Mixtures and Random Sums

C. CHATFIELD and C. M. THEOBALD, *University of Bath*

A brief review is given of the terminology used to describe two types of probability distribution, which are often described as "compound" and "generalized" distributions. In view of the confusion in terminology, it is suggested that these terms be discarded and that the more descriptive terms suggested by Feller, namely "mixtures" and "random sums", should be generally adopted. Some remarks on "contagious" models and related topics are made. Attention is drawn to the fact that a mixture may be regarded as a marginal distribution of a bivariate distribution and that a random sum may be regarded as a special type of mixture. General formulae for the mean and variance of a mixture and of a random sum are also given.

1. Mixtures

Following Feller (1966, page 52) we can generate a class of probability distributions in the following way. Let F_X be a distribution function depending on a parameter θ, and let F be another distribution function. Then

$$F_Y(y) = \int_{-\infty}^{\infty} F_X(y|\theta)\, dF(\theta) \qquad -(1)$$

is also a distribution function. Feller describes distributions of this type as mixtures. Distributions arise in the above way in a wide variety of situations, and have numerous practical applications.

Some other authors (e.g. Kemp, 1970; Gurland, 1957) describe distributions of this type as compound distributions, while Patil and Joshi (1968, page 5) and Johnson and Kotz (1969, pages 27 and 183) use both terms. In fact Feller (1943) also originally used the term compound distribution. However, the term mixture is much more descriptive and is coming more into fashion (e.g. Haight, 1967). In view of the fact that the term "compound distribution" is used by some authors in a different way (see section 2), it would be advantageous for the term "mixture" to be generally adopted.

For some purposes it is useful to think of a mixture as being a marginal distribution of a bivariate distribution. Suppose we have two random variables Y and T. Suppose further that we are given the marginal distribution of T, and also that we are given the conditional distribution of Y given T = t for every value of t. Then we can evaluate the marginal distribution of Y, which is a mixture. For a particular conditional distribution of Y given a particular value t of the random variable T, we may regard t as a parameter of this conditional distribution.

It is clear from the above approach that every probability distribution can be represented as a mixture (Feller, 1966, page 72), but in practice the term is usually reserved for distribution arising in the case where all the conditional distributions are of the same type. For example, a mixed Poisson distribution arises when the conditional distribution of Y given T = t is Poisson with mean t for all positive values of t.

Perhaps the best-known example of a mixture is that obtained by "mixing a Poisson distribution with a gamma distribution" to give a negative binomial distribution. Suppose that the conditional distribution of Y given T = t is a Poisson distribution with mean t for all positive t. In other words we have a family of conditional distributions which are all Poisson. If the marginal distribution of the random variable T is gamma with probability density function

$$f(t) = e^{-t/a}\, t^{k-1}/\{a^k\, \Gamma(k)\} \qquad (0<t<\infty)$$

where $a > 0$, $k > 0$, then the resulting marginal distribution of Y is given by

$$P(Y = r) = \int_0^\infty e^{-t} \frac{t^r}{r!} f(t)\, dt$$

$$= (1+a)^{-k} \frac{\Gamma(k+r)}{\Gamma(k)\, r!} \left(\frac{a}{1+a}\right)^r$$

$$(r = 0, 1, ---)$$

which is a negative binomial distribution. If k is an integer and $\theta = a/(1+a)$, then $P(Y = r)$ may be rewritten as

$$P(Y = r) = \binom{k+r-1}{r} \theta^r (1-\theta)^k$$

which is the familiar form arising from inverse binomial sampling (e.g. Feller, 1968, p. 165).

A distribution of this type arises naturally from a "spurious contagion" model discussed for example by Feller (1943) in a paper which is still well worth reading. In a "true contagion" model, the occurrence of an event increases (or decreases) the probability of the event occurring again, whereas apparent or spurious contagion arises because of inhomogeneity in the population. For example, suppose that a population consists of individuals for each of whom the number of "events" recorded in a given time-period has a Poisson distribution. Further suppose that the Poisson mean event rate is constant for an individual but varies from individual to individual and has a gamma distribution in the whole population. Then in the whole population of individuals the distribution of the number of events recorded per individual in a given time-period will follow the negative binomial distribution.

The moment generating function of a mixture may be written

$$m_Y(s) = E(e^{sY})$$
$$= \underset{T}{E}\left[E(e^{sY}|T = t)\right]$$
$$= \int m_c(s|t)dF(t)$$

where m_Y, m_c are the moment generating functions of the marginal distribution of Y and the conditional distribution of Y given T = t. In particular it is often useful to express the mean and variance of a mixture in the following way

$$E(Y) = \underset{T}{E}[E(Y|T = t)]$$

which by convention is written $\underset{T}{E}[E(Y|T)]$.

$$V(Y) = \underset{T}{E}[V(Y|T)] + \underset{T}{V}[E(Y|T)]$$

= mean conditional variance+variance of the conditional mean. —(2)

These are "well-known" results if a mixture is regarded as a marginal distribution of a bivariate distribution (e.g. Rao, 1965, page 79), and can also be derived directly from the above expression for the moment generating function of a mixture. The formulae are also tucked away in Feller (1966, page 164) as an exercise, but deserve wider appreciation.

