# Power Series Distributions and their Successors

**Adrienne W. (Freda) Kemp**

University of St Andrews, Scotland

## Resume

Most discrete statistical distributions are lattice distributions — here the values that the random variable (rv) can take lie on a lattice. The number of values can be either infinite, e.g. $0, 1, \dots$ for the Poisson distribution, or finite, e.g. $0, 1, \dots, n$ for the binomial distribution.

Classes of discrete distributions were studied intermittently pre-1960. However in the 1960's G. P. Patil made great advances in the theory of the power series subclass (PSD) and the generalized power series subclass (GPSD). PSD's have probability mass functions (pmf's) of the form

$$\Pr[X = x] = \frac{a_x \theta^x}{\eta(\theta)}, \qquad x = 0, 1, \dots, \qquad \theta > 0,$$

where $a_x \geq 0$, and $\eta(\theta) = \sum_{x=0}^{\infty} a_x \theta^x$. We call $\theta$ the power parameter of the distribution and $\eta(\cdot)$ its series function. For GPSD's the set of values that the variable can take is any nonempty enumerable set of nonnegative integers. The class of modified power series distributions (MPSD) was created by R. C. Gupta in the 1970's; these have pmf's of the form $\Pr[X = x] = a_x[u(\theta)]^x / \sum_x a_x[u(\theta)]^x$. They include distributions derived from Lagrangian expansions by P. C. Consul and his colleagues.

PSD's and GPSD's have many global properties (although some may fail to hold if $\theta = 1$). Beheading or truncating a GPSD creates another GPSD, as does the convolution of two GPSD's. Moreover, given a sample of independent observations from a PSD, then equating the observed and theoretical means gives the maximum likelihood estimate of $\theta$. If a pmf can be put into the form $\Pr[X = x] = \exp[a(\phi)b(x) + c(\phi) + d(x)]$, where $\phi$ is a parameter and $a(\cdot),\ b(\cdot),\ c(\cdot)$ and $d(\cdot)$ are known functions, then the distribution belongs to the important exponential family.

Much of the theory of PSD's and GPSD's can be deduced from the probability generating function (pgf)

$$G(z) = \sum_x a_x \theta^x z^x / \sum_x a_x \theta^x = \eta(\theta z)/\eta(\theta).$$

For the Poisson distribution $\eta(\theta) = e^\theta$ and for the binomial distribution $\eta(\theta) = (1+\theta)^n$, $n \in \mathbb{Z}^+$, $\theta = p/(1-p)$. Some researchers have focussed on the characteristic function (cf) $G(e^{it})$, and others on the Laplace transform $G(e^{-it})$.

The moment generating function (mgf) is

$$G(e^t) = \eta(\theta e^t)/\eta(\theta) = 1 + \sum_{r \geq 1} \mu_r^{'} t^r / r!$$

where $\mu_r^{'}$ are the uncorrected moments. Often it is easier to obtain the moment properties from the factorial moment generating function (fmgf)

$$G(1+t) = \eta(\theta + \theta t)/\eta(\theta) = 1 + \sum_{r \geq 1} \mu_{[r]}^{'} t^r / r!$$

where $\mu_{[r]}^{'}$ are the factorial moments, using the relationships $\mu = \mu_1^{'} = \mu_{[1]}^{'}$, $\mu_2 = \mu_2^{'} - \mu^2 = \mu_{[2]}^{'} + \mu - \mu^2$, etc. The cumulant generating function (cgf), $\ln G(e^t) = \sum_{r \geq 1} \kappa_r^{'} t^r / r!$, and the factorial cumulant generating function (fcgf), $\ln G(1+t)$, are also useful theoretically. For example, $\kappa_{r+1} = \theta \; d\kappa_r / d\theta$. So if $\mu = \kappa_1$ is known *as a function of* $\theta$, then all the other cumulants and hence all the moments are known.

The shapes of discrete distributions (dispersion, skewness, kurtosis) are usually described by the moment ratios $\mu_2/\mu,\ \mu_3/\mu_2^{3/2}$, and $\mu_4/\mu_2^2$. Expressions for the mode, median and mean deviation are often complicated.

The cumulative distribution function (cdf), $F_X(x) = \Pr[X \leq x] = \sum_{r \leq x} a_r \theta^r / \eta(\theta)$, is a step function with an enumerable number of steps. The survival function, hazard function (failure rate), mean residual life function, etc, are obtained from it in the usual way. There are simple relationships between certain PSD lower tail probabilities and the upper tail probabilities for related continuous distributions, e.g. Poisson and gamma tails, negative binomial and Beta tails.

PSD's, GPSD's, and MPSD's arise in several ways, including (i) sampling (e.g. the binomial and classical hypergeometric distributions), (ii) from the solution of difference equations, (iii) from stochastic processes (e.g. the assumption of exponential interarrival times leads to Poisson numbers of arrivals, the lack of memory process also characterizes the Poisson distribution), and (iv) the construction of new mathematical expressions for $\eta(\theta)$ when fitting data and when approximating other distributions.

Many different types of function for $\eta(\theta)$ have been found useful. When the function has many mathematical properties, the resultant distribution generally has many statistical properties, e.g. the exponential function, $\eta(\theta) = e^{\theta}$, and the Poisson distribution. The nineteenth century "wooden plough" of generalized hypergeometric functions was one of the first classes of mathematical functions to be explored. Lagrangian expansions have proved fruitful. Order $-k$ distributions are related to success runs models. Recently generalizations of hypergeometric series known as $q$-series have been investigated.

New distributions are constantly being created from existing ones by mixing, stopped sums, weighting, conditioning, and convolution. Many mixture and stopped sum distributions were originally created in response to problems in agriculture, biometry, industrial statistics and medicine. During and since the late 1900's the field of applications has expanded enormously; it now includes biostatistics, econometrics, astronomy and financial statistics. Of very great importance is our ability to gather enormous amounts of information and to process it using modern computers. This is leading to much research concerning discrete parametric and semiparametric regression models, for example the Tweedie-Poisson, Poisson polynomial, and double exponential models.