Ann. Inst. Statist. Math.
31 (1979), Part A, 373-390

GENERALIZED HYPERGEOMETRIC, DIGAMMA AND TRIGAMMA DISTRIBUTIONS

MASAAKI SIBUYA

(Received Sept. 11, 1978; revised Sept. 21, 1979)

Summary

In this paper we introduce and study new probability distributions named "digamma" and "trigamma" defined on the set of all positive integers. They are obtained as limits of the zero-truncated Type B3 generalized hypergeometric distributions (inverse Pólya-Eggenberger or negative binomial beta distributions), and also by compounding the logarithmic series distributions.

The family of digamma distributions has the logarithmic series as a limit and the trigamma as another limit. The trigamma distributions are very close to the zeta (Zipf) distributions. Thus, our new distributions are useful as substitutes of the logarithmic series when the observed frequency data have such a long tail that cannot be fitted by the latter distributions.

In the beginning sections we summarize properties of the Type B3 generalized hypergeometric distributions. It is emphasized that the distributions are obtained by compounding a Poisson distribution by "gamma product-ratio" distributions.

1. Type B3 generalized hypergeometric distributions

Generalized hypergeometric distributions are discrete distributions defined on an integer interval having probabilities expressed as

$$\text{(1)} \qquad \Pr[X=x]=\frac{\Gamma(\gamma-\alpha)\Gamma(\gamma-\beta)}{\Gamma(\gamma-\alpha-\beta)\Gamma(\alpha)\Gamma(\beta)}\frac{\Gamma(x+\alpha)\Gamma(x+\beta)}{\Gamma(x+1)\Gamma(x+\gamma)},$$

which we shall refer to as $F(\alpha, \beta; \gamma)$. There are three major (or statistically useful) subfamilies in the generalized hypergeometric distributions:

Type A1. $F(-\xi, -n; \zeta)$, $\zeta>0$, $\xi>n-1$, $x=0, 1, \cdots, n$.
Type A2. $F(\xi, -n; -\zeta)$, $\xi>0$, $\zeta>n-1$, $x=0, 1, \cdots, n$.
Type B3. $F(\xi, \eta; \zeta)$, $\xi>0$, $\eta>0$, $\zeta>\xi+\eta$, $x=0, 1, 2, \cdots$.

The Type A1 distributions are positive hypergeometric distributions including ordinary hypergeometric; the Type A2 distributions are negative hypergeometric, Pólya-Eggenberger or binomial beta distributions; the Type B3 distributions are inverse Pólya-Eggenberger, generalized Waring or negative binomial beta distributions. In this paper we study exclusively the last Type B3 distributions. See Shimizu [14] and Sibuya and Shimizu [15] for the generalized hypergeometric distributions and their classification.

Changing parameters slightly, we shall deal with a probability distribution $F(\alpha, \beta; \alpha+\beta+\gamma)$, namely

$$p(x;\alpha,\beta,\gamma)=\frac{\Gamma(\alpha+\gamma)\Gamma(\beta+\gamma)}{\Gamma(\alpha+\beta+\gamma)\Gamma(\gamma)}\frac{(\alpha)_x(\beta)_x}{x!(\alpha+\beta+\gamma)_x}, \qquad x=0,1,2,\cdots;\ \alpha,\beta,\gamma>0, \tag{2}$$

where $(a)_x$ is Pochhammer's symbol

$$(a)_x=\begin{cases}1, & x=0,\\ a(a+1)\cdots(a+x-1), & x=1,2,\cdots.\end{cases}$$

We shall refer to the distribution simply as GHgB3. This subfamily is called inverse Pólya-Eggenberger being obtained by inverse sampling in Pólya's urn model (cf., e.g. Johnson and Kotz, [7]). It is called the generalized Waring because the special case $\alpha=1$ (or $\beta=1$) was named the Waring previously being obtained by Waring's expansion into inverse factorial series (Irwin, [4], [5], [6]). The Waring distributions contain the Yule distributions, the case $\alpha=\beta=1$ (cf. Simon, [16], [17]), and also the Mizutani distributions, the case $\beta=\gamma=1$ (cf. Mizutani, [11]). The name "negative binomial beta" is discussed in Section 2.

We describe some elementary facts on the GHgB3 of the expression (2) for completeness. It is unimodal with mode at the integral part of $m=(\alpha-1)(\beta-1)/(\gamma+1)$ or at $m-1$ and m if m is an integer, and has factorial moments, if $\gamma>r$,

$$\mathrm{E}[X^{(r)}]=(\alpha)_r(\beta)_r/(\gamma-1)^{(r)}, \qquad r=1,2,\cdots. \tag{3}$$

Central moments can be computed like those of the hypergeometric distributions (cf., e.g. Kendall and Stuart [8]),

$$\mu_2=\frac{\alpha\beta(\alpha+\gamma-1)(\beta+\gamma-1)}{(\gamma-2)(\gamma-1)^2}=\mu\left(1+\frac{\mu+\alpha+\beta+1}{\gamma-2}\right)>\mu$$

$$\mu_3=\frac{\mu_2(2\alpha+\gamma-1)(2\beta+\gamma-1)}{(\gamma-3)(\gamma-1)}=\frac{\mu_2}{\gamma-3}\left\{2(\gamma-2)\frac{\mu_2}{\mu}+2\mu-(\gamma-1)\right\} \tag{4}$$

$$\mu_4=\frac{\mu_2}{2(\gamma-4)}\left\{3(\gamma-3)\left(\frac{\mu_3}{\mu_2}\right)^2+6(\gamma-2)\mu_2-(\gamma-1)\right\} .$$

The ratio

$$p(x+1)/p(x)=(\alpha+x)(\beta+x)/(x+1)(\alpha+\beta+\gamma+x)$$

is increasing for sufficiently large x depending on parameter values. If the ratio is increasing, then the probabilities are log-convex in the sense that $p(x)\leqq\sqrt{p(x-1)p(x+1)}\leqq(p(x-1)+p(x+1))/2$. For some parameter values the ratio increases at all x, and $p(x)$ reveals J-shaped. This is the case if, e.g., $-(\alpha-1)(\beta-1)/(\alpha+\beta+\gamma-1)$ and $(\alpha+\gamma)(\beta+\gamma)/(\alpha+\beta+\gamma-1)$ are positive.

