the original distribution is Poisson with parameter λ and the survival distribution is Binomial with parameter p, the resulting distribution is also Poisson with parameter λp. This means that in the resulting distribution the parameters of the original and the survival distribution get confounded; in other words, they cannot be separately estimated. (Analogous results hold for other forms of the original distribution such as the Binomial and the Negative Binomial). In this situation it is natural to ask the following question. Would it be possible to recover some of the information, if it could be ascertained whether an observation is undamaged or whether it has some sort of damage? However, the R-R condition tells us precisely that this ascertainment would not offer any help since in all cases the resulting distributions are identical, i.e.

$$P(Y=r) = P(Y=r|X=Y) = P(Y=r|X > Y)$$

The R-R characterization, on the other hand, shows that this happens only if the original distribution is Poisson. This property of this particular form of the damage model generates an interesting statistical problem, namely that of finding a method of estimating λ and p.

Another point which is worth pointing out here is the importance of making sure that the parameter p of the Binomial survival mechanism is independent of the process X(t), before applying damage model theory. To illustrate this, consider the example given by Parzen (1962) concerning the paralyzable (or type II) nuclear particle counter with a constant locking time L. This is a counter in which a particle arriving at the counter locks the counter for a time L, regardless of whether or not it was registered. Let X(t) be the number of particles arriving at the counter in the time interval (0,t) and let Y(t) be the number of particles registered if and only if no particles arrived during the preceeding time interval of length L.

Consequently, the probability that a particle is registered is $p = e^{-\lambda L}$. Clearly, the time intervals between the arrivals of succesive particles are independent. Therefore, the registering of a particle is independent of the registering of other particles. It might be thought that the damage model theory applies here, and hence, because $\{X(t), t \geq 0\}$ is a Poisson process, $\{Y(t), t \geq 0\}$ is also a Poisson process. However, this is not the case, since the event that a particle is registered is not independent of the process $\{X(t), t \geq 0\}$. (The process Y(t) has been studied by Parzen (1962)).

The above example indicates the need for a modification of the damage model theory in order to cope with such situations.

9.2.3. On the Utility of the Characterizations Concerning the Poisson Distribution.

Characterizations concerning the Poisson distribution - as indeed characterizations in general - may be useful in applied statistics for a number of reasons. Such reasons are the following:-

First of all, a characterization involves a particular property which is unique for the characterized distribution. In our results for example, the R-R condition for the Poisson model with Binomial survival distribution was proved to be valid only if the original distribution is Poisson. The statistical significance of this fact has been demonstrated in the examples presented in the previous section.

A characterization will also be useful because it can guide the choice of assumptions that we have to impose in a given problem. It tells us, for example, whether in a certain situation a particular assumption is redundant or not. The R-R characterization is a typical example. Suppose that a person is not aware of the R-R characterization of the Poisson distribution in the way that it is stated by Theorem 2.1.1 (p26). Instead, suppose that he only knows the similar result proved by Van-der-Vaart and others (see p21). Then he would probably try to see whether the functional equation $G(q + t) = CG(t)$ is satisfied for all values of q in $(0, 1)$, if he wished to show that the distribution corresponding to G is Poisson. However, this is not necessary. As it was shown by Theorem 2.1.1, we need to verify the equation for only one value of q.

One of the most significant contributions of characterizations to applied statistics is in model building. From this aspect, they enable us to know whether one set of conditions is equivalent to another set of conditions. This may reduce a complicated problem to an equivalent but possibly simpler one. The Poisson process for instance has several equivalent formulations. It is a Renewal process with exponential intervals. It can also be viewed as a process X(t) with stationary independent increments such that $X(1)$ has a Poisson distribution. It is then clear that it is better for the investigator to work with the formulation which is more easily verifiable. If, for example, it is easy to show that the intervals are independent and identically distributed with exponential fit, one can in that context say immediately that the process can be viewed as Poisson. In another case it may happen that the investigator will find it easy to see that the process has independent increments such that the conditional distribution of $X(\tau)|X(t)$ for $\tau < t < \infty$ is binomial with parameter p independent of the given value of $X(t)$. One would then ask whether one can conclude just from this information that the process is non-homogeneous Poisson. The result of Moran studied in the thesis provides the required answer and tells us that the process is indeed Poisson of the required type.

