Ann. Inst. Statist. Math.
37 (1985), Part A, 1-16

THE EXACT AND APPROXIMATE DISTRIBUTIONS OF LINEAR COMBINATIONS OF SELECTED ORDER STATISTICS FROM A UNIFORM DISTRIBUTION

T. MATSUNAWA

(Received Nov. 24, 1983; revised Apr. 26, 1984)

Summary

The exact probability density function is given for linear combinations of $k=k(n)$ order statistics selected from whole order statistics based on random sample of size n drawn from a uniform distribution. Normal approximation to the linear combinations is made with the aid of Berry-Esseen's theorem. Necessary and sufficient conditions of the asymptotic normality for the statistic are obtained, too. An exact distribution and its normal approximation of linear combination of mutually independent gamma variables with integer valued parameters are also given as associated consequences.

1. Introduction

Distributions of linearly combined random variables have been often used to carry out statistical inference in various problems. In such distributions, those of the linear combinations of order statistics have taken important parts and many works have appeared from the aspect of limiting or asymptotic theory. (e.g. Chernoff, Gastwirth and Johns [2], Eiker and Puri [5], Hecker [6], Sholack [9], Stigler [10], van Zwet [12], etc.) On the other hand, the exact distributions of linear combinations of whole order statistics have not been investigated thoroughly without a few exceptions (e.g. Dempster and Kleyle [3]), still less those of *selected* order statistics. So far as the present author knows, only Weisberg [13] considered the exact cumulative distribution function of the linear combinations of *selected* order statistics from a uniform distribution, but his main interest was to give an algorithm for recursive computation of the distribution function. Thus, to find an closed form of the exact distribution has been unsettled even in such simple

Key words and phrases: Linear combination, order statistics, uniform distribution, exact distribution, normal approximation.

case of uniform distribution. Indeed, the case is not only theoretically interesting in itself, but also significant to the effect that we can obtain an exact distribution of linear combinations of jointly Dirichlet distributed random variables whose application will be useful in various statistical analysis. The purpose of this article is two fold. One is to derive the exact pdf. of the linear combinations of selected order statistics from a uniform distribution. Another is to investigate normal approximations to the exact distribution.

Let $U_1<U_2<\cdots<U_n$ be order statistics based on a random sample of size n from the uniform distribution $U(0,1)$. Suppose that among the above whole order statistics we select k order statistics $U_{n_1}<U_{n_2}<\cdots<U_{n_k}$, where $k=k(n)$, $1\leqq k\leqq n$ and $0<n_1<n_2<\cdots<n_k<n+1$. Let us denote the linear combination of this selected order statistics by

$$L_n=\sum_{l=1}^{k} a_l U_{n_l}\ , \tag{1.1}$$

where coefficients $a_l=a_l(n)$ $(l=1,\cdots,k)$ are real constants with $\sum_{l=1}^{k} a_l^2\neq 0$ for each k. Further, let

$$b_j=b_j(n)=\sum_{l=j}^{k} a_l\ ,\qquad (j=1,\cdots,k)\ , \tag{1.2}$$

$$d_j=d_j(n)=n_j-n_{j-1}-1\ ,\qquad (j=1,\cdots,k+1)\ , \tag{1.3}$$

$$V_j=U_{n_j}-U_{n_{j-1}}\ ,\qquad (j=1,\cdots,k+1)\ , \tag{1.4}$$

where in the above we have put the conventions $n_0=0$, $n_{k+1}=n+1$, $U_{n_0}=0$, $U_{n_{k+1}}=1$. Then, $U_{n_j}=V_j+V_{j-1}+\cdots+V_1$ and we can rewrite (1.1) as

$$L_n=\sum_{j=1}^{k} b_j V_j\ , \tag{1.5}$$

from which our problem is reduced to study the distributions of linear combinations of the Dirichlet random variables $(V_1,\cdots,V_k)$ with the pdf.

$$f_V(v_{(k)})=\frac{\Gamma(n+1)}{\prod_{j=1}^{k+1}\Gamma(d_j+1)}\prod_{j=1}^{k} v_j^{d_j}\cdot\Bigl(1-\sum_{j=1}^{k} v_j\Bigr)^{d_{k+1}}\ , \tag{1.6}$$

at any point in the simplex $\left\{v_{(k)}=(v_1,\cdots,v_k)\,|\,v_j\geqq 0,\ j=1,\cdots,k,\ \sum_{j=1}^{k} v_j\leqq 1\right\}$ in the k-dimensional real space R_k and zero outside.

In the next section we prepare some necessary lemmas for later analysis. The exact pdf. of the statistic L_n is derived in Section 3, where the exact pdf. of the linear combination of unordered mutually independent gamma random variables with integer-valued parameters is given, too. In Section 4 normal approximations are discussed to the above distributions. Some necessary and sufficient conditions of the asymptotic normality for L_n and related statistics are obtained in the final part of the section.

