REFERENCES

[1] H. Hotelling, "Some new methods in matrix calculation," *Annals of Math Stat.*, Vol. 14 (1943), pp. 1–35.

[2] P. S. Dwyer, "The solution of simultaneous equations," *Psychometrika*, Vol. 6 (1941), p. 101.

A CLASS OF RANDOM VARIABLES WITH DISCRETE DISTRIBUTIONS

Albert Noack

Cologne, Germany

1. General results. A large class of random variables with discrete probability distributions can be derived from certain power series. Let

$$f(z) = \sum_{x=0}^{\infty} a_x z^x, \quad a_x \text{ real}, \quad |z| < r.$$

We may have either non-negative coefficients a_x or we may have $(-1)^x a_x \geq 0$. In the first case take $0 < z < r$; and in the second case take $-r < z < 0$. Define a random variable with the distribution

$$P\{\xi = x\} = \frac{a_x z^x}{f(z)}; \qquad x = 0, 1, 2, \cdots. \tag{1}$$

The above conditions insure $P\{\xi = x\} \geq 0$ for all x; besides

$$\sum_x P\{\xi = x\} = \frac{1}{f(z)} \sum_x a_x z^x = 1.$$

The distribution of ξ may be called the power series distribution (p.s.d.). The mean of such a distribution is

$$E(\xi) = \sum_x xP\{\xi = x\} = \frac{1}{f(z)} \sum_x x a_x z^x.$$

Hence it follows that

$$E(\xi) = z\frac{f'(z)}{f(z)} = z\frac{d}{dz}\log f(z). \tag{2}$$

We have for the moments about the origin

$$\mu_r' = \sum_x x^r P\{\xi = x\} = \frac{1}{f(z)} \sum_x x^r a_x z^x,$$

and hence

$$z\frac{d\mu_r'}{dz} = \frac{1}{f(z)} \sum_x x^{r+1} a_x z^x - z\frac{f'(z)}{f(z)} \frac{1}{f(z)} \sum_x x^r a_x z^x.$$

Thus we have the recurrence relation

$$(3) \qquad \mu'_{r+1} = z\frac{d\mu'_r}{dz} + \mu'_1\mu'_r .$$

The central moments are

$$\mu_r = \sum_x (x - \mu'_1)^r P\{\xi = x\} = \frac{1}{f(z)}\sum_x (x - \mu'_1)^r a_x z^x,$$

and hence

$$z\frac{d\mu_r}{dz} = \frac{1}{f(z)}\sum_x x(x - \mu'_1)^r a_x z^x - zr\frac{d\mu'_1}{dz}\frac{1}{f(z)}\sum_x (x - \mu'_1)^{r-1} a_x z^x$$
$$- z\frac{f'(z)}{f(z)}\cdot\frac{1}{f(z)}\sum_x (x - \mu'_1)^r a_x z^x.$$

The sum of the first and third term will be found to be μ_{r+1}, hence

$$z\frac{d\mu_r}{dz} = \mu_{r+1} - rz\frac{d\mu'_1}{dz}\mu_{r-1},$$

whence we have for the central moments of a p.s.d. the recurrence relation

$$(4) \qquad \mu_{r+1} = z\left[\frac{d\mu_r}{dz} + r\frac{d\mu'_1}{dz}\mu_{r-1}\right].$$

Putting $r = 1$, $\mu_0 = 1$, $\mu_r = 0$, we get the variance of ξ

$$(5) \quad \mu_2 = \sigma^2(\xi) = z\frac{d\mu'_1}{dz} = z^2\frac{d^2}{dz^2}\log f(z) + \mu'_1 = z^2\frac{f''(z)}{f(z)} - z^2\left[\frac{f'(z)}{f(z)}\right]^2 + z\frac{f'(z)}{f(z)}.$$