The comments presented in the thesis concerning the R-R characterization of the Poisson distribution provides the investigator with a means of constructing equivalent formulations regarding the Poisson distribution. In particular, they suggest that if the conditional distribution of $(Y|X=n)$ is binomial with p independent of the parameter(s) of the distribution of X, then testing the hypothesis that X follows Poisson is equivalent to testing the hypothesis that Y and $X-Y$ are independent. (Tests of independence for discrete distributions have been considered in the literature, e.g. Ahmed (1961)). This is also equivalent to testing the hypothesis that the distribution of the damaged observation is the same as the distribution of the undamaged observation. In addition this is equivalent to testing the hypothesis that the distribution of the undamaged observation is the same as the distribution of the observation when the classification as damaged or undamaged is not known. Hence if the investigator wants to test any of the above hypotheses it is sufficient for him to test any of the equivalent ones; he will choose the one which is simpler for his particular case.

9.2.4 Shanbhag's Extension of the R-R Characterization

The extention of the R-R characterization given by Shanbhag is, in the general case, of mathematical rather than statistical interest. However, the results we obtained as special cases are of some statistical value. It is the aim of this section to discuss these particular cases.

The most important is the one concerning the model where the survival distribution is Negative Hypergeometric and the original distribution is Negative Binomial.

At first glance it might seem that a Negative Hypergeometric form for the survival distribution is not as meaningful as a Binomial form, as far as practical applications are concerned. However, the derivation of the Negative Hypergeometric as a Binomial$\wedge$Beta distribution indicates that this is not so. Quite to the contrary, it happens that certain real life situations are better explained by assuming that the parameter p of the Binomial distribution is not a constant, but instead that it is a variable following a Beta distribution. A good example of a situation like this is given by Griffiths (1973). He took as the Binomial distribution the distribution of the total number of cases if a non-infectious disease arising in households of a given size. He then derived the Beta-Binomial distribution allowing for variation between households by letting p have Beta distribution. In terms of our model, one can consider the distribution of the total number of cases arising in households of size n to be the distribution of $Y|(X=n)$, where the distribution of X corresponds to the distribution of the size of families.

The Negative Binomial was first used as an empirical distribution. This was the result of realising that in many cases a better agreement between observed facts and mathematical theory could be obtained by use of the Negativ Binomial distribution rather than by the use of the Poisson distribution. Experience showed for example that accidents do not always fall in the category of randon Poisson phenomena. Since then the Negative Binomial distribution has very often been the first choice as alternative when it has been felt that a Poisson distribution might be inadequate. While it does not have the same flexibility as certain contagious distributions with more than two assignable parameters, it often gives an adequate representation when the strict randomness required for the Poisson distribution does not hold sufficiently well.

As a direct consequence of the empirical evidence, attempts were made to find mathematical explanations for the Negative Binomial result. This led to the discovery of a number of models giving rise to the Negative Binomial distribution. Among the most important are those related to the Poisson-Gamma mixture, birth-death processes, Poisson-Logarithmic generalisation and the inverse binomial sampling.

The Negative Binomial distribution as Poisson-Gamma mixture was first adopted by Greenwood and Yule (1920) in connection with accident theory. They assumed that the number of accidents for each individual was Poisson, but that the mean value of accidents varied from individual to individual, due to psychophysical factors, according to a Gamma probability law. (The parameter λ is now called accident proneness). In terms of the damage model, we can consider the distribution of the number of accidents as the original Negative Binomial distribution. We can then assume that each accident is reported with probability p which varies from accident to accident, depending on the nature of the accident - social or financial pressures may encourage the individual not to report all the accidents he incurs . If we allow p to be Beta distributed, then our Negative Binomial model with Hypergeometric survival is applicable. The effect of the R-R condition here is identical to the effect it had in the accident example discussed in Section 9.2.2.

As pointed out by Bartko (1961) the use of the Negative Binomial in connection with birth-death processes was considered by Yule (1924), Furry (1937) and Kendall (1949). Yule is concerned with the mathematical theory of evolution, Furry with cosmic-ray showers and Kendall with the process in general.

Quenouille (1949) proved that the Negative Binomial distribution arises as the distribution of the random sum of n independent variables each having the same logarithmic series distribution, where n is a Poisson random variable.