The formulae for the mean and variance of a mixture are particularly simple when there is linear regression of Y on T so that

$$E(Y|T = t) = a+bt \qquad \text{for all } t.$$

Then we have

$$E(Y) = a+b\mu_T \qquad \text{—(3)}$$

and $$\underset{T}{V}[E(Y|T)] = b^2\sigma_T^2 \qquad \text{—(4)}$$

where μ_T, σ_T^2 denote the mean and variance of the random variable T.

This class of mixtures includes random sums (see section 2), mixed Poisson and mixed binomial distributions, and the latter will be used as an example. Suppose the random variable T is defined on (0, 1). Further, suppose that the conditional distribution of Y given T = t is binomial with index n and parameter t for all values of t in (0, 1). Then we say that the marginal distribution of Y is a mixed binomial distribution.

Since

$$E(Y|T = t) = nt$$

is a linear function of t with a = 0, b = n and

$$V(Y|T = t) = nt\,(1-t)$$

then if μ_T, σ_T^2, denote the mean and variance of T, we find

$$\begin{aligned} E(Y) &= n\,\mu_T \\ V(Y) &= \int_0^1 nt\,(1-t)\,dF\,(t) + n^2\sigma_T^2 \\ &= n\mu_T - n\,(\sigma_T^2 + \mu_T^2) + n^2\,\sigma_T^2 \\ &= n\mu_T\,(1-\mu_T) + n\,(n-1)\,\sigma_T^2. \end{aligned}$$

Mixed binomial distributions have applications in several areas including consumer purchasing behaviour (Chatfield and Goodhardt, 1970) and readership and televiewing behaviour.

2. Random Sums

Let X_1, X_2, ---- be independent, identically distributed random variables and let N be a random variable independent of the X_j which is defined on the non-negative integers. Then we can consider a random variable of the following type:

$$S = \begin{cases} X_1 + X_2 + \text{-----} + X_N. & N > 0 \\ 0. & N = 0 \end{cases}$$

Feller (1966, page 53) describes random variables which arise in this way as "random sums".

If the X_j and S are also defined on the non-negative integers it is easy to show that

$$g_S\,(t) = g_N\,(g_X\,(t))$$

where g_S, g_N, g_X are the probability generating functions of S, N, X respectively. This relationship between probability generating functions is often used as a definition of a class of distributions called generalized distributions (e.g. Patil and Joshi, 1968, page 5; Johnson and Kotz, 1969, page 202; Kemp, 1970) and in fact this description is the one originally used by Feller (1943). However, some confusion has arisen from the fact that Feller changed his terminology and in his volume 1 (1968 and also in earlier editions) described distributions of this type as compound distributions. Some other authors (e.g. Bailey, 1964) also use "compound" in this way. However, in his volume 2 (1966) Feller uses "random sum". This confusion in terminology has been briefly mentioned by Haight (1967, page 36). Clearly the reader should ascertain exactly what is meant by the term "compound distribution" when it is used by different authors. To add to the confusion, the expression "generalized distribution" has also been used in quite different situations (e.g. Johnson and Kotz, 1969, page 109). Thus there is something to be said for avoiding the terms "compound" and "generalized" altogether. Boswell and Patil (1970) have recently adopted Feller's term "random sum" which is much more descriptive than either "compound" or "generalized" and we hope that this term will be adopted more widely. (Recently Douglas (1970) has suggested the term "randomly-stopped sum" which is perhaps even more descriptive.)

Part of the confusion probably arises from the fact that a "random sum" can be considered to be a special type of mixture by regarding the number of terms, n, as a parameter of the conditional distribution of S given $N = n$. This result is given by Feller (1966, page 53) and Moran (1968, page 69) but is not always clear in other sources. Suppose each X_j has distribution function F, then, given $N = n$, the distribution function of

$$S = X_1 + \cdots + X_n$$

is the n-fold convolution of F, which we will denote by F^{n*} (Feller, 1966, page 157). Then, from equation (1), we have that the distribution function of the unconditional random variable S is given by

$$F_s(s) = \sum_{n=0}^{\infty} F^{n*}(s) \, P(N = n)$$

$$\text{where } F^{0*}(s) = \begin{cases} 0 & s < 0. \\ 1 & s \geq 0. \end{cases}$$

A well-known example of a random sum is that arising when N is a Poisson variable and each X_j is a logarithmic series variable with

$$P(X_j = r) = \frac{-q^r}{r \log_e(1-q)} \quad (r = 1, 2, \cdots)$$

and $0 < q < 1$. Then it can be shown that the random sum S has a negative binomial distribution.