If the parameter α (or β) is an integer, then the sum of the probabilities (2) is expressed by that of a negative hypergeometric distribution, Type A2 generalized hypergeometric $F(\beta, -\alpha-x; -\gamma-\alpha-x+1)$;

$$(5)\qquad \sum_{s=0}^{x} p(s;\alpha,\beta,\gamma)=\sum_{s=\alpha}^{\alpha+x}\binom{-\beta}{\alpha+x-s}\binom{-\gamma}{s}\Big/\binom{-\beta-\gamma}{\alpha+x}$$
$$=\sum_{s=0}^{x}\binom{\beta+s-1}{s}\binom{\gamma+\alpha+x-s-1}{\alpha+x-s}\Big/\binom{\beta+\gamma+\alpha+x-1}{\alpha+x} .$$

This is shown from the fact that the cumulative sum of the negative binomial distribution can be expressed by that of the binomial distribution;

$$\sum_{s=0}^{x}\binom{\alpha+s-1}{s}p^{\alpha}(1-p)^{s}=\sum_{s=\alpha}^{\alpha+x}\binom{\alpha+x}{s}p^{s}(1-p)^{\alpha+x-s} .$$

Assume p to be a beta variable with parameters γ and β, take expectation with respect to p, and the terms of the left-hand side become GHgB3 probabilities as discussed below and the terms of the right-hand side become those of the expression (5).

2. GHgB3 distributions as compound Poisson

It is known that GHgB3, the expression (2), is obtained by assuming that the parameter p of a negative binomial distribution, NBn(α, p),

$$\binom{\alpha+x-1}{x}p^{\alpha}(1-p)^{x} , \qquad x=0, 1, 2, \cdots;\ \alpha>0,\ 0<p<1$$

is distributed as a beta distribution Be(γ, β) with the probability density

$$\frac{1}{B(\gamma,\beta)}p^{\gamma-1}(1-p)^{\beta-1} , \qquad 0<p<1;\ \beta,\gamma>0 .$$

That is, GHgB3 is a beta compound of a negative binomial. Symbolically, we express this fact by writing

$$(6)\qquad F(\alpha, \beta; \alpha+\beta+\gamma)=\text{GHgB3}=\text{NBn}(\alpha, p)\underset{p}{\wedge}\text{Be}(\gamma, \beta)\ .$$

It is also well known that the negative binomial distribution NBn $(\alpha, 1/(1+c))$ is obtained when a Poisson distribution, Po(λ),

$$e^{-\lambda}\lambda^{x}/x!\ , \qquad x=0, 1, 2, \cdots;\ \lambda>0\ ,$$

is compounded by a gamma distribution, Ga(α, c), with the density

$$\frac{1}{\Gamma(\alpha)c^{\alpha}}\lambda^{\alpha-1}e^{-\lambda/c}\ , \qquad 0<\lambda<\infty;\ \alpha, c>0\ .$$

If $p=1/(1+c)$ is distributed as Be(γ, β), then $c=(1-p)/p$ is distributed as the second type beta distribution, BeII(β, γ), with the density

$$\frac{1}{B(\beta, \gamma)}\frac{c^{\beta-1}}{(1+c)^{\beta+\gamma}}\ , \qquad 0<c<\infty;\ \beta, \gamma>0\ .$$

Thus, (6) is rewritten as

$$(7)\qquad F(\alpha, \beta; \alpha+\beta+\gamma)=\text{GHgB3}=\text{NBn}(\alpha, 1/1+c)\underset{c}{\wedge}\text{BeII}(\beta, \gamma)$$
$$=\text{Po}(\lambda)\underset{\lambda}{\wedge}\text{Ga}(\alpha, c)\underset{c}{\wedge}\text{BeII}(\beta, \gamma)\ .$$

This shows that GHgB3 is also a compound Poisson, since compounding is associative provided compounding operations are well-defined.

This fact was noticed in Irwin [5], where he fitted GHgB3 to accident data and analyzed the fluctuation of occurrences into three components; randomness, individual's internal proneness, and his external liability. Here, we are concerned with the Poisson compounder leading to GHgB3, on which Irwin wrote very briefly.

DEFINITION 1. Let V_α, V_β and V_γ be standard gamma random variables with shape parameter α, β and γ, respectively. We call the distribution of $V_\alpha V_\beta/V_\gamma$, a "gamma product-ratio" distribution, and refer to it as GaPR(α, β, γ).

THEOREM 1. *The generalized hypergeometric distribution* (2) *is a compound Poisson distribution compounded by a gamma product-ratio distribution*:

$$(8)\qquad F(\alpha, \beta; \alpha+\beta+\gamma)=\text{GHgB3}=\text{Po}(\lambda)\underset{\lambda}{\wedge}\text{GaPR}(\alpha, \beta, \gamma)\ .$$

PROOF. Compounding with respect to scale parameter is equivalent to obtaining the distribution of product of two random variables; the compounded with unit scale and the compounder. So our compounder is the product of a $\mathrm{Ga}(\alpha, 1)$ variable V_α, and a $\mathrm{BeII}(\beta, \gamma)$ variable which is the ratio of a $\mathrm{Ga}(\beta, 1)$ variable to a $\mathrm{Ga}(\gamma, 1)$ variable.

3. Gamma product-ratio distributions

From the genesis of $\mathrm{GaPR}(\alpha, \beta, \gamma)$, we have

$$(9)\qquad \begin{aligned}\mathrm{GaPR}(\alpha, \beta, \gamma) &= \mathrm{Ga}(\alpha, c) \bigwedge_c \mathrm{BeII}(\beta, \gamma) = \mathrm{Ga}(\beta, c) \bigwedge_c \mathrm{BeII}(\alpha, \gamma) \\ &= \mathrm{Ga}(\alpha, c) \bigwedge_{1/c} \mathrm{BeII}(\gamma, \beta) = \mathrm{Ga}(\beta, c) \bigwedge_{1/c} \mathrm{BeII}(\gamma, \alpha) .\end{aligned}$$

The last two expressions mean that the inverse scale of a gamma distribution follows a second type beta. Some other expressions are obtained by changing the order of compounds and by compounding three variables.