2. Preliminary lemmas

The following lemma enables us to represent the statistic L_n by a ratio of linear combinations of mutually independent unordered gamma variates. Let $\{Y_j\}$ $(j=1,\cdots,k+1)$ be such gamma variables whose pdf.'s are given by

$$g_j(y_j)=y_j^{d_j}e^{-y_j}/\Gamma(d_j+1) \qquad \text{for } y_j>0\,, \tag{2.1}$$

zero otherwise, $j=1,\cdots,k+1$ and put $S=Y_1+Y_2+\cdots+Y_{k+1}$. Then, we have

LEMMA 2.1. *The random vectors* $(V_1, V_2,\cdots, V_{k+1})$ *and* $(Y_1/S, Y_2/S, \cdots, Y_{k+1}/S)$ *are equidistributed. Further,* S *and* $(Y_1/S, Y_2/S,\cdots, Y_k/S)$ *are independently distributed according to the gamma distribution with mean* $n+1$ *and the Dirichlet distribution with the pdf.* (1.6), *resectively.*

Remark 2.1. In case of all $d_j=0$, $j=1,\cdots,k+1$, this lemma is reduced to the wellknown result on uniform spacings (see e.g. Section 4 in Pyke [8]).

PROOF. The pdf. of $Y_{(k+1)}=(Y_1, Y_2,\cdots, Y_{k+1})$ is expressed as

$$f_Y(y_{(k+1)})=\left\{\prod_{j=1}^{k+1} y_j^{d_j}/\Gamma(d_j+1)\right\}\cdot\exp\left(-\sum_{j=1}^{k+1} y_j\right) \tag{2.2}$$

at any point in $\{y_{(k+1)}=(y_1,\cdots,y_{k+1})\,|\,y_j\geqq 0,\ j=1,\cdots,k+1\}$ in R_{k+1} and zero outside. Let $Z_j=Y_j/S$ $(j=1,\cdots,k)$, $Z_{k+1}=S$ and correspondingly consider the transformation

$$z_j=y_j/(y_1+\cdots+y_{k+1})\,, \qquad (j=1,\cdots,k)\,,$$
$$z_{k+1}=y_1+\cdots+y_{k+1}\,,$$

then, since the Jacobian of the transformation $J((y_1,\cdots,y_{k+1})\to(z_1,\cdots,z_{k+1}))$ is equal to z_{k+1}^k, the joint pdf. of $Z_{(k+1)}=(Z_1, Z_2,\cdots, Z_{k+1})$ is

$$f_Z(z_{(k+1)}) = \begin{cases} \left\{\prod_{j=1}^{k+1}\dfrac{1}{\Gamma(d_j+1)}\prod_{j=1}^{k} z_j^{d_j}\cdot\left(1-\sum_{j=1}^{k} z_j\right)^{d_j+1}\right\}\left(1-\sum_{j=1}^{k} z_j\right)^{d_{k+1}}(z_{k+1}^n e^{-z_{k+1}}) \\ \qquad \text{for } z_j\geqq 0,\ j=1,\cdots,k+1 \text{ and } \sum_{j=1}^{k} z_j\leqq 1 \\ 0 \qquad \text{otherwise}\,. \end{cases} \tag{2.3}$$

Thus, it is easily seen that the variables Z_{k+1} and $Z_{(k)}=(Z_1,\cdots,Z_k)$ are independently distributed according to respective marginal pdf.'s

$$g(z_{k+1})=\frac{1}{\Gamma(n+1)}z_{k+1}^{n}e^{-z_{k+1}} \qquad \text{for } z_{k+1}>0\,, \tag{2.4}$$

zero otherwise, and

$$f_1(z_{(k)})=\frac{\Gamma(n+1)}{\prod_{j=1}^{k+1}\Gamma(d_j+1)}\prod_{j=1}^{k}z_j^{d_j}\cdot\left(1-\sum_{j=1}^{k}z_j\right)^{d_{k+1}}, \tag{2.5}$$

at any point in $\left\{z_{(k)}=(z_1,\cdots,z_k)\,|\,z_j\geqq 0,\ j=1,\cdots,k,\ \sum_{j=1}^{k}z_j\leqq 1\right\}$ in R_k and zero outside, which coincides with (1.6).

Incidentally, since $Y_1/S+Y_2/S+\cdots+Y_{k+1}/S=1=\left(\sum_{j=1}^{k+1}V_j\right)$, we have consequently proved the lemma.

Next lemma is an exact representation of partial fraction expansions in multiplicity cases, which will be very useful to many general problems required to use the expansions. Different from usual expressions (see, e.g. Doetsch [4]) coefficients of our expansion do not contain derivatives of the underlying function.

LEMMA 2.2. *Let s be a complex variable and $G(s)$ be a rational function of s:*

$$G(s)=[(s+\lambda_1)^{\nu_1}(s+\lambda_2)^{\nu_2}\cdots(s+\lambda_k)^{\nu_k}]^{-1}\,, \tag{2.6}$$

where ν_j $(j=1,\cdots,k)$ denote the multiplicity of the corresponding zeros of $G^{-1}(s)$, $-\lambda_j$ $(j=1,\cdots,k)$, which are assumed to be finite and distinct constants if $k\geqq 2$. We have then the following expansion:

$$G(s)=\sum_{l=1}^{k}\sum_{m=1}^{\nu_l}C_{l,m}(s+\lambda_l)^{-m}\,, \tag{2.7}$$

with coefficients $C_{1,\nu_1}=1$ and $C_{1,m}=0$ $(1\leqq m\leqq \nu_1-1)$ for $k=1$, and