By (5), (4) assumes the form

$$(4') \qquad \mu_{r+1} = z\frac{d\mu_r}{dz} + r\mu_2\mu_{r-1}.$$

The characteristic function of ξ is

$$\varphi(t) = \sum_x e^{itx}P\{\xi = x\} = \frac{1}{f(z)}\sum_x a_x e^{itx}z^x,$$

or

$$(6) \qquad \varphi(t) = \frac{f(e^{it}z)}{f(z)}.$$

To get a relation connecting the cumulants κ_n and the moments μ'_r about the origin, we differentiate both sides of the identity

$$\sum_{r=1}^{\infty}\frac{\kappa_r}{r!}(it)^r = \log\sum_{\rho=0}^{\infty}\frac{\mu'_\rho}{\rho!}(it)^\rho$$

with respect to (it), identifying coefficients in $(it)^{r-1}$ we get[1]

$$\mu_r' = \sum_{j=1}^{r} \binom{r-1}{j-1} \mu_{r-j}' \kappa_j . \tag{7}$$

Differentiation of (7) with respect to z gives

$$\frac{d\mu_r'}{dz} = \sum_{j=1}^{r} \binom{r-1}{j-1} \left[\frac{d\mu_{r-j}'}{dz} \kappa_j + \mu_{r-j}' \frac{d\kappa_j}{dz} \right]. \tag{7'}$$

Substitution of (7) and (7′) in (3) gives

$$\sum_{j=1}^{r+1} \binom{r}{j-1} \mu_{r+1-j}' \kappa_j = \sum_{j=1}^{r} \binom{r-1}{j-1} \left\{ \left[z \frac{d\mu_{r-j}'}{dz} + \mu_1' \mu_{r-j}' \right] \kappa_j + z\mu_{r-j}' \frac{d\kappa_j}{dz} \right\},$$

or by (3) after a little re-arrangement

$$\kappa_{r+1} = z \sum_{j=1}^{r} \binom{r-1}{j-1} \mu_{r-j}' \frac{d\kappa_j}{dz} - \sum_{j=2}^{r} \binom{r-1}{j-2} \mu_{r+1-j}' \kappa_j . \tag{8}$$

2. Special cases.

(a) Choosing $f(z) = e^z$, ξ has Poisson-distribution

$$P\{\xi = x\} = \frac{z^x e^{-z}}{x!}. \tag{1a}$$

(2) and (5) are the well known relations $E(\xi) = \sigma^2(\xi) = z$; the recurrence formula (4) assumes the form[2]

$$\mu_{r+1} = z \left[\frac{d\mu_r}{dz} + r\mu_{r-1} \right]. \tag{4a}$$

(b) Taking $f(z) = (1 - z)^{-k}$, $k > 0$, $0 < z < 1$ we get the so-called negative binomial distribution

$$P\{\xi = x\} = \frac{\Gamma(k + x)}{x!\,\Gamma(k)} z^x (1 - z)^k, \qquad x = 0, 1, 2, \cdots . \tag{1b}$$

The mean is

$$E(\xi) = \frac{kz}{1 - z}, \tag{2b}$$

while the recurrence formula for the central moments is

$$\mu_{r+1} = z \left[\frac{d\mu_r}{dz} + \frac{rk}{(1 - z)^2} \mu_{r-1} \right], \tag{4b}$$

hence the first three moments of this distribution are

$$\sigma^2(\xi) = \mu_2 = \frac{kz}{(1 - z)^2},$$

[1] Cf. M. G. KENDALL, *The Advanced Theory of Statistics*, Vol. I, p. 87.

[2] Cf. CRAIG, *Am. Math. Soc. Bull.*, Vol. 40 (1934), p. 262.

(5b)
$$\mu_3 = \frac{kz(1+z)}{(1-z)^3},$$

$$\mu_4 = \frac{kz(1+4z+z^2+3kz)}{(1-z)^4}.$$

The characteristic function of the distribution is

(6b)
$$\varphi(t) = \left(\frac{1-e^{it}z}{1-z}\right)^{-k}.$$

Writing $z = \eta/(1+\eta)$, $k = h/\eta$, $\eta > 0$, $h > 0$ we get the so-called Polya-Eggenberger distribution for rare contagious events[3].