The gamma product-ratio distribution has the probability density.

$$(10)\qquad \begin{aligned} f(w) &= \frac{\Gamma(\beta+\gamma)}{\Gamma(\alpha)\Gamma(\beta)\Gamma(\gamma)} w^{\alpha-1} \int_0^\infty e^{-(w/t)} \frac{t^{\beta-\alpha-1}}{(1+t)^{\beta+\gamma}} dt \\ &= \frac{\Gamma(\beta+\gamma)}{\Gamma(\alpha)\Gamma(\beta)\Gamma(\gamma)} w^{\alpha-1} \int_0^\infty e^{-wt} \frac{t^{\alpha+\gamma-1}}{(1+t)^{\beta+\gamma}} dt \\ &= \frac{\Gamma(\alpha+\gamma)\Gamma(\beta+\gamma)}{\Gamma(\alpha)\Gamma(\beta)\Gamma(\gamma)} w^{\alpha-1} U(\alpha+\gamma, \alpha-\beta+1, w) ,\end{aligned}$$

where U is the function defined by the above integrations, is a solution of Kummer's equation, and relates to Kummer's confluent hypergeometric function (e.g., Slater [18]). In (10) α and β can be exchanged.

The family $\mathrm{GaPR}(\alpha, \beta, \gamma)$ has the moments of the order up to γ which are given by,

$$(11)\qquad F[X^r] = (\alpha)_r(\beta)_r/(\gamma-1)^{(r)} , \qquad r = 1, 2, \cdots .$$

This is equal to the factorial moments (3) of GHgB3, a general fact for a compound Poisson. Central moments are

$$(12)\qquad \begin{aligned} \mu_2 &= \frac{\mu(\mu+\alpha+\beta+1)}{\gamma-2} \\ \mu_3 &= \frac{2\mu}{\gamma-2}\{2\mu^2+3\mu(\alpha+\beta+\gamma)+\alpha^2+\beta^2+3(\alpha+\beta)+2\} \\ &= \frac{2}{\gamma-3}\left\{(\gamma-2)\frac{\mu_2^2}{\mu}+\mu_2(\mu+1)+\mu^2\right\} \\ \mu_4 &= \frac{3}{\gamma-4}\left\{\mu_3\left(\frac{\mu_2(\gamma-2)}{\mu}+2-(\gamma-4)\mu\right)+3\mu_2^2(\gamma-2)+2\mu_2\mu^2+6\mu_2\mu+2\mu^3\right\} .\end{aligned}$$

The probability density is unimodal. In fact, it is log-concave, since it is a compound of a gamma function with log-concave density. (See, e.g., Artin [1].) It is J-shaped if α and/or $\beta \leqq 1$.

4. Some limits of GHgB3

As was pointed out in Section 3, a GHgB3 distribution is not only a compound negative binomial but a compound Poisson. If the compounder degenerates into a one-point distribution, then the GHgB3 returns to the original negative binomial or Poisson. Therefore we can expect that GHgB3 approaches either a negative binomial or a Poisson distributions as the parameters tend to some limits. In fact, we have the following limiting processes:

(i) GHgB3$\to$NBn(α, p) as $\beta, \gamma \to \infty$ with $\gamma/(\beta+\gamma)=p$.

(ii) GHgB3$\to$Po(λ) as $\alpha, \beta, \gamma \to \infty$ with $\alpha\beta/\gamma=\lambda$.

Other limiting processes are as follows. We denote by $\varepsilon(x)$ a distribution degenerated into one point x.

(iii) GHgB3$\to\varepsilon(0)$ as $\gamma \to \infty$ while α and β remain finite.

(iv) GHgB3$\to\varepsilon(0)$ as β and/or $\alpha \to 0$ while γ remains finite.

(v) $p(x; \alpha, \beta, \gamma) \to 0$ for all x, as $\gamma \to 0$.

(vi) $p(x; \alpha, \beta, \gamma) \to 0$ for all x, as α and/or $\beta \to \infty$ while γ remains finite.

The process (v) and (vi) were further studied by Irwin [6]. We are concerned with (iii) and (iv). In the case (iii) if $x=0$ is truncated, the limit distribution degenerates again into $\varepsilon(1)$. While in the case (iv), the limit of zero-truncated distribution exists, which we shall study in the next section.

For $x \to \infty$, using Stirling's formula, we see that

$$p(x; \alpha, \beta, \gamma) \to \text{constant} \times x^{-(1+\gamma)} . \tag{13}$$

This behavior is regarded as one of Zipf's laws, so any justification of the law supports partially the application of GHgB3. It is disputable whether zeta or Zipf distributions (27) should be defined on positive integers or nonnegative integers. In the latter case, GHgB3 with smaller α and β, including Yule distributions, is close to zeta. In the former case, the following trigamma function is close to it.

5. Zero-truncated GHgB3

Before going into the discussion on the limits of zero-truncated distributions, we remark upon the range of its parameters. It is known that the range $0<\alpha<\infty$ of a negative binomial distribution NBn(α, p) is extended to $(-1, 0)\cup(0, \infty)$ when zero-truncated, the limit-case $\alpha=0$ being excluded for the logarithmic series distribution. The situation is similar to GHgB3.

THEOREM 2. *The zero-truncated probabilities*

$$q(x;\alpha,\beta,\gamma)=p(x;\alpha,\beta,\gamma)/(1-p(0;\alpha,\beta,\gamma)) \tag{14}$$

are positive (i) *if* $0>\alpha>\max(-\gamma,-1)$, (*or* $0>\beta>\max(-\gamma,-1)$), *or* (ii) *if* $0>\alpha,\beta>-1$ *and* $\alpha+\beta+\gamma>0$.

PROOF. Write

$$q(x;\alpha,\beta,\gamma)=\frac{p(0;\alpha,\beta,\gamma)}{1-p(0;\alpha,\beta,\gamma)}\frac{\alpha(\alpha+1)_{x-1}\beta_x}{(\alpha+\beta+\gamma)_x x!}, \qquad x=1,2,\cdots,$$

and remark that $p(0;\alpha,\beta,\gamma)<1$ implies

$$\Gamma(\alpha+\gamma)\Gamma(\beta+\gamma)<\Gamma(\gamma)\Gamma(\alpha+\beta+\gamma), \qquad \alpha,\beta,\gamma>0, \tag{15}$$

which means actually the log-convexity of the gamma function (cf. [1]).
If $-\gamma<\alpha<0$, then from (15)

$$\Gamma(\alpha+\gamma)\Gamma(\beta+\gamma)>\Gamma(\gamma)\Gamma(\alpha+\beta+\gamma), \qquad \beta,\gamma>0.$$

So, further if $\alpha+1>0$, then $q(x;\alpha,\beta,\gamma)>0$. Thus the case (i) is proved. The case (ii) is similarly proved. Actually this is the zero-truncation of Type B1 generalized hypergeometric distributions $F(-n+\varepsilon,-n+\delta;\zeta)$, $0<\varepsilon,\delta<1$, $\zeta>0$ and $n=1$.