$$C_{l,m}=\prod_{\substack{r=1\\ r\neq l}}^{k}\frac{1}{(\lambda_r-\lambda_l)^{\nu_r}}\sum_{q=0}^{\nu_l-m}\Sigma^*\prod_{j=0}^{\nu_l-m}\frac{1}{q_j!\,j^{q_j}}\left\{\sum_{\substack{p=1\\ p\neq l}}^{k}(-1)^j\frac{\nu_p}{(\lambda_p-\lambda_l)^j}\right\}^{q_j}, \quad (k\geqq 2)\,, \tag{2.8}$$

where Σ^* stands for the sum over all non-negative integers $\{q_0,q_1,\cdots,q_{\nu_l-m}\}$ such that $q_0+q_1+\cdots+q_{\nu_l-m}=q$ and $0q_0+1q_1+2q_2+\cdots+(\nu_l-m)q_{\nu_l-m}=\nu_l-m$, with $q_0=0$ for $\nu_l-m=0$.

PROOF. For $k=1$ the result is clear. For $k\geqq 2$, consider

(2.9) $$G_l(s)\equiv(s+\lambda_l)^{\nu_l}G(s)=\prod_{\substack{r=1\\r\neq l}}^{k}(s+\lambda_r)^{-\nu_r}\equiv\exp[h_l(s)]\,,$$

with

(2.10) $$h_l(s)=-\sum_{\substack{r=1\\r\neq l}}^{k}\nu_r\ln(s+\lambda_r)\,,\qquad(l=1,\cdots,k;\ l\neq r)\,,$$

where $\ln(\cdot)$ denotes the principal branch of the natural logarithmic function for complex variables. Applying limiting and differential techniques repeatedly to $G(s)$ it is easily seen that the coefficients of (2.7) can be firstly represented as

(2.11) $$C_{l,m}=G_l^{(\nu_l-m)}(-\lambda_l)/(\nu_l-m)!\ .$$

In respect to derivatives appeared in (2.11), making use of the Bell polynomials (see, e.g. Andrews [1]) we have

(2.12) $$G_l^{(\nu_l-m)}(-\lambda_l)=\sum_{q=0}^{\nu_l-m}\left[\frac{d^qG_l(h_l)}{dh_l^q}\right]_{h_l=h_l(-\lambda_l)}\times\Sigma^*\frac{(\nu_l-m)!}{q_1!\cdots q_{\nu_l-m}!}\left[\frac{1}{1!}\frac{dh_l(-\lambda_l)}{ds}\right]^{q_1}\cdots\left[\frac{1}{(\nu_l-m)!}\frac{d^{\nu_l-m}h_l(-\lambda_l)}{ds^{\nu_l-m}}\right]^{q_{\nu_l-m}},$$

where the second summation is extended over all non-negative integers $(q_0,q_1,\cdots,q_{\nu_l-m})$ such that $q_0+q_1+\cdots+q_{\nu_l-m}=q$ and $0q_0+1q_1+2q_2+\cdots+(\nu_l-m)q_{\nu_l-m}=\nu_l-m$. Incidentally, from (2.9) we get for all q

(2.13) $$\left[\frac{d^qG_l(h_l)}{dh_l^q}\right]_{h_l=h_l(-\lambda_l)}=G_l(-\lambda_l)=\prod_{\substack{r=1\\r\neq l}}^{k}(\lambda_r-\lambda_l)^{-\nu_r}\,,\qquad(l=1,\cdots,k)$$

and from (2.10)

(2.14) $$\frac{d^jh_l(s)}{ds^j}=\sum_{\substack{p=1\\p\neq l}}^{k}(-1)^j\nu_p\frac{(p-1)!}{(s+\lambda_p)^p}\,,\qquad(j=1,\cdots,\nu_l-m)\,.$$

Therefore, combining (2.11)–(2.14) with (2.10), we get the expression (2.8).

The following two complex integral formulas are wellknown and they will be respectively used in calculating a characteristic function and its inversion.

LEMMA 2.3. (i) *For* $a, c>0$ *and* $i=\sqrt{-1}$,

(2.15) $$\int_0^\infty e^{-(a+ib)x}x^{c-1}dx=\frac{\Gamma(c)}{(a+ib)^c}\,,\qquad-\infty<b<\infty\,.$$

(ii) *For* $a, c>0$ *and* $i=\sqrt{-1}$,

$$(2.16)\qquad \int_{-\infty}^{\infty}\frac{e^{ibx}}{(a+ix)^c}dx=\begin{cases}\dfrac{2\pi}{\Gamma(c)}\cdot\dfrac{b^{c-1}}{e^{ab}} & b>0\,,\\ 0\,, & b<0\,.\end{cases}$$

3. The exact distribution of the statistic L_n

In this section the exact pdf. of L_n will be derived by a characteristic function technique. In view of Lemma 2.1 the statistic L_n is equidistributed to

$$(3.1)\qquad T_n\equiv\sum_{j=1}^{k}b_jY_j\Big/\sum_{j=1}^{k+1}Y_j\equiv W/S\,,$$

where Y_j's ($j=1,\cdots,k+1$) are mutually independent gamma random variables with the pdf. given by (2.1). So, instead of L_n depending on "order", it is easy for us to consider the statistic T_n being free of the dependency.