An example of the above model is the problem examined by Ashford (1972) concerning patient contacts with the doctor. He assumed that episodes of illness occur as events in a Poisson process with parameter λ. Each episode may be assumed to give rise to a variable number of contacts of a given type with a doctor (e.g. home visits) which can be considered to have a form of the logarithmic series dsitribution. On the assumption that the number of contacts arising out of different episodes are independent, and also independent of the number of episodes, then the Negative Binomial can be obtained as the distribution of the number of contacts. In our set-up the distribution of the total number of contacts with the GP can be considered as the distribution of X. Assume that each contacting patient is referred to a consultant with probability p. If p is assumed to vary from patient to patient (depending on the seriousness of his illness) according to a Beta distribution, then the conditions for our model are satisfied. The R-R condition and the characterization in this case may help in building equivalent formulations for the model, and also in examining the distribution of the number Y of subsequent contacts with the consultant.

Inverse Binomial sampling is of primary importance in the consideration of the Negative Binomial distribution. It occurs in the following general type of situation. Consider a population consisting of individuals belonging to two different classes. Then the number of individuals of the second class drawn before the k-th, say, individual of the first class, follows a Negative Binomial distribution. (Drawings are made without replacement).

In our model suppose that in a survey concerning the Government's pay policy X represents the number of Labour voters questioned before k Conservative voters have been interviewed. Suppose that each of these Labour voters favours the Government's pay policy with probability p. Suppose that p varies from Labour voter to Labour voter depending on his Trade Union affiliation, and that p follows a Beta distribution. Then Y will represent the number of Labour voters in favour of the Government's pay policy who were questioned before k Conservative voters had been interviewed. Our results show that Y will also follow Negative Binomial distribution. They also suggest that $Y|X=Y$, $Y|X>Y$ and Y are identically distributed. They also prove that Y and X-Y are independent.

The examples presented in this section illustrate how some of our results based on Shanbhag's extension of the R-R characterization can be used in practice. It is important to point out here that the R-R condition for the Negative Binomial model with Negative Hypergeometric damage has the same implications as for the Poisson model with binomial damage.

9.2.5 The R-R Condition in Relation to Truncated Distributions

Truncated distributions are not used as often in practice as untruncated ones. They arise when some of the non-negative integers are omitted from the range of possible values. This happens either because these values are theoretically meaningless under the model,or else because they are unobservable in practice.

Two distinct classes of left-truncated distributions were dealt with in Chapters 3 and 4. These corresponded to truncation of the first c frequencies (truncation at c-1), and truncation of the zero frequency (truncation at zero). The first class is more general and includes the second class as a special case. We dealt with this first class mainly because of its mathematical interest. However,the zero truncated class has an important statistical interpretation. This happens because cases where the zero class is either missing or non-observable arise very frequently in practice. In the sequel,we focus attention on practical situations where the zero truncated Poisson and the zero-truncated Negative Binomial distributions are applicable. (The truncated Poisson and the truncated Negative Binomial were the forms we dealt with in Chapter 3). We also give a practical instance of a (c-1)-truncated Poisson distribution. These examples are examined in relation to the results studied in the thesis.

An example of a zero-truncated Poisson distribution is given by Haight (1967)(Section 3.1). Assuming that the number of courses taken by university students is Poisson distributed we have (from the definition of a student) that at least one course is taken by each student. Hence the zero category would not be germane to the discussion and the random variable would be defined over the truncated domain $n = 1,2,3....$. This is an example where the zero class is theoretically meaningless for the model. In connection with our study we may consider as X the number of courses taken by university students and as Y the number of these courses which are of theoretical nature (p is the probability that a course chosen will be theoretical). The R-R condition here implies that the distribution of the number of courses taken by students who have chosen only theoretical courses is the same as the distribution of the number of theoretical courses taken by all students with the zero class excluded. On the other hand, if these two distributions are the same, then the truncated Poisson is the only appropriate distribution for X.

A classical example of zero-truncated distribution where the zero class is practically unobservable is given by Rao (1963). It arises in connection with the birth of children with a specific genetic defect (for example, albinism in families carrying the appropriate gene. If the genetic character of the parents can be observed only by means of the birth of such children, all families with normal children, or with no children, will escape notice even though they may well belong to the population being sampled. In this case, the domain of definition will be n = 1, 2,.... defective children. In terms of the damage model,X may be considered to represent the number of defective children and Y the number of male defective children. (p will be the probability that a defective child is a boy).