6. Digamma and trigamma distributions

As mentioned previously, the zero probability $p(0;\alpha,\beta,\gamma)$ approaches one if β and/or α approach zero, and GHgB3 degenerates. However, the zero-truncated distribution $q(x;\alpha,\beta,\gamma)$ in (14) converges to a non-degenerate distribution. The new distribution is related to the digamma (psi) function $\psi(z)=d\log\Gamma(z)/dz$ and the trigamma function $\psi'(z)=d\psi(z)/dz$.

THEOREM 3.

$$\lim_{\beta\to 0} q(x;\alpha,\beta,\gamma)=q_1(x;\alpha,\gamma)=\frac{1}{\psi(\alpha+\gamma)-\psi(\gamma)}\frac{(\alpha)_x}{x(\alpha+\gamma)_x}, \tag{16}$$
$$x=1,2,\cdots;\ \gamma>0,\ \alpha>-1\ (\alpha\neq 0),\ \alpha+\gamma>0$$

and

$$\lim_{\alpha,\beta\to 0} q(x;\alpha,\beta,\gamma)=q_2(x;\gamma)=\frac{1}{\psi'(\gamma)}\frac{(x-1)!}{x(\gamma)_x}, \tag{17}$$
$$x=1,2,\cdots;\ \gamma>0.$$

In each case the ratio of both sides tends to 1 *uniformly in* x.

The proof is based on the Taylor expansion

$$\Gamma(z+u)=\Gamma(z)\left\{1+\psi(z)u+\frac{1}{2}(\psi'(z)+\psi^2(z))u^2+O(u^3)\right\} .$$

Because of the uniformity of convergence, q_1 and q_2 are probability distributions. It can be also proved that

$$\lim_{\alpha\to 0} q_1(x;\alpha,\gamma)=q_2(x;\gamma) .$$

DEFINITION 2. We call $q_1(x;\alpha,\gamma)$ in (16) and $q_2(x;\gamma)$ in (17) a "digamma" and a "trigamma" distributions, and refer to them as DGa(α, γ) and TGa(γ) respectively.

Theorem 3 gives a new proof of the formulas

$$\psi(\alpha+\gamma)-\psi(\gamma)=\sum_{n=1}^{\infty}\frac{(\alpha)_n}{n(\alpha+\gamma)_n} , \tag{18}$$

and

$$\psi'(\gamma)=\sum_{n=1}^{\infty}\frac{(n-1)!}{n(\gamma)_n} . \tag{19}$$

The formula (18) was shown in Nörlund [12], and the formula (19) was obtained by Matsunawa [10] and Ruben [13]. Our proof is essentially the same as Ruben's.

Distributions DGa(α, γ) and TGa(γ) are J-shaped. More precisely, $q_i(x)$ $(i=1, 2)$ are decreasing in x and $q_i(x)\leqq\sqrt{q_i(x-1)q_i(x+1)}$ for all x. The moments of DGa are obtained by considering the form of $F(\alpha, 1; \alpha+1+\gamma)$:

$$\begin{aligned} \mathrm{E}\,[X(\alpha+X)_r]&=\frac{1}{\psi(\alpha+\gamma)-\psi(\gamma)}\sum_{x=1}^{\infty}\frac{(\alpha)_{x+r}}{(\alpha+\gamma)_x} \\ &=\frac{1}{\psi(\alpha+\gamma)-\psi(\gamma)}\,\frac{(\alpha)_{r+1}}{\gamma-r-1} , \\ &\qquad r=0, 1, 2,\cdots;\ \gamma>r+1,\ \alpha>-1\ (\alpha\neq 0) . \end{aligned} \tag{20}$$

Or considering the form of $F(\alpha+r, r; \alpha+\gamma+r)$, we get the factorial moments,

$$\begin{aligned} \mathrm{E}\,[X^{(r)}]&=\frac{1}{\psi(\alpha+\gamma)-\psi(\gamma)}\sum_{x=r}^{\infty}\frac{(x-1)!(\alpha)_x}{(x-r)!(\alpha+\gamma)_x} \\ &=\frac{(r-1)!}{\psi(\alpha+\gamma)-\psi(\gamma)}\,\frac{(\alpha)_r}{(\gamma-1)^{(r)}} , \\ &\qquad r=1, 2,\cdots;\ \gamma>r,\ \alpha>-1\ (\alpha\neq 0) . \end{aligned} \tag{21}$$

From either of these expressions,

$$\begin{aligned}
&\mu=\frac{1}{\psi(\alpha+\gamma)-\psi(\gamma)}\frac{\alpha}{\gamma-1}\,, \\
&\mu_2'=\mu\Big[1+\frac{\alpha+1}{\gamma-2}\Big] \\
(22)\quad &\mu_3'=\mu\Big[1+\frac{2}{\gamma-3}(\alpha+1)(\alpha+2)-\frac{1}{\gamma-2}(\alpha+1)(2\alpha+1)\Big] \\
&\mu_4'=\mu\Big[1+\frac{3}{\gamma-4}(\alpha+1)(\alpha+2)(\alpha+3)-\frac{6}{\gamma-3}(\alpha+1)^2(\alpha+2) \\
&\qquad +\frac{1}{\gamma-2}(\alpha+1)(3\alpha^2+3\alpha+1)\Big]\,.
\end{aligned}$$

The moments of TGa are obtained from Waring's expansion;

$$(23)\qquad \mathrm{E}\,[X(X)_r]=\frac{1}{\psi'(\gamma)}\frac{r!}{\gamma-r-1}\,,\qquad \gamma>r+1;\ r=0,1,2,\cdots\,.$$

Or, considering the form of $F(r, r; \gamma+r)$, we get the factorial moments,

$$(24)\qquad \mathrm{E}\,[X^{(r)}]=\frac{1}{\psi'(\gamma)}\frac{((r-1)!)^2}{(\gamma-1)^{(r)}}\,,\qquad r=1,2,\cdots;\ \gamma>r\,.$$

From either of these,

$$\begin{aligned}
&\mu=\frac{1}{\psi'(\gamma)(\gamma-1)}\,,\qquad \mu_2'=\frac{1}{\psi'(\gamma)(\gamma-2)}=\mu\frac{\gamma-1}{\gamma-2}\,, \\
(25)\quad &\mu_3'=\frac{1}{\psi'(\gamma)}\Big[\frac{2}{\gamma-3}-\frac{1}{\gamma-2}\Big]=\mu\frac{(\gamma-1)^2}{(\gamma-2)^{(2)}}\,, \\
&\mu_4'=\frac{1}{\psi'(\gamma)}\Big[\frac{6}{\gamma-4}-\frac{6}{\gamma-3}+\frac{1}{\gamma-2}\Big]=\mu\frac{\gamma(\gamma-1)^2}{(\gamma-2)^{(3)}}\,.
\end{aligned}$$

As mentioned above, the trigamma is a limit of the digamma:

$$\mathrm{DGa}\,(\alpha,\gamma)\to\mathrm{TGa}\,(\gamma)\qquad(\alpha\to 0)\,.$$

Remark that this limit process is not restricted to positive α, thus the trigamma is a boundary between the digamma distributions with positive α and negative α.