The characteristic function of the joint random variable (W, S) is

$$\varphi_{W,S}(t_1,t_2)=\int_0^\infty\cdots\int_0^\infty\exp\left(it_1\sum_{\gamma=1}^{k}b_\gamma y_\gamma+it_2\sum_{\gamma=1}^{k+1}y_\gamma\right)\cdot\prod_{\gamma=1}^{k+1}\frac{y_\gamma^{d_\gamma}}{\Gamma(d_\gamma+1)}$$

$$\times\exp\left(-\sum_{\gamma=1}^{k+1}y_\gamma\right)dy_1\cdots dy_{k+1}$$

$$=\prod_{\gamma=1}^{k}\int_0^\infty\frac{1}{\Gamma(d_\gamma+1)}y_\gamma^{d_\gamma}\exp\left[-\{(1-it_2)-b_\gamma it_1\}y_\gamma\right]dy_\gamma$$

$$\times\int_0^\infty\frac{1}{\Gamma(d_{k+1}+1)}y_{k+1}^{d_{k+1}}\exp\left[-(1-it_2)y_{k+1}\right]dy_{k+1}\,,$$

$$(i^2=-1)\,.$$

Making use of the formula (2.15)

$$(3.2)\qquad \varphi_{W,S}(t_1,t_2)=\prod_{\gamma=1}^{k}\{1-i(b_\gamma t_1+t_2)\}^{-d_\gamma(+1)}\cdot(1-it_2)^{-(d_{k+1}+1)}\,.$$

Now, from the indices γ's ($1\leqq\gamma\leqq k$) pick up such ones that $b_\gamma\neq0$, and combine the terms among $\{1-i(b_\gamma t_1+t_2)\}^{-(d_\gamma+1)}$ ($1\leqq\gamma\leqq k$) if the corresponding b_γ's take the same value. Denote by r, b_r^* and d_r^* ($1\leqq r\leqq k^*$) the resultant quantities corresponding to γ, b_γ and d_γ, respectively, rearranged as the above. Then, the RHS. of (3.2) can be rewritten as

$$(3.3)\qquad \varphi_{W,S}(t_1,t_2)=\prod_{r=1}^{k^*}b_r^{*-(d_r^*+1)}\left(\frac{1}{b_r^*}-\frac{it_1}{1-it_2}\right)^{-(d_r^*+1)}\cdot(1-it_2)^{-(n+1)}\,,$$

which can be expanded into partial fractions by using Lemma 2.2. Replacing k, λ_r, ν_r and s, in the lemma, by k^*, $1/b_r^*$, d_r^*+1 and $-it_1/$

$(1-it_2)$, respectively, with additional changes, we have the following expansion

$$(3.4)\qquad \varphi_{W,S}(t_1, t_2)=\prod_{r=1}^{k^*} b_r^{*-(d_r^*+1)} \sum_{l=1}^{k^*} \sum_{m=1}^{d_l^*+1} C_{l,m}^* b_l^{*m}\{(1-it_2)-ib_l^* t_1\}^{-m} \times(1-it_2)^{-(n+1-m)},$$

where $C_{l,m}^*$ is the constant coefficients corresponding to (2.8).

Let us now invert the characteristic function with the aid of formula (2.16). We have the joint pdf. of (W, S) as

$$(3.5)\qquad f_{W,S}(w, s)=\sum_{l=1}^{k^*} \sum_{m=1}^{d_l^*+1} C_{l,m}^* \prod_{r=1}^{k^*} b_r^{*-(d_r^*+1)} \frac{1}{\Gamma(m)\Gamma(n-m+1)} \times\chi_l(w/b_l^*)\cdot\chi_l(s-w/b_l^*)w^{m-1}e^{-s}(s-w/b_l^*)^{n-m},$$

where χ_l stands for the unit impulse function defined by

$$(3.6)\qquad \chi_l(z)=\begin{cases} 1, & \text{if } z>0, \\ 0, & \text{if } z\leqq 0, \end{cases}$$

for $l=1, 2,\cdots, k^*$.

Thus, the pdf. of $T=W/S$ can be calculated as

$$\begin{aligned} f(t)&=\int_{-\infty}^{\infty} |s| f_{W,S}(ts, s)ds \\ &=\sum_{l=1}^{k^*} \sum_{m=1}^{d_l^*+1} C_{l,m}^* \prod_{r=1}^{k^*} b_r^{*-(d_r^*+1)}\cdot\frac{1}{\Gamma(m)\Gamma(n-m+1)} \\ &\quad\times\int_0^{\infty} \chi_l(ts/b_l^*)\chi_l(s-st/b_l^*)s^n e^{-s}ds\cdot t^{m-1}(1-t/b_l^*)^{n-m}. \end{aligned}$$

Namely, we have

$$(3.7)\qquad f(t)=\sum_{l=1}^{k^*} \sum_{m=1}^{d_l^*+1} C_{l,m}^\dagger \chi_l(t/b_l^*)\chi_l(1-t/b_l^*)t^{m-1}(1-t/b_l^*)^{n-m}/B(m, n-m+1),$$

with the coefficients $C_{1,d_1^*+1}^\dagger=1$ and $C_{1,m}^\dagger=0$ $(1\leqq m\leqq d_l^*)$ for $k^*=1$, and