Another model using the zero truncated Poisson distribution has been studied by Placket (1953). He was interested in the distribution of the number of workers X in a factory having n accidents in a given period of time. He concentrated on the values 1,2... of n. He reasoned that, while it was a simple matter to count over that period of time the number of workers sustaining one, two, or more accidents, the number of persons incurring no accidents could not be enumerated owing to fluctuations in the size of the factory population during that period. In the damage model set up Y can be considered to represent the number of severe accidents in the same period of time (p is the probability that an accident will be severe). Here again the R-R condition may be used to provide information about the distribution of Y truncated at zero, $Y|(X=Y)$ and $Y|(X>Y)$.

A model for a (c-1) truncated distribution would hold in the study of the number of students attending a particular course. Universities are known to be under pressure to abandon courses attended only by a few students. A (c-1) truncated distribution seems therefore to be appropriate. Our model would be useful if we wanted to study the number of students who are likely to get an 'A' mark in that course. (p here is the probability that a student taking the course will get 'A'). The characterization concerning the (c-1) truncated Poisson distribution provides useful information in such a study.

An example when the zero truncated Negative Binomial distribution is applicable is given by Bartko (1961). He suggests this is connection with his use of the Negative Binomial distribution to represent the number of representatives of different species of butterflies obtained in a collection. He points out that in such a study only the frequencies of numbers greater than zero will be observable. This happens because the collection by itself gives no indication of the number of species which have not been collected. (In cases like the above an alternative choice has been the logarithmic distribution). Our characterization of thr truncated Negative Binomial distribution may be useful for discriminating between these two models. The characterization tells us that, under the assumption that $Y|X$ is Negative Hypergeometric, the hypothesis that X is zero truncated Negative Binomial has the following equivalent formulations: X-Y and Y truncated at zero are independent; Y truncated at zero and $Y|X=Y$ are identically distributed; $Y|X=Y$ and $Y|X>Y$ are identically distributed. If the nature of a particular problem enables the investigator to prove or disprove one of the equivalent formulations, then, in the light of the characterization, he can conclude whether the distribution of X is zero truncated Negative Binomial or not. (In the latter case the research worker might then turn to the logarithmic distribution).

The model with truncated survival distribution is mainly of mathematical interest. However, it could also arise in practice in the following way. For example, in the Griffiths model examined in section 9.2.4, suppose that it is impossible to observe the zero class for the number of cases of the non-infectious disease arising in a household of a given size. (This will happen if we have data of those households in which there is at least one case, but cannot observe the number of households with no cases at all). In such a situation, as Griffiths points out, the zero truncated Binomial-Beta will apply.

It is evident from the examples given in this section that the R-R condition for the cases where either the original or the survival distribution is truncated gives the same sort of information as for the untruncated case mentioned in 9.2.2. The only difference is that here the information refers to the resulting distribution truncated at the same point as either the original or the survival distribution.

9.2.6 Finite Distributions and the Damage Model

In this section we consider the applicability of the results obtained in Chapter 5 concerning finite distributions. A damage model interpretation of these results will be described.

It has been pointed out in the thesis (Chapter 5) that the R-R condition $(P(Y=r) = P(Y=r|X=Y))$ does not characterise finite distributions. Our results demonstrated that in such cases the required condition was

$$P(Y=r|X=Y) = P(Y=r|X=Y+1) = \dots = P(Y=r|X=Y+\ell) \qquad (9.2.1)$$

where ℓ is fixed and $1 \leq \ell \leq N-m$.

Thus the initial motivation for the study of this problem was purely mathematical. However, its importance from the statistical viewpoint should not be neglected.

Relationship (9.2.1) was proved to be a characterising condition for the first ℓ frequencies of finite discrete distributions ('modified distributions') This can be useful in some statistical problems where one is interested in knowing what happens within a reasonable distance from the origin of a distribution, e.g. when the bulk of the probability mass is concentrated towar(zero.