(1b₁)
$$w\{\xi = x\} = \frac{\Gamma\left(\frac{h}{\eta}+x\right)}{x!\,\Gamma(h\eta^{-1})}\left(\frac{\eta}{1+\eta}\right)^x (1+\eta)^{-h/\eta}, \qquad x = 0, 1, 2, \cdots.$$

The first four moments of this distribution are

(2b₁)
$$\mu_1' = h$$

(5b₁)
$$\mu_2 = h(1+\eta)$$

$$\mu_3 = h(1+\eta)(1+2\eta)$$

$$\mu_4 = h(1+\eta)[1+3(1+\eta)(h+2\eta)].$$

To obtain a recurrence relation for the moments consider

$$\frac{d\mu_r}{dz} = \frac{\partial\mu_r}{\partial\eta}\frac{d\eta}{dz} + \frac{\partial\mu_r}{\partial h}\frac{dh}{\partial z} = (1+\eta)^2\left[\frac{\partial\mu_r}{\partial\eta} + \frac{h}{\eta}\frac{\partial\mu_r}{\partial h}\right];$$

hence we find for this distribution by (4) and (4b)

(4b₁)
$$\mu_{r+1} = (1+\eta)\left[\eta\frac{\partial\mu_r}{\partial\eta} + h\frac{\partial\mu_r}{\partial h} + rh\mu_{r-1}\right].$$

It follows from (4b₁), that μ_r is a polynomial in η and h. The characteristic function of this distribution is

(6b₁)
$$\varphi(t) = [1+\eta(1-e^{it})]^{-h/\eta}.$$

(c) The coefficients of the series $-\log(1-z) = \sum_{x=1}^{\infty} z^x/x$ are positive; the associated distribution derived is

(1c)
$$P\{\xi = x\} = -\frac{z^x}{x\log(1-z)}, \qquad 0 < z < 1; \quad x = 1, 2, \cdots,$$

and has the mean

(2c)
$$E(\xi) = -\frac{z}{(1-z)\log(1-z)}.$$

[3] Cf. *Zeits. f. angew. Math. und Mech.*, Vol. 3 (1923), p. 279–289.

Recurrence formula (4) has for this distribution the form

(4c) $$\mu_{r+1} = z\left[\frac{d\mu_r}{dz} - r\frac{z + \log(1-z)}{(1-z)^2[\log(1-z)]^2}\mu_{r-1}\right],$$

while the variance and the characteristic function of this distribution are

(5c) $$\mu_2 = \sigma^2(\xi) = -\frac{z^2 + z\log(1-z)}{(1-z)^2[\log(1-z)]^2},$$

(6c) $$\varphi(t) = \frac{\log(1-e^{it}z)}{\log(1-z)}.$$

(d) The coefficients of the series $\log(1+z)/(1-z) = 2\sum_{x=1}^{\infty}(z^{2x+1})/(2x+1)$ are positive, so we can derive a random variable ξ with the distribution

(1d) $$P\{\xi = 2x+1\} = \frac{2z^{2x+1}}{(2x+1)\log\frac{1+z}{1-z}}, \quad 0 < z < 1, x = 1, 2, 3, \cdots.$$

ξ has the mean

(2d) $$E(\xi) = \frac{2z}{(1-z^2)\log\frac{1+z}{1-z}},$$

the recurrence formula (4) assumes the form

(4d) $$\mu_{r+1} = z\left(\frac{d\mu_r}{dz} + 2r\cdot\frac{(1+z^2)\log\frac{1+z}{1-z} - 2z}{(1-z^2)^2\left[\log\frac{1+z}{1-z}\right]^2}\mu_{r-1}\right),$$

while the variance and the characteristic function of ξ are

(5d) $$\sigma^2(\xi) = 2z\frac{(1+z^2)\log\frac{1+z}{1-z} - 2z}{(1-z^2)^2\left[\log\frac{1+z}{1-z}\right]^2},$$