$$(3.8)\qquad C_{l,m}^\dagger=b_l^{*\Sigma^\dagger} \prod_{\substack{r=1\\ r\neq l}}^{k^*} (b_l^*-b_r^*)^{-(d_r^*+1)} \sum_{q=0}^{d_l^*+1-m} \Sigma^* \sum_{j=0}^{d_l^*+1-m} \frac{1}{q_j!\, j^{q_j}} \times\left\{\sum_{\substack{p=1\\ p\neq l}}^{k^*} (-1)^j(d_p^*+1)\left(\frac{b_p^* b_l^*}{b_l^*-b_p^*}\right)^j\right\}^{q_j}, \qquad (k\geqq 2),$$

where Σ^* stands for the sum over all non-negative integers $\{q_0, q_1,\cdots, q_{d_l^*+1-m}\}$ such that $q_0+q_1+\cdots+q_{d_l^*+1-m}=q$ and $0q_0+1q_1+2q_2+\cdots+(d_l^*+1-m)q_{d_l^*+1-m}=d_l^*+1-m$, and where

$$(3.9)\qquad \Sigma^*=\sum_{r=1,\, r\neq l}^{k^*} (d_r^*+1)-(d_l^*+1).$$

Consequently, we have proved the following

THEOREM 3.1. *The linear combination of k selected order statistics L_n defined in* (1.1), *or the linear combination of k jointly Dirichlet distributed random variables defined in* (1.5), *is equidistributed to the distribution of T_n defined in* (3.1). *The exact pdf. of those statistics is a mixture of scaled beta distributions given by* (3.7) *with* (3.8) *and* (3.9), *where k^* denotes the number of non-zero and distinct b_j's in* (1.5) *or* (3.1).

In case of $k^*=n$ and therefore $d_l^*=0$ $(l=1,\cdots,k^*)$, we have the following

COROLLARY 3.1. *The linear combination of whole order statistics based on random sample of size n from the uniform distribution* $U(0,1)$ *has the pdf.*

$$f(t)=\sum_{l=1}^{n} b_l^{n-2} \prod_{\substack{r=1 \\ r\neq l}}^{n} (b_l-b_r)^{-1}\chi_l(t/b_l)\chi_l(1-t/b_l)n(1-t/b_l)^{n-1} . \tag{3.10}$$

Remark. A special case of (3.10) just coincides with the pdf. obtained by differentiating the cdf. given by Dempster and Kleyle [3].

In the remaining part of this section we shall give an exact distribution of linear combination of independent gamma variables with integer parameters. That is to say, let us consider the exact pdf. of the statistic

$$W=\sum_{j=1}^{k^*} b_j^* Y_j , \tag{3.11}$$

where, as before, Y_j's $(j=1,\cdots,k)$ are mutually independent gamma variables with pdf. (2.1). The desired result is immediately obtained by integrating out $f_{W,S}(w,s)$ in (3.5) with respect to s:

$$f_W(w)=\sum_{l=1}^{k^*} \sum_{m=1}^{d_l^*+1} C_{l,m}^{\dagger}\chi_l(w/b_l^*)w^{m-1}\exp(-w/b_l^*)/\Gamma(m) , \tag{3.12}$$

where k^*, $C_{l,m}^{\dagger}$ and $\chi_l(\cdot)$ are the same ones as that of those in the preceding theorem.

THEOREM 3.2. *The exact pdf. of the linear combination W of mutually independent gamma variables with integer parameters is a mixture of scaled gamma distributions given by* (3.12).

Remark. If $k^*=n$ and hence $d_l^*=0$ $(l=1,\cdots,k^*)$, the random variable W is distributed according to the generalized Erlang distribution, whose exact pdf. follows immediately from (3.12).

4. Asymptotic normality of L_n

From the definition of L_n its mean and variance are easily calculated as

$$\mu_n \equiv \mathrm{E}(L_n) = \frac{1}{n+1} \sum_{l=1}^{k} a_l n_l = \frac{1}{n+1} \sum_{j=1}^{k} b_j (d_j+1) \tag{4.1}$$

and

$$\begin{aligned} \sigma_n^2 \equiv \mathrm{Var}(L_n) &= \frac{1}{n+2} \sum_{l,m=1}^{k} a_l a_m \left\{ \min\left(\frac{n_l}{n+1}, \frac{n_m}{n+1} \right) - \frac{n_l n_m}{(n+1)^2} \right\} \\ &= \frac{1}{n+2} \sum_{j,r=1}^{k} b_j b_r \left(\frac{d_j+1}{n+1} \right) \left(\delta_{j,r} - \frac{d_r+1}{n+1} \right) \end{aligned} \tag{4.2}$$

$(\delta_{j,r}$: Kronecker's delta, $\sigma_n > 0)$.

However, as was seen in the preceding section, the exact pdf. of L_n is fairly complicated. Especially, in case of large k^*, it seems that calculation of exact numbers of the partitions appeared in Σ^* of (3.7) will become very laborious. So, let us consider to approximate the distribution of L_n by a normal distribution when k may increase as $n \to \infty$.