Consider, for example, a population consisting of N individuals inoculated against a disease. The number $X=n$ of individuals attacked by the disease, even though they have been inoculated, will have probabilities which will tend to be negligible as n approaches N. Then one can ignore the behaviour of the tail of the distribution beyond a certain point $\ell+m$ and concentrate on the firs $\ell+m$ probabilities. Corollary 5.2.1 can then be utilised. Suppose that X represents the number of individuals who will be attacked even though they have been inoculated. Assume that m individuals out of N have been attacked in the past. Further denote by Y the number of individuals who will be re-attacked (clearly $Y \leq X$). It is then reasonable to assume that for a given sample of n attacked out of N individuals, the distribution of the number $Y=r$ of the individuals who will be re-attacked will be hyper-geometric. The conditions for corollary 5.2.1 will then be satisfied. Accordingly, condition (9.2.1) for $\ell \leq N-m$ will characterize the first $\ell+m$ probabilities P_n of the distribution of X; these will be proportional to Binomial probabilities. Note that P_n will tend to the Binomial probability as $\ell \to N-m$. This implies that the finite original distribution is such that truncating the frequencies beyond $\ell+m$ gives rise to the Binomial distribution truncated beyond $\ell+m$. In other words relation (9.2.1) provides only partial information about the original distribution. However, if we know beforehand that the distribution of X has its frequencies truncated beyond $\ell+m$ ($P_n=0, n>\ell+m$ then (9.2.1) will characterize all the frequencies of the distribution of X. In relation to our example X will be characterized as a random variable having the Binomial distribution truncated beyond $\ell+m$.

This was an example concerning the application of the results of Chapter 5.

9.2.7 Bivariate and Multivariate Distributions

Our purpose in this section is to demonstrate how the results on multivariate distributions can be utilised in practice.

We will investigate cases giving rise to the multivariate distributions which have been dealt with in the thesis. In particular, we will discuss models leading to the positive and Negative Multinomial and to the Multiple Poisson distributions. Statistical inference will then be made in the light of the information provided by the relevant results studied in the thesis.

Multivariate distributions are very often applicable in real-life situations in which we deal with populations consisting of several types of individuals (e.g. a population of members classified according to height, weight, intelligence, financial status etc.)

The extension of our results concerning modified finite distributions to the multivariate case led to the characterization of the Multinomial distribution.

This is one of the most important multivariate distributions from the point of view of practical applications. It is applied in circumstances similar to those in which a binomial distribution might be used when there are multiple categories of events instead of a simple dichotomy. Consider, for instance, a population exhibiting s+1 qualities $Q_1, \dots Q_s$ and neither Q_1, nor $Q_2, \dots$ nor Q_s which is denoted by Q_o. Then, if the corresponding proportions are $p_1, \dots p_s, p_o$ the joint distribution of the numbers r_i of individuals exhibiting the i^{th} quality $i = 1,2,\dots, s$ in a sample of n is the multinomial distribution. An illustrative example is that of tossing s dice and asking for the probability of getting the i^{th} face r_i times, $i = 1,\dots 6$. Clearly, this is multinomial with $p_i = \frac{1}{6}$ i.e. it is given by $s!\ 6^{-s}/r_1!r_2! \dots r_s!$

Conditions for a Multinomial distribution are also satisfied whenever data obtained by random sampling are grouped in a finite number of mutually exclusive groups. This suggests that one can apply the Multinomial distribution to many problems requiring estimation of a population distribution say $F_o(x)$. In fact, since $F_o(x)$ is specified, we may calculate the probability p_{oi} of an observation falling in the i^{th} class. Then, if n_i denotes the observed frequencies in the i^{th} class ($\Sigma n_i = n$), the n_i's are Multinomially distributed and the likelihood function will be given by $n! \prod_i p_{oi}^{n_i}/n_i!$

Consider the characterization of the Multinomial distribution as given by the corollary 6.5.3. This could be of statistical interest in the following way.

Assume that we have a population of size N consisting of individuals of various intelligence standards (slow, slow but able to concentrate, fairly intelligent,...). Let X_i $i = 1,2,\dots,3$ be the number of individuals of the i^{th} standard $i = 1,2,\dots,s$ and Y_i be the number of the i^{th} standard individuals who are unemployed.

Then, for given $\underset{\sim}{X} = \underset{\sim}{n}$ out of N, $\underset{\sim}{Y}$ can be reasonably assumed to be distributed in the Multivariate Hypergeometric form.

This set up corresponds to the theoretical model studied by Corollary 6.5.3. Hence, this model can be used for deducing a Multinomial form for $P_{\underset{\sim}{n}}$. Similarly, if $P_{\underset{\sim}{n}}$ is assumed to be Multinomial we can have information about the distributions of $\underset{\sim}{Y} \mid (\underset{\sim}{X} = \underset{\sim}{n})$ and $\underset{\sim}{Y}$.