(6d) $$\varphi(t) = \frac{\log(1+e^{it}z) - \log(1-e^{it}z)}{\log(1+z) - \log(1-z)}.$$

(e) Likewise the coefficients of the series

$$\sin^{-1} z = z + \sum_{x=1}^{\infty}\frac{1\cdot3\cdot5\cdots(2x-1)}{2\cdot4\cdot6\cdots(2x)}\frac{z^{2x+1}}{2x+1}$$

are positive, the derived variable ξ with the distribution

$$P\{\xi = 1\} = (\sin^{-1} z)^{-1},$$

(1e) $$P\{\xi = 2x + 1\} = \frac{1\cdot 3\cdots(2x-1)}{2\cdot 4\cdot 6\cdots(2x)} \cdot \frac{z^{2x+1}}{2x+1}(\sin^{-1} z)^{-1},$$
$$0 < z < 1, x = 1, 2, 3, \cdots,$$

has the mean

(2e) $$E(\xi) = \frac{z}{\sqrt{1-z^2}\sin^{-1} z}.$$

The recurrence formula for the moments

(4e) $$\mu_{r+1} = z\left[\frac{d\mu_r}{dz} + r\,\frac{\sin^{-1} z - z\sqrt{1-z^2}}{\sqrt{1-z^2}^3(\sin^{-1} z)^2}\,\mu_{r-1}\right]$$

gives the variance

(5e) $$\sigma^2(\xi) = z\,\frac{\sin^{-1} z - z\sqrt{1-z^2}}{\sqrt{1-z^2}^3(\sin^{-1} z)^2}.$$

The characteristic function assumes the form

(6e) $$\varphi(t) = \frac{\sin^{-1} e^{it} z}{\sin^{-1} z}.$$

(f) It is well known, that series (b), (c), (d), and (e) are special cases of the hypergeometric function $F(a, b, c; z)$. This function gives a p.s.d., if $abc > 0$. If $a > 0, b > 0, c > 0$ or if $a < 0, b < 0, c > 0$, a, b integers, there exist no further restrictions on these parameters. Suppose $a < 0, b < 0, c > 0$, a integer, b not, we must have $[b] \leq a$[4]; if neither a nor b are integers, we must have $[a] = [b]$. Suppose $a < 0, b > 0, c < 0$. If c is an integer, a must be an integer $> c$. If a is an integer, but c not, we must have $[c] \leq a$. Finally if neither a nor c are integers, we must have $[a] = [c]$. Corresponding conditions are valid, if $a > 0, b < 0, c < 0$. Regarding

$$\frac{d}{dz}F(a, b; c; z) = \frac{ab}{c}F(a+1, b+1; c+1; z),$$

the mean of a random variable ξ with hypergeometric distribution is

(2f) $$E(\xi) = z\frac{ab}{c}\,\frac{F(a+1, b+1; c+1; z)}{F(a, b; c; z)}.$$

Considering the differential equation

$$z(1-z)f''(z) + [c - (a+b+1)z]f'(z) - abf(z) = 0,$$

(5) gives the variance of ξ

(5f) $$\sigma^2(\xi) = \frac{ab}{c}\cdot\frac{z}{1-z}\left\{c + [1 - c + (a+b)z]\,\frac{F(a+1, b+1; c+1; z)}{F(a, b; c; z)} - z(1-z)\frac{ab}{c}\left[\frac{F(a+1, b+1; c+1; z)}{F(a, b; c; z)}\right]^2\right\}.$$

The higher moments of this distribution can now derived from (4′).

[4] $[b]$ means as usual the greatest integer $\leq b$.