To this end we utilize again the property that L_n is equidistributed to the statistic:

$$T_n = \sum_{j=1}^{k} b_j Y_j / S \qquad \text{with} \qquad S = \sum_{j=1}^{k+1} Y_j \, ,$$

where Y_j's $(j=1, \cdots, k+1)$ are mutually independent gamma distributed random variables with mean d_j+1, respectively. Since, $\mathrm{E}(S) = \mathrm{Var}(S) = n+1$ and S can be decomposed into $n+1$ independent negative exponential random variables, then by the strong law of large numbers it is seen that

$$\frac{S}{n+1} \to 1 \qquad \text{w.p.1.} \ (n \to \infty) \, . \tag{4.3}$$

So, let us express T_n as

$$T_n = T_n^* \Big/ \left(\frac{S}{n+1} \right) \tag{4.4}$$

where

$$T_n^* = \frac{1}{n+1} \sum_{j=1}^{k} b_j Y_j \, . \tag{4.5}$$

From (4.3) and (4.4) it is expected that T_n and T_n^* are asymptotically equivalent as $n \to \infty$. To prove this we shall consider first the normal approximation to T_n^*. Let

$$\mu_n^* = \mathrm{E}(T_n^*) = \frac{1}{n+1} \sum_{j=1}^{k} b_j(d_j+1) = \mu_n , \tag{4.6}$$

$$\sigma_n^{*2} = \mathrm{Var}(T_n^*) = \frac{1}{(n+1)^2} \sum_{j=1}^{k} b_j^2(d_j+1) , \qquad (\sigma_n^* > 0) , \tag{4.7}$$

and

$$Z_n^* = (T_n^* - \mu_n^*)/\sigma_n^* . \tag{4.8}$$

In the present case we can state the following

THEOREM 4.1. (Berry-Esseen).

$$\sup_{-\infty<z<\infty} |\mathrm{P}(Z_n^* < z) - \Phi(z)| \leqq C(B_{3n}/\sigma_n^*)^3 , \tag{4.9}$$

where $\Phi(z)$ denotes the cdf. of the standard normal distribution,

$$B_{3n}^3 = (n+1)^{-3} \sum_{j=1}^{k} |b_j|^3 \mathrm{E}|Y_j - (d_j+1)|^3 \tag{4.10}$$

and C is an absolute constant satisfying $(3+\sqrt{10})/(6\sqrt{2\pi}) \leqq C < 0.7975$ (*cf. van Beek* [11]).

The following is an immediate consequence of the theorem.

COROLLARY 4.1.

$$\sup_{-\infty<t<\infty} |\mathrm{P}(T_n^* < t) - \Phi_{\mu_n^*, \sigma_n^*}(t)| \leqq C(B_{3n}/\sigma_n^*)^3 , \tag{4.11}$$

where $\Phi_{\mu_n^*, \sigma_n^*}(\cdot)$ denotes the cdf. of the normal distribution $N(\mu_n^*, \sigma_n^{*2})$.

For the variable S, we have the following

COROLLARY 4.2.

$$\sup_{-\infty<z<\infty} |\mathrm{P}(S-(n+1) < \sqrt{n+1}\,z) - \Phi(z)| < C(A_{3n}/\sqrt{n+1})^3 \tag{4.12}$$

where

$$A_{3n}^3 = \sum_{j=1}^{k+1} \mathrm{E}|Y_j - (d_j+1)|^3 . \tag{4.13}$$

Now, we show a uniform asymptotic equivalence between T_n and T_n^*. For any given real t and for any given sequence of positive numbers $\{\varepsilon_n\}$ such that $\varepsilon_n \to 0$ and $n\varepsilon_n^2 \to \infty$, as $n \to \infty$, it holds that

$$|\mathrm{P}(T_n < t) - \mathrm{P}(T_n^* < t)| < \sup_{|x-x'|<2|t|\varepsilon_n} \mathrm{P}(x \leqq T_n^* < x') \tag{4.14}$$

$$+\mathrm{P}\left(\left|\frac{S}{n+1}-1\right|\geqq\varepsilon_n\right)$$
$$=\sup_{|z-z'|<2\varepsilon_n|t-\mu_n^*|/\sigma_n^*}\mathrm{P}(z\leqq Z_n^*<z')+\mathrm{P}(|S-(n+1)|\geqq(n+1)\varepsilon_n)\ .$$

In view of Theorem 4.1, there exists a sequence of positive numbers $\{K_n\}$ such that $K_n\to\infty$ as $n\to\infty$, and such that

$$\mathrm{P}(|T_n^*-\mu_n^*|\geqq K_n\sigma_n^*)\leqq 2[1-\Phi(K_n)+C(B_{3n}/\sigma_n^*)^3]\ . \tag{4.15}$$

Further, for the above ε_n and K_n there exists a sequence of positive numbers $\{\delta_n\}$ such that $\delta_n=2\varepsilon_nK_n\to 0$ as $n\to\infty$, which will be seen in (4.19) later. Then,

$$\sup_{|z-z'|<2\varepsilon|t-\mu_n^*|/\sigma_n^*}\mathrm{P}(z\leqq Z_n^*<z') \tag{4.16}$$
$$\leqq\sup_{|z-z'|<\delta_n}|\Phi(z)-\Phi(z')|+2C(B_{3n}/\sigma_n^*)^3+\mathrm{P}(|T_n^*-\mu_n^*|\geqq K_n\sigma_n^*)\ .$$

Moreover, by Bernstein's inequality

$$\mathrm{P}(|S-(n+1)|>(n+1)\varepsilon_n)\leqq\mathrm{P}(|S-(n+1)| \tag{4.17}$$
$$\geqq 2(\sqrt{n+1}\,\varepsilon_n/2)\sqrt{\mathrm{Var}(S)})\leqq 2\exp\{-(n+1)\varepsilon_n^2/4\}\ .$$