The Negative Multinomial distribution has also been examined in chapter 4 in relation to the results concerning the multivariate extension of the R-R condition. It arises in situations where we go on making trials until exactly n occurences of the s^{th} outcome have been noted and we require the joint probabilities of n_i occurences of the i^{th} outcome $(i = i,\ldots,s - 1)$ noted before the n^{th} occurence of the s^{th} outcome. Clearly it is a generalization of the Negative Binomial and just as the latter can be deduced from a number of different models, so can the Negative Multinomial.

Thus, Bates and Neyman (1952) arrive at this distribution in the context of accidents, as a result of mixing s independent Poisson random variables with parameters proportional to a gamma variable.

In particular, they assumed that the various kinds of accidents incurred by an individual were independently distributed as Poisson random variables with parameters λa_i, $i = 1,2,\ldots,s$. They thus obtained the Negative Multinomial by assuming that λ was a Gamma random variable associated with the individual's proneness to accidents.

Consider now the importance of the characterization of the distribution as given by corollary 4.5.4. In a given problem where conditions are satisfied for a multivariate Inverse Hypergeometric distribution to be the distribution of the conditional r.v. $\underset{\sim}{Y} \mid (\underset{\sim}{X} = \underset{\sim}{n})$ we are able to deduce a Negative Multinomial distribution for $\underset{\sim}{X}$. Such a form for the distribution of $\underset{\sim}{Y}|(\underset{\sim}{X}=\underset{\sim}{n})$ may not seem feasible in practice. However, it is a distribution used in connection with pollen analysis. Janardan (1973b) for instance, assumed that counts of various kinds of pollen grains found at a given depth in sediment follow independent Binomial distributions with constant proportion p. He then allowed p to vary from depth to depth, according to a Beta distribution. Averaging over all depths in this manner he obtained the multivariate Inverse Hypergeometric as the joint distribution for counts of various kinds of pollen grains. Therefore, in a problem of pollen analysis with multivariate Inverse Hypergeometric survival mechanism the results of corollary 4.5.4 may indicate that the counts $\underset{\sim}{X}$ of the different pollen species have a Negative Multinomial distribution. Similarly, they might indicate that $\underset{\sim}{X}$ has a distribution which is definitely not the Negative Multinomial.

A special class of multivariate distributions which has been examined in the thesis, is the class of multiple distributions. These are multivariate distributions with independent components.

One has to admit that the assumption of independent marginal distribution is unlikely to be satisfied exactly. There are, however, situations where for reasons of mathematical simplicity one considers such distributions as first approximation. This approach has the advantage of making the problem mathematically tractable.

In an accident study, for example, we could divide the time interval from $t = 0$ to $t = T$ into s non-overlapping sub-intervals. Secondly, we could assume that accidents are due to pure chance. In other words we could assume that what happens in one time period has no effect upon what happens in any other period. Furthermore, let us assume that an individual can have at most one accident in the short interval of time Δt with probability depending on time alone. Then the joint distribution of the numbers of accidents for the s sub-intervals is the multiple Poisson.

The examples and discussion of this section show the degree of the statistical importance of the results produced in the thesis concerning multivariate distributions.

9.2.8 The Damage Model with a Mixed Poisson as the Original Distribution

In this section we examine practical situations leading to the mixed Poisson distribution. We also discuss applications of damage model theory in such situations. Finally we indicate how the characterizations related to this model and obtained in Chapter 8 can be used for making statistical inference.

Situations arise very often in practice where a Poisson distribution with a constant parameter does not fully describe the phenomenon under investigation. This happens frequently for example in entomological and bacteriological studies, as well as in absenteeism studies and accident studies. This is the result of inhomogeneity in the underlying population. A possibility in these circumstances is that the population consists of two or more homogeneous populations. Therefore the underlying distribution will be given by

$$P(X=n) = \int_0^{\infty} e^{-\lambda} \frac{\lambda^n}{n!} \, dF(\lambda)$$

where $F(\lambda)$ is an arbitrary distribution function.