On the other hand, if $\alpha,\gamma\to\infty$ keeping $p=\gamma/(\alpha+\gamma)$ constant, then the limit is a logarithmic series distribution (28);

$$\mathrm{DGa}\,(\alpha,\gamma)\to\mathrm{LSr}\,(1-p)\,.$$

This is a natural result because of the parallelism between NBn and

LSr on one hand and between GHgB3 and DBa/TGa on the other hand. The parallelism becomes clearer in Section 7.

For negative α another limit is obtained by letting $\alpha+\gamma\to 0$.

$$\lim_{\alpha+\gamma\to 0} q_1(x;\alpha,\gamma)=q_3(\gamma)=\frac{\gamma(1-\gamma)_{x-1}}{x!}, \qquad x=1,2,\cdots;\ 0<\gamma<1, \tag{26}$$

since $x\psi(x)\to -1$ as $x\to 0$. This is a right shift by one of $F(1,1-\gamma;2)$ and has a very long tail and does not have a finite mean.

A trigamma distribution is similar to a Zipf (or zeta) distribution, $\mathrm{Z}(\rho)$,

$$\frac{1}{\zeta(1+\rho)x^{1+\rho}}, \qquad x=1,2,\cdots;\ \rho>0. \tag{27}$$

Apparently $\mathrm{Z}(1)=\mathrm{TGa}(1)$ and the families are close to each other for parameters around $\rho=\gamma=1$. There is difference of at most several percent of probabilities between the closest $\mathrm{Z}(\theta)$ and $\mathrm{TGa}(\gamma)$ as discussed in Section 9. It is difficult to distinguish them practically from a sample of moderate size.

7. Other geneses of digamma and trigamma distributions

It is interesting to note that DGa and TGa are obtained by compounding the logarithmic series distribution, $\mathrm{LSr}(\theta)$

$$\frac{1}{-\log(1-\theta)}\frac{\theta^x}{x}, \qquad x=1,2,\cdots;\ 0<\theta<1. \tag{28}$$

As the compounder we need,

Definition 3. We call a distribution with the probability density

$$h(p;\lambda,\mu)=\frac{1}{C(\lambda,\mu)}(-\log p)p^{\lambda-1}(1-p)^{\mu-1}, \qquad 0<p<1;\ \lambda>0,\ \mu\geqq 0, \tag{29}$$

where

$$C(\lambda,\mu)=\int_0^1(-\log p)p^{\lambda-1}(1-p)^{\mu-1}dp = \begin{cases} B(\lambda,\mu)\{\psi(\lambda+\mu)-\psi(\lambda)\}, & \lambda,\mu>0, \\ \psi'(\lambda), & \lambda>0,\ \mu=0. \end{cases} \tag{30}$$

"(zero-) end accented beta" distribution, and refer to it as $\mathrm{EABe}(\lambda,\mu)$.

By symmetry of the beta distribution, we can define the one end accented beta with factor $-\log(1-p)$. Remark that parameter value $\mu=0$ is included. This is, in fact, a probability density function as shown below.

THEOREM 4. *Digamma and trigamma distributions are compound logarithmic series distribution in the sense that*

$$\text{DGa}(\alpha, \gamma)=\text{LSr}(1-p)\underset{p}{\wedge}\text{EABe}(\gamma, \alpha)\ , \tag{31}$$

and

$$\text{TGa}(\gamma)=\text{LSr}(1-p)\underset{p}{\wedge}\text{EABe}(\gamma, 0)\ .$$

PROOF. Consider the integral

$$\int_0^1 \frac{1}{-\log p}\frac{(1-p)^x}{x}(-\log p)p^{\lambda-1}(1-p)^{\mu-1}dp=\frac{B(\lambda, \mu+x)}{x} \tag{32}$$
$$=\begin{cases} B(\lambda, \mu)\{\psi(\mu+\lambda)-\psi(\lambda)\}q_1(x; \mu, \lambda)\ , & \lambda, \mu>0\ , \\ \psi'(\lambda)q_2(x; \lambda)\ , & \lambda=0,\ \mu>0\ , \end{cases}$$
$$x=1, 2,\cdots .$$

We sum up the integrand and the right-hand side over $x=1, 2,\cdots$, since the integral is uniformly convergent with respect to x. We get (30) as the result, showing that (29) is a probability density.

Kendall and Stuart [8] computed the integral (30) for $\lambda, \mu>0$, to obtain geometric mean of the beta distribution, by differentiating Euler's first integral $B(\lambda, \mu)$ with respect to λ. Changing the integration variable, we have

$$C(\lambda, \mu)=\int_0^\infty te^{-\lambda t}(1-e^{-t})^{\mu-1}dt\ , \tag{33}$$

whose $\mu=0$ case is very well known.

It is known that a logarithmic series distribution is obtained by compounding a zero-truncated Poisson distribution. In fact, if the parameter λ of a zero-truncated Poisson

$$\frac{e^{-\lambda}}{1-e^{-\lambda}}\lambda^x/x!\ , \qquad x=1, 2,\cdots,\ \lambda>0\ , \tag{34}$$

is a random variable with the probability density

$$(1-e^{-\lambda})\lambda^{-1}e^{-\lambda/\omega}/\log(1+\omega)\ , \tag{35}$$

then the compounded distribution is a logarithmic series distribution $\text{LSr}(\omega/(1+\omega))$.

This fact, together with the above discussion, shows a digamma distribution to be a compounded zero-truncated Poisson distribution.