It is easily seen from (4.14)–(4.17) that

$$\sup_{-\infty<t<\infty}|\mathrm{P}(T_n<t)-\mathrm{P}(T_n^*<t)| \tag{4.18}$$
$$\leqq 4C(B_{3n}/\sigma_n^*)^3+\sqrt{\frac{2}{\pi}}\frac{1}{K_n}e^{-(K_n)^2/2}+\frac{1}{\sqrt{2\pi}}e^{-1/2}\delta_n$$
$$+2\exp\{-(n+1)\varepsilon_n^2/4\}\qquad(\equiv\eta_n,\ \text{say})\ .$$

Thus, if we take $2\{(\ln\sqrt{n+1})/(n+1)\}^{1/2}$ as ε_n in (4.14) and choosing the sequences $\{K_n\}$ and $\{\delta_n\}$ properly such that $K_n\to\infty$ and $\delta_n=2\varepsilon_nK_n\to 0$, as $n\to\infty$, for instance, such as

$$K_n=(n+1)^{(1-\alpha)/2}(\ln\sqrt{n+1})^{1/2}\quad\text{and}\quad\delta_n=4(n+1)^{-\alpha/2}\ln\sqrt{n+1} \tag{4.19}$$

with some positive constant α in $(0,1]$, then from (4.18) we get the following result:

THEOREM 4.2. *If the condition*

$$(B_{3n}/\sigma_n^*)^3\to 0\ ,\qquad(n\to\infty) \tag{4.20}$$

is satisfied, then T_n and T_n^* are uniformly asymptotically equivalent in the sense of

$$\sup_{-\infty<t<\infty}|\mathrm{P}(T_n<t)-\mathrm{P}(T_n^*<t)|\to 0\ ,\qquad(n\to\infty)\ . \tag{4.21}$$

Next, we shall prove

COROLLARY 4.3. *Under the condition* (4.20) *it holds that*

$$\left(\frac{\sigma_n}{\sigma_n^*}\right)^2 \to 1\,, \qquad (n\to\infty)\,. \tag{4.22}$$

PROOF. For any constant $\tau>0$

$$\begin{aligned}\left|\left(\frac{\sigma_n}{\sigma_n^*}\right)^2-1\right| &\leqq \frac{1}{\sigma_n^{*2}}\int_{|t-\mu_n^*|<\tau\sigma_n^*}(t-\mu_n^*)^2|d\{\mathrm{P}(T_n<t)-\mathrm{P}(T_n^*<t)\}| \\ &\quad+\frac{1}{\sigma_n^{*2}}\int_{|t-\mu_n^*|\geqq\tau\sigma_n^*}(t-\mu_n^*)^2|d\{\mathrm{P}(T_n<t)-\mathrm{P}(T_n^*<t)\}| \\ &\leqq \tau^2\cdot\sup_{-\infty<t<\infty}|\mathrm{P}(T_n<t)-\mathrm{P}(T_n^*<t)| \\ &\quad+\frac{1}{\sigma_n^{*2}}\int_{|t-\mu_n^*|\geqq\tau\sigma_n^*}(t-\mu_n^*)^2 d\,\mathrm{P}(T_n^*<t) \\ &\leqq \tau^2\eta_n+\frac{1}{\sigma_n^{*2}}\times \\ &\quad\times\sum_{j=1}^{k}\int_{|b_j|/(n+1)\cdot|y_j-(d_j+1)|\geqq\tau\sigma_n^*}\left|\frac{b_j}{n+1}\{y_j-(d_j+1)\}\right|^2 d\,\mathrm{P}(Y_j<y) \\ &\leqq \tau^2\eta_n+(B_{3n}/\sigma_n^*)^3/\tau\,,\end{aligned}\tag{4.23}$$

where η_n is the same quantity defined in (4.18). Hence, the condition (4.20) assures us the validity of (4.22).

Furthermore, we have the following

COROLLARY 4.4. *Under the condition* (4.20)

$$\sup_{-\infty<t<\infty}|\Phi_{\mu_n,\sigma_n}(t)-\Phi_{\mu_n^*,\sigma_n^*}(t)|\to 0\,, \qquad (n\to\infty)\,. \tag{4.24}$$

PROOF. Since $\mu_n=\mu_n^*$ the Kullback-Leibler mean information between the two distributions is calculated as

$$\begin{aligned}I(\Phi_{\mu_n,\sigma_n}:\Phi_{\mu_n^*,\sigma_n^*}) &= I(\Phi_{0,\sigma_n}:\Phi_{0,\sigma_n^*}) \\ &=\ln(\sigma_n/\sigma_n^*)+(\sqrt{2\pi}\sigma_n^*)^{-1}\int_{-\infty}^{\infty}\exp\{-t^2/(2\sigma_n^{*2})\} \\ &\quad\times(1/\sigma_n^2-1/\sigma_n^{*2})t^2/2dt \\ &=\ln(\sigma_n/\sigma_n^*)+(\sigma_n^{*2}/\sigma_n^2-1)/2\leqq(\sigma_n^2-\sigma_n^{*2})^2/(2\sigma_n^2\sigma_n^{*2})\,.\end{aligned}$$

Then, using the inequality given in Matsunawa [7], it follows that

$$\begin{aligned}&\sup_{-\infty<t<\infty}|\Phi_{\mu_n,\sigma_n}(t)-\Phi_{\mu_n^*,\sigma_n^*}(t)| \\ &\quad\leqq\{I(\Phi_{\mu_n,\sigma_n}:\Phi_{\mu_n^*,\sigma_n^*})/2\}^{1/2} \\ &\quad\leqq|\sigma_n^2-\sigma_n^{*2}|/(2\sigma_n\sigma_n^*)=0.5|(\sigma_n/\sigma_n^*)^2-1|(\sigma_n^*/\sigma_n)\,.\end{aligned}\tag{4.25}$$

Hence, by Corollary 4.3, we get the desired result (4.24).