Applications of mixed Poisson distributions go as far back as 1920 when Greenwood and Yule used the Poisson-Gamma distribution to describe accident data (see section 9.2.4). Purchasing behaviour is another phenomenon which has been studied using a Poisson-Gamma model. Chalfield and Goodhart (1970) assumed that for each individual household the number of packets bought in successive periods of equal length, can be regarded as a random variable, having a Poisson distribution with mean λ. They then allowed λ to vary from household to household having a Gamma distribution in the whole population.

In many applications of mixed Poisson distribution the phenomena in question correspond to a Poisson process. This is the case for example in the model of the Poisson-Truncated Gamma used by Kemp (1968b) in relation to collective risk theory. She argued that for insurance applications it is more realistic to assume limited risk having the form of a tail-truncated Gamma distribution.

The Poisson-Beta is another mixture of the Poisson distribution which we have considered. This is a special case of Gurland's distribution (1958). He assumed that the distribution of observed larvae survivors in a plot of a field is Poisson (λp) with p fluctuating from plot to plot according to a Beta distribution.

Our results in Chapter 7 provide the form of the resulting distribution after Binomial damage in all these cases. Consequently, in practice, one can readily obtain the resulting distribution for different forms of the original distribution.

The form of the resulting distribution when no damage has occurred has been derived for various cases (i.e. the distribution of $Y|X=Y$). A second interpretation (other than the damage interpretation) can be given to these forms of distribution. For each of the cases which have been examined, $P(Y=r|X=Y) = P_r\, s(r,r)/\sum_r P_r\, s(r,r)$. We can then consider the probability $s(r,r)$ (e.g. the probability $P(Y=r|X=r)$ to be the weight in a weighted model. The weight represents the probability that an observation is actually included in the sample. Weighted models have been considered by Rao (1963), Kemp (1973) and Patil and Rao (1976). In this case the distribution of $Y|(X=Y)$ will be a weighted form of the original distribution.

The results of Chapter 8 have the following statistical interpretation. Suppose that in a practical situation we know that a Binomial survival mechanism is operating, and that the observed distribution is Poisson. Then we are able to infer that the original distribution is also mixed Poisson. What is more important is that the original mixed Poisson distribution has the same mixing distribution as the observed one.

The results of Chapters 7 and 8 apply also to other forms of mixed Poisson distributions which we have not mentioned specifically. Such mixtures include the Poisson$\wedge$Rectangular studied by Bhattacharya and Holla (1965). They include also the Poisson$\wedge$Poisson (Neyman type A) which was studied by Neyman (1939) in connection with the distribution of larvae in a field.

Thus it is evident that the results concerning mixed Poisson distribution are of immense use in applied statistics.

9.2.9 The Damage Model with a Mixed Binomial as the Survival Distribution

In this section we deal with the statistical significance of the results of Chapters 7 and 8 of the thesis, which are based on the assumption that the survival distribution is of a mixed Binomial form. We will give some examples where such a situation arises in practice, and we will try to find their relation to our results.

The Mixed Binomial distribution arises when the parameter p of the Binomial distribution is not a constant but varies according to some distributioı This happens in practice when sampling is taking place over an extended area or an extended period of time. It has been observed that data derived in this way do not conform to the simple Binomial type.

The Binomial-Beta model has already been mentioned earlier as a model for the Negative Hypergeometric distribution. Another interesting model for the Binomial-Beta distribution has been given by Kemp and Kemp (1956). The authors derive it as the distribution followed by point quadrat percentage cover data. Their theoretical consideration was that for a given location of a frame of n pins the number of contacts of the species under study is Binomial. The expected proportion of contacts however was assumed to vary from location to location, and was assumed to have a Beta distribution.

The same model has been adopted by Chatfield and Goodhart (1970). They assumed that the number of weeks out of n in which a consumer makes at least one purchase is Binomially distributed with parameters n and p where the value of p varies from consumer to consumer, and has a Beta distribution over the whole population.

Other forms of mixed Binomial distribution such as the Binomial-Right truncated Gamma do not seem to arise in practice. They are mostly of mathematical interest.

The forms of the distribution of Y and $Y|(X=Y)$ were derived for many different forms of mixed Binomial, under the hypothesis of a Poisson original distribution.

As far as the characterizations based on mixed Binomial survival distribution are concerned apart from their mathematical interest, they contribute in a positive way in some applied problems.