THEOREM 5. *If the parameter λ of a zero-truncated Poisson distribution is a random variable with the probability density function*

$$h(\lambda;\alpha,\beta)=\frac{(1-e^{-\lambda})^{\lambda-1}}{C(\gamma,\alpha)}\int_0^\infty e^{-\omega\lambda}\frac{\omega^{\gamma-1}}{(1+\omega)^{\alpha+\gamma}}d\omega$$
$$=\frac{(1-e^{-\lambda})^{\lambda-1}}{C(\gamma,\alpha)}U(\gamma,-\alpha,\lambda)\,,\qquad \lambda>0\,;\ \alpha,\gamma>0\,, \tag{36}$$

where $C(\gamma,\alpha)$ was defined by (30) *and $U(\gamma,-\alpha,\lambda)$ was introduced in* (10), *then the compound distribution is* DGa(α,γ). *Compounding by $h(\lambda;0,\gamma)$ we obtain* TGa(γ).

The proof is by direct computation. The fact that h of (36) is a probability density function is also checked by direct computation.

Let X be a random variable with the probabilities $\Pr[X=x]=p(x-1;1,\alpha+1,\gamma-1)$, $x=1,2,\cdots$, $\alpha>0$, $\gamma>1$, which is a GHgB3 distribution with particular parameter values and right shifted by one. If $X=x$ is observed, then it is retained with a probability c/x and completely neglected with the complementary probability $1-c/x$. The resulting probability distribution of X is

$$\frac{\Gamma(\gamma)\Gamma(\alpha+\gamma)}{\Gamma(\alpha+\gamma+1)\Gamma(\gamma-1)}\,\frac{(1)_{x-1}(\alpha+1)_{x-1}}{(\alpha+\gamma+1)_{x-1}(x-1)!}\times\frac{c}{x}=\frac{c(\gamma-1)}{\alpha}\,\frac{(\alpha)_x}{x(\alpha+\gamma)_x}\,,$$

which is DGa(α,γ) if the probabilities are normalized.

This process is similar to a genesis of the logarithmic series from a shifted geometric distribution.

It is instructive to compare chance mechanisms generating the digamma distributions and those generating the logarithmic series distributions (cf. Boswell and Patil [2]).

8. Parameter estimation

We give here some estimators for the digamma and the trigamma distributions without discussing their properties, which are not fully known and further investigations are desired.

For the parameters α and γ of DGa(α,γ), the first estimator is based on the factorial moments $m_r=\sum_i x_i^{(r)}$, $i=1$, 2, and 3. To avoid solving nonlinear equations we use the three moments rather than the first two, and obtain the linear equations,

$$m_2\gamma-m_1\alpha=m_1+2m_2\,,\qquad m_3\gamma-2m_2\alpha=4m_2+3m_3\,, \tag{37}$$

which cannot be used when the population moments do not exist.

The second estimator is based on the relative sample frequencies f_x for sample values $x=1$, 2, and 3. Here also we use the three frequencies rather than the first two to obtain the linear equations,

$$(2f_2-f_1)\alpha+2f_2\gamma+2f_2-f_1=0\ , \tag{38}$$
$$(3f_3-2f_2)\alpha+3f_3\gamma+6f_3-4f_2=0\ .$$

If one does not want to use the higher sample moments but the sample mean, he can use m_1, f_1 and f_2 and compute the third estimator

$$\gamma=\left(\frac{m_1}{f_1}-1\right)\Big/\left(\frac{m_1}{f_1}-\frac{f_1}{f_1-2f_2}\right)\ ,\qquad \alpha=\frac{2f_2}{f_1-2f_2}\gamma-1\ . \tag{39}$$

These estimators may be used also as the starting values for solving iteratively the maximum likelihood equations

$$\frac{\psi'(\gamma)}{\psi(\alpha+\gamma)-\psi(\gamma)}=\sum_{r=1}^{\infty}\frac{v_r}{\alpha+r-1}\ , \tag{40}$$
$$-\frac{\psi'(\alpha+\gamma)-\psi'(\gamma)}{\psi(\alpha+\gamma)-\psi(\gamma)}=\sum_{r=1}^{\infty}\frac{v_r}{\alpha+\gamma+r-1}\ ,$$

where $v_r=\sum_{j=r}^{\infty}f_j$ are the relative frequencies of observations such that $x_i\geqq r$ $(1=v_1\geqq v_2\geqq v_3\geqq\cdots)$.

Some estimators for the parameter γ of TGa(γ) based on the same principle as above and using m_r and/or f_x are

$$\gamma=\frac{m_1}{m_2}+2\ , \tag{41}$$

$$\gamma=\frac{f_1}{2f_2}-1\ ,$$

and

$$\gamma=\frac{m_1}{m_1-f_1}\ .$$

The maximum likelihood equation is

$$-\frac{\psi''(\gamma)}{\psi'(\gamma)}=\sum_{r=1}^{\infty}\frac{v_r}{\gamma+r-1}\ . \tag{42}$$

9. Graphs showing features of digamma and trigamma distributions

It is natural to ask how digamma and trigamma distributions are close to or different from logarithmic series or zeta distributions. Some graphs are presented to answer the question.

Figure 1 shows the graphs of (p_1, p_2), where $p_x = \Pr[X=x]$, for logarithmic, digamma and trigamma distributions. The graph for zeta

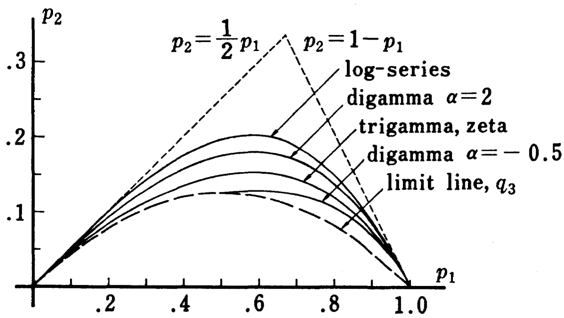

Fig. 1. (p_1, p_2) curves

distribution is almost the same as that for trigamma. For all distributions $p_2 \doteqdot p_1/2$ as $p_1 \to 0$ while $p_2 \doteqdot 1 - p_1$ evidently as $p_1 \to 1$. For digamma distributions with negative α the value of p_1 is limited by $p_1 > \gamma = -\alpha$. At this limit $p_2 = p_1(1-p_1)/2$ whose graph is shown by a broken line and corresponds to the limit distribution $q_3(x)$ of (26).