This last result spotlights another potentially confusing problem. Some distributions can arise in several different ways from quite different models. For example the negative binomial distribution can arise from a "gamma mixture of Poisson distributions" or a "Poisson sum of logarithmic series distributions". It can also arise from a "true contagion" model called the Polya-Eggenberger urn scheme. This has led to severe difficulty in interpreting data because, on the basis of a distribution in a single time period, it is impossible to distinguish between several models giving rise to the same distribution. For many years there was argument about "accident proneness" as to whether the good fit of the negative binomial distribution to empirical distributions of accidents was an indication of true or apparent contagion. In recent years it has become possible, though not easy, to distinguish between different types of model by studying the multivariate distribution of events in different time periods (Bates and Neyman, 1952; Kemp, 1970).

In passing it is worth pointing out that it may be misleading to describe a distribution, such as the negative binomial distribution, as a "contagious" distribution. The adjective contagious should really be applied to the term "model" rather than "distribution" as there are several models other than contagious ones which give rise to the negative binomial distribution.

Finally, we note that although a random sum can always be represented as a mixture, the reverse statement clearly does not generally hold. However, Gurland (1957) has given a useful result which suggests when a mixture may also be represented as a random sum.

General formulae for the mean and variance of a random sum may easily be derived from equations (2), (3) and (4). Suppose that each X_j has mean μ_X and variance σ_X^2. Then the conditional distribution of S given $N = n$ has mean $n\mu_X$ and variance $n\sigma_X^2$. The mean value of this conditional distribution is a linear function of the "parameter" n for all values of n with $a = 0$, $b = \mu_X$. If the random variable N has mean μ_N, variance σ_N^2, then we have from equations (2), (3) and (4) that

$$E(S) = \mu_X \mu_N$$

$$V(S) = \underset{N}{E}[n\sigma_X^2] + \mu_X^2 \sigma_N^2$$

$$= \mu_N \sigma_X^2 + \mu_X^2 \sigma_N^2.$$

These useful results are also given by Feller (1966, page 164) as an exercise.

Acknowledgements

It is a pleasure to acknowledge some useful discussions with Mr. G. J. Goodhardt, Aske Research.

REFERENCES

BAILEY, N. T. J. (1964). *The Elements of Stochastic Processes with Applications to the Natural Sciences.* New York: Wiley.

BATES, G. E. and NEYMAN, J. (1952). "Contributions to the theory of accident proneness. Parts I and II." *University of California Publications in Statistics,* **1**, 215–75.

BOSWELL, M. T. and PATIL, G. P. (1970). "Chance mechanisms generating the negative binomial distribution." In: *Random Counts in Scientific Work.* Vol. 1, 3–22. Ed. G. P. Patil. Penn. State University Press.

CHATFIELD, C. and GOODHARDT, G. J. (1970). "The beta-binomial model for consumer purchasing behaviour." *Appl. Statist.*, **19**, 240–50.

DOUGLAS, J. B. (1970). "Statistical models in discrete distributions." In: *Random Counts in Scientific Work*. Vol. 3, 203–32. Ed. G. P. Patil. University Park: Penn. State University Press.

FELLER, W. (1943). "On a general class of contagious distributions." *Ann. Math. Statist.*, **14**, 389–400.

FELLER, W. (1966). *An Introduction to Probability Theory and its Applications*. Vol. 2. New York: Wiley.

FELLER, W. (1968). *An Introduction to Probability Theory and its Applications*. Vol. 1. New York: Wiley.

GURLAND, J. (1957). "Some interrelations among compound and generalized distributions." *Biometrika*, **44**, 265–8.

HAIGHT, F. A. (1967). *Handbook of the Poisson Distribution*. New York: Wiley.

JOHNSON, N. L. and KOTZ, S. (1969). *Discrete Distributions*. Boston: Houghton Mifflin Co.

KEMP, C. D. (1970). "Accident proneness and discrete distribution theory." In: *Random Counts in Scientific Work*. Vol. 2, 41–66. Ed. G. P. Patil. University Park: Penn. State University Press.

MORAN, P. A. P. (1968). *An Introduction to Probability Theory*. Oxford: Oxford University Press.

PATIL, G. P. and JOSHI, S. W. (1968). *A Dictionary and Bibliography of Discrete Distributions*. Edinburgh: Oliver and Boyd.

RAO, C. R. (1965). *Linear Statistical Inference and its Applications*. New York: Wiley.