Consider, for example, the number X of cars passing, in a given period of time, through a junction with traffic lights. Let the number of cars out of n which pass while the lights are red be Binomially distributed with parameter p. Assume further than p is not a constant, but that it is a random variable associated with the driver's tendencies to commit an offence. Then the number of cars out of n passing when the red light is on, will have the Beta-Binomial distribution.

If it can be ascertained that the distribution of cars passing against a red light is Poisson $(\lambda p)\wedge$Beta then we can deduce a Poisson distribution for X.

In the same example, assuming a Poisson distribution for X, then we can deduce a Poisson^Beta for the number of drivers who cross against the lights.

It seems that the Binomial^Beta is the form of mixed Binomial survival distribution most likely to occur in practice. The mixed Poisson model with mixed Binomial survival distribution is a generalization of the previous two models. It corresponds to the case where the parameters of the original and the survival distributions are both random variables.

9.2.10 Concluding Remarks

This section has examined practical situations where results studied in this thesis can be useful. Although the study was mainly mathematically motivated, it is clear from this discussion that the results can be applied to many practical problems.

On the other hand, this section has disclosed that some of the results presented suggested other interesting problems of a statistical nature which need attention. These include the problem of finding a method for estimating the parameters λ and p in the damage model when the original distribution is Poisson and the survival distribution is Binomial. They also include the problem of studying the necessary modifications in the damage model theory to deal with cases where p depends on the process X(t).

Finally this section has demonstrated the extent to which our characterizations may be used to simplify statistical problems.

9.3 Scope for Further Research

This section suggests directions in which the work of this thesis could be extended.

1. Theorem 3.1.1 in Chapter 3 provides the basis for characterizations of discrete distributions. It would be interesting to extend this result to the continuous case.

2. The R-R characterization and all its variants and extensions have been examined under the assumption that the conditional distribution of $Y|X$ is independent of the parameter λ of the distribution of X. What would be the effect of having the distribution of $Y|X$ dependent on λ?

3. Characterizations of Bivariate and Multivariate distributions with independent components were derived in Chapters 4 and 6. It is possible to obtain similar characterizations for Multivariate distributions whose components are not indepedent although of course the methods required are more complicated. However, in most of these cases characterizations based on the conditions introduced in Chapters 4 and 6 will result in characterizing distributions of non-standard form. Perhaps some changes in the conditions used in Theorems 4.1.1, 4.5.1, 6.1.1 and 6.4.1 would help to obtain meaningful characterizations for this kind of distributions

4. The derivation of simpler proofs for Theorems 4.5.1 and 6.1.1 may be of some interest.

5. A "generating model" as opposed to the damage model would arise if one assumes the distribution of $Y|X$ to be Pascal of the form

$$P(Y=r|X=n) = \binom{r-1}{n-1} p^n q^{r-n} \quad r=n,\ n+1,\dots;\ n=1,2\dots$$

instead of Binomial; in this case of course the resulting r.v. Y will be larger than X, i.e. we will have $X=Y-Z$. Such a situation can arise for example if X represents the number of motorway accidents resulting in injuries in a given locality and for a given period, and Y stands for the number of injuries in these X accidents.

This model could be studied along lines similar to those for the damage model. (It is evident, for example, that in the case where X is also Pascal, Y comes out to be Pascal with its parameter confounded with the parameter of the distribution of X.) It would be interesting therefore to find out how many of the results obtained for the damage model could be transferred to this "generating model". It seems however that the R-R condition as it stands is not very useful for obtaining characterizations based on this model. Another kind of condition may be more helpful.

6. Should such results be obtainable then one could go on to combine the two situations and to study a more complicated model in which $Y|X$ follows a birth-death process.

7. To characterize the compound Poisson distribution by Theorem 8.2.2, we had to impose the condition that the distribution of X is uniquely determined by its factorial moments; this condition is, of course, very stringent. Is it possible to relax this condition ?

8. The results of Chapter 7 and 8 can be extended easily to the Bivariate and hence to the Multivariate case. Similar extensions would be possible for the truncated case.

9. All the results in Chapters 7 and 8 were derived with either the distribution of X having a compound Poisson form or the distribution of $Y|X$ having a compound Binomial form.

   Other forms of compound distributions could be studied and parallel results could possibly be obtained. (One obvious choice would be the compound Binomial for X and the compound Hypergeometric for $Y|X$).