Figure 2 shows the graphs of $(r_1 = p_2/p_1,\ r_2 = p_3/p_2)$. All of them move from the origin (the limit $p_1 \to 1$, $p_2 \to 0$ and $p_3 \to 0$) to the point (1/2, 2/3) (the limit $p_1, p_2, p_3 \to 0$). Analytic expression of the graphs are,

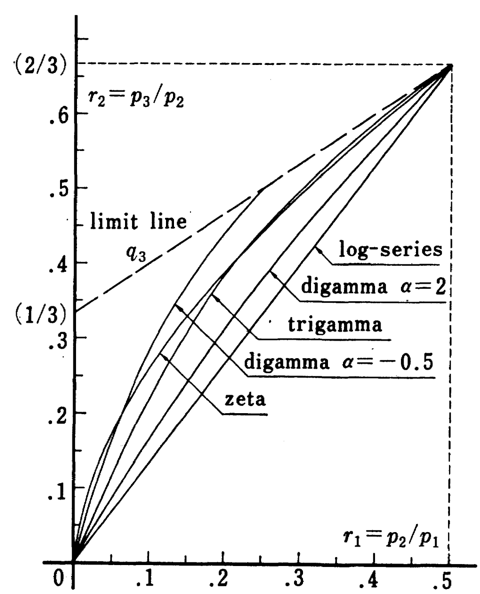

Fig. 2. (r_1, r_2) curves $(r_1 = p_2/p_1,\ r_2 = p_3/p_2)$

$$
(43) \qquad \begin{array}{lll} & \text{for LSr} & r_2=4r_1/3 \ , \\ & \text{for TGa} & r_2=8r_1/3(1+2r_1) \ , \\ & \text{for DGa} & r_2=4(2+\alpha)r_1/3(1+\alpha+2r_1) \ , \\ \text{and} & \text{for Z} & \log r_2=(\log_2 3-1)\log r_1 \ . \end{array}
$$

The graphs for digamma distributions with negative α values are limited by that for q_3 which is a straight line shown by a broken line.

It is known that for logarithmic series probabilities $(x+1)p_{x+1}/p_x$ is linear in x, and this fact characterizes the distribution. For other distributions the ratio is as follows.

$$
(44) \qquad \begin{array}{lll} & \text{for LSr} & \theta x \ , \\ & \text{for TGa} & x^2/(\gamma+x) \ , \\ & \text{for DGa} & x(\alpha+x)/(\alpha+\gamma+x) \ , \\ \text{and} & \text{for Z} & x^{1+\rho}/(1+x)^{\rho} \ . \end{array}
$$

Table 1 shows that actually all the values are so close that drawing the points are almost meaningless if the parameter values are suitably adjusted. In Table 1 they are chosen so that $p_1=0.4$ or 0.2. For smaller values of p_1 the ratios are closer. It is not practical to use the ratios of sample frequencies to identify its population distribution.

Factorial moments of the first and the second order in Figure 3 show the difference among these distributions. Mean-variance curves behave similarly. Logarithmic series distributions have finite mean and vari-

Table 1 $(x+1)p_{x+1}/p_x$ for logarithmic series, digamma, trigamma and zeta distributions

$p_1=0.4$

x	LSr	DGa		TGa	Z
		$\alpha=6$	$\alpha=2$		
1	.893	.839	.786	.671	.692
2	1.785	1.712	1.661	1.606	1.613
3	2.678	2.609	2.579	2.578	2.575
4	3.571	3.525	3.521	3.563	3.554
5	4.463	4.454	4.478	4.553	4.539
6	5.356	5.394	5.444	5.546	5.529
7	6.249	6.343	6.418	6.541	6.521
8	7.141	7.298	7.396	7.538	7.516
9	8.034	8.258	8.378	8.535	8.511
10	8.926	9.223	9.363	9.532	9.507
15	13.390	14.096	14.313	14.525	14.495
20	17.853	19.015	19.284	19.521	19.489
25	22.316	23.959	24.266	24.519	24.485
30	26.779	28.918	29.254	29.517	29.483
	LSr (.8926)	DGa (6, 1.3471)	DGa (2, .8165)	TGa (.4908)	Z (.5303)

Table 1 (Continued)

$p_1=0.2$

x	LSr	DGa $\alpha=6$	DGa $\alpha=2$	TGa	Z
1	.993	.952	.915	.826	.854
2	1.986	1.916	1.870	1.809	1.824
3	2.979	2.887	2.842	2.803	2.810
4	3.972	3.864	3.823	3.799	3.802
5	4.965	4.845	4.810	4.797	4.797
6	5.958	5.829	5.799	5.796	5.794
7	6.951	6.816	6.791	6.795	6.791
8	7.944	7.804	7.784	7.794	7.789
9	8.937	8.794	8.779	8.794	8.787
10	9.930	9.785	9.774	9.793	9.786
15	14.895	14.753	14.759	14.792	14.782
20	19.860	19.733	19.751	19.791	19.780
25	24.826	24.720	24.746	24.791	24.778
30	29.791	29.710	29.742	29.790	29.778
	LSr (.9930)	DGa (6, .3517)	DGa (2, .2771)	TGa (.2112)	Z (.2269)

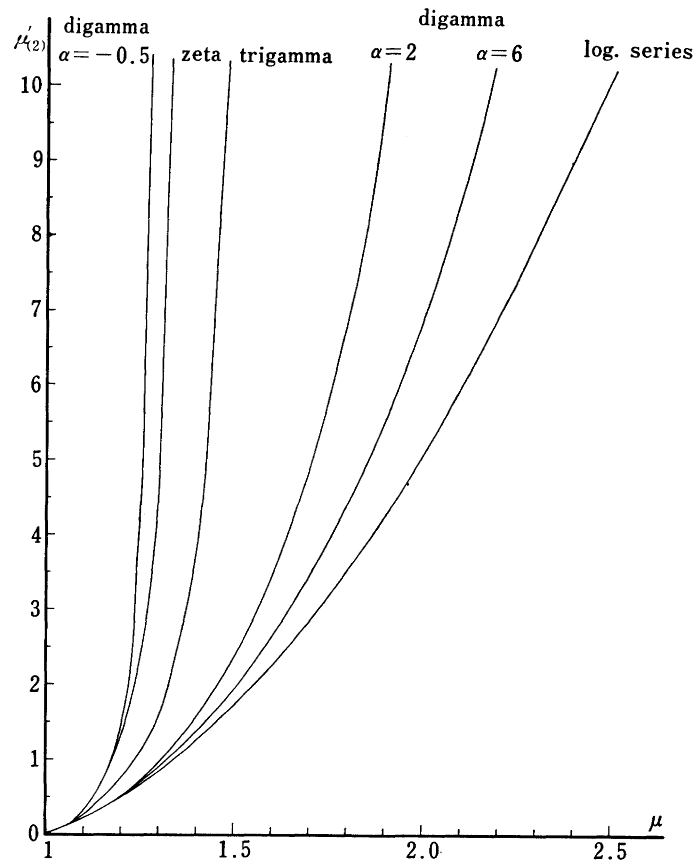

Fig. 3. Factorial moments of the first and the second orders

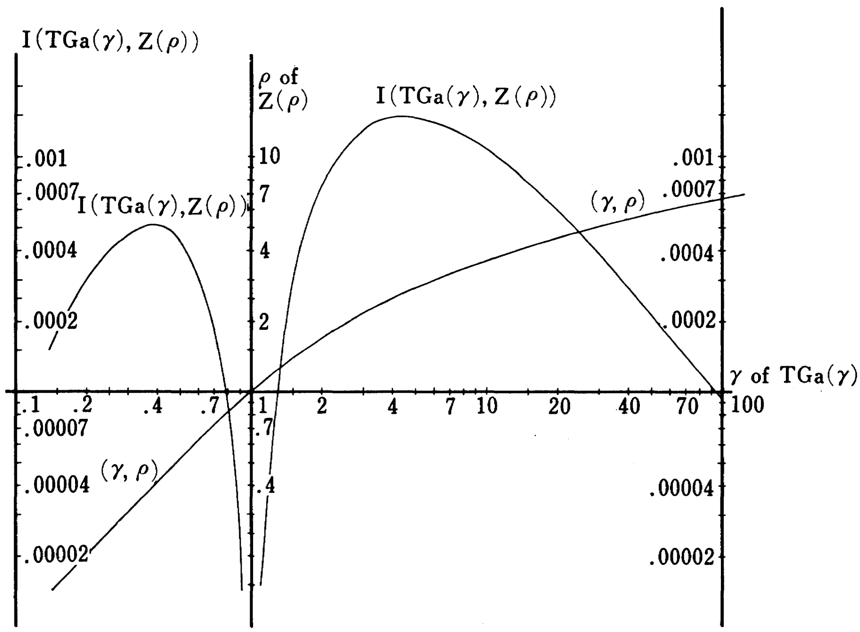

Fig. 4. The closest TGa (γ) and Z (ρ) and the minimum entropy distance I (TGa (ρ), Z (γ))

ance for all parameter values. Trigamma, zeta and digamma with negative α have relatively larger variance even if it is finite.

Figure 4 shows the closeness between zeta and trigamma distributions. The distance from a probability distribution $f(x)$ (here TGa(γ)) to another $g(x)$ (here Z(ρ)) is measured by the entropy distance

$$\mathrm{I}(f, g)=\sum_{x=1}^{\infty} f(x) \log (f(x)/g(x)) . \tag{45}$$

The distance of another direction $\mathrm{I}(g, f)$ can be used as well as this. For a given value of γ of TGa(γ) we obtain the value of ρ of Z(ρ) minimizing the distance (45) and the minimized distance, and plot both values in Figure 4. To simplify computation of the distance for smaller values of γ and ρ, terms for $\Pr[X>1000]$ are grouped together and distances between the grouped distributions are computed. For larger values of θ and γ the probabilities concentrate to smaller values of x and they match well if the parameter values are suitably chosen. For smaller values of θ and γ, all probabilities are smaller and they match well when global tendencies are similar.

For computing the digamma and the trigamma functions refer to de Medeiros and Schwachheim [3], and for the zeta function refer to Markman [9].

In summary, digamma distributions with trigamma as a limit are situated between logarithmic series and zeta distributions. So they are useful for data which cannot be fitted by the latter distributions.

IBM JAPAN

REFERENCES

[1] Artin, E. (1964). *The Gamma Function*, Holt, Rinehart and Winston.

[2] Boswell, M. T. and Patil, G. P. (1971). Chance mechanisms generating the logarithmic series distribution used in the analysis of number of species and number of individuals, *Statistical Ecology*, Vol. I, (eds. G. P. Patil, E. C. Pielu and W. E. Waters), Pennsylvania State University Press, 99-130.

[3] de Medeiros, A. T. and Schwachheim, G. (1969). Polygamma functions with arbitrary precision, Algorithm 349, *Commun. Ass. Comput. Math.*, **12**, 213-214.

[4] Irwin, J. O. (1965). Inverse factorial series as frequency distributions, Classical and Contagious Discrete Distributions, *Proc. Intnl. Symposium, Montreal*, 1963, Calcutta Statistical Publ., 159-174.

[5] Irwin, J. O. (1968). The generalized Waring distribution applied to accident theory, *J. Roy. Statist. Soc.*, A, **131**, 205-225.

[6] Irwin, J. O. (1975). The generalized Waring distribution, *J. Roy. Statist. Soc.*, A, **138**, Part I, 18-31; Part II, 205-225; Part III, 374-384.

[7] Johnson, N. L. and Kotz, S. (1977). *Urn Models and Their Application*, John Wiley, New York.

[8] Kendall, M. G. and Stuart, A. (1958). *The Advanced Theory of Statistics*, Vol. 1, Charles Griffin Co.

[9] Markman, B. (1965). The Riemann zeta function, *BIT*, **5**, 138-141.

[10] Matsunawa, T. (1976). Some inequalities based on inverse factorial series, *Ann. Inst. Statist. Math.*, **28**, 291-305.

[11] Mizutani, S. (1953). Vocabulary in women's magazines, *Kokken-Hokoku*, **4**, National Language Research Institute, Tokyo, (in Japanese).

[12] Nörlund, N. E. (1924). *Vorlesungen über Differenzenrechnung*, Springer, Berlin.

[13] Ruben, H. (1976). A note on the tri-gamma function, *Amer. Math. Monthly*, **83**, 622-623.

[14] Shimizu, R. (1968). Generalized hypergeometric distributions, *Proc. Inst. Statist. Math.*, **16**, 147-165, (in Japanese).

[15] Sibuya, M. and Shimizu, R. (1979). Classification of generalized hypergeometric distributions, to appear.

[16] Simon, H. A. (1955). On a class of skew distribution functions, *Biometrika*, **42**, 425-440.

[17] Simon, H. A. (1960). Some further notes on a class of skew distribution functions, *Inf. Control*, **3**, 80-88.

[18] Slater, L. J. (1964). Confluent hypergeometric functions, *Handbook of Mathematical Functions*, (eds. Abramowitz, M. and Stegun, I. A.), Dover, New York.