Now, we are in a position to state the following

THEOREM 4.3. *Under the conditions* (4.20)

$$\sup_{-\infty<t<\infty} |\mathrm{P}(T_n<t)-\Phi_{\mu_n,\sigma_n}(t)|\to 0\ , \qquad (n\to\infty)\ . \tag{4.26}$$

PROOF. By the inequalities (4.18), (4.11) and (4.25), we get the following estimation:

$$\begin{aligned}(4.27)\quad &\sup_{-\infty<t<\infty} |\mathrm{P}(T_n<t)-\Phi_{\mu_n,\sigma_n}(t)| \\ &< \sup_{-\infty<t<\infty} \{|\mathrm{P}(T_n<t)-\mathrm{P}(T_n^*<t)|+|\mathrm{P}(T_n^*<t)-\Phi_{\mu_n^*,\sigma_n^*}(t)| \\ &\qquad +|\Phi_{\mu_n^*,\sigma_n^*}(t)-\Phi_{\mu_n,\sigma_n}(t)|\} \\ &\leqq \eta_n+C(B_{3n}/\sigma_n^*)^3+0.5|(\sigma_n/\sigma_n^*)^2-1|\cdot(\sigma_n^*/\sigma_n)\ ,\end{aligned}$$

from which (4.26) immediately follows.

In the above theorem we have seen that the condition (4.20) is sufficient for T_n to be asymptotically normally distributed to $N(\mu_n, \sigma_n^2)$ in the sense of (4.26). The sufficient condition is nothing but a special case of Lyapunov's condition in the central limiting theorem. In practical point of view, however, the condition is not so manageable, because the third absolute moments appeared in B_{3n}^3 are fairly difficult in calculation. So, it is interesting to provide other sufficient conditions for T_n and T_n^*, if exist.

Indeed, it is possible to give such conditions. In addition, we can prove those conditions are necessary for T_n and T_n^* to be asymptotically normally distributed according to $N(\mu_n^*, \sigma_n^{*2})$ in the same sense as (4.26), if we put certain restrictions on d_j's later.

Let us consider the condition

$$\max_{1\leqq j\leqq k} |b_j|\sqrt{d_j+1}/\{(n+1)\sigma_n^*\}\to 0\ , \qquad (n\to\infty)\ , \tag{4.28}$$

under which the condition (4.20) automatically holds, because

$$\begin{aligned}\mathrm{E}|Y_j-(d_j+1)|^3 &\leqq \{\mathrm{E}[Y_j-(d_j+1)]^4\}^{3/4}=\{3(d_j+1)^2+6(d_j+1)\}^{3/4} \\ &\leqq 3\sqrt{3}\,(d_j+1)^{3/2}\end{aligned}$$

for each j, and hence

$$\begin{aligned}(4.29)\quad 0<\left(\frac{B_{3n}}{\sigma_n^*}\right)^3 &\leqq 3\sqrt{3}\sum_{j=1}^{k}|b_j|^3(d_j+1)^{3/2}/\{(n+1)\sigma_n^*\}^3 \\ &\leqq 3\sqrt{3}\max_{1\leqq j\leqq k}|b_j|\sqrt{d_j+1}/\{(n+1)\sigma_n^*\}\ .\end{aligned}$$

Thus, by Theorem 4.3 the condition (4.28) is sufficient for the asymptotic normality of T_n.

Conversely, suppose that T_n is asymptotically normally distributed according to $N(\mu_n^*, \sigma_n^{*2})$ in the same sense as (4.26). Then, by (4.3) and (4.4) T_n^* is also asymptotically distributed to the same distribution, and we have

$$\ln \varphi_n(t) \to -t^2/2\,, \qquad (n\to\infty)\,, \tag{4.30}$$

where $\varphi_n(t)$ denotes the characteristic function of $Z_n^*=(T_n^*-\mu_n^*)/\sigma_n^*$ and is given by

$$\varphi_n(t)=\prod_{j=1}^{k}\left[\exp\left\{-\frac{itb_j(d_j+1)}{(n+1)\sigma_n^*}\right\}\cdot\left(1-\frac{itb_j}{(n+1)\sigma_n^*}\right)^{-(d_j+1)}\right]. \tag{4.31}$$

We have from (4.30) with (4.31)

$$\sum_{j=1}^{k}[itb_j(d_j+1)/\{(n+1)\sigma_n^*\}+(d_j+1)\ln(1-itb_j/\{(n+1)\sigma_n^*\})]\to t^2/2\,, \qquad (n\to\infty)\,.$$

Taking the real part of the above

$$\frac{1}{2}\sum_{j=1}^{k}(d_j+1)\ln[1+b_j^2t^2/\{(n+1)\sigma_n^*\}^2]\to t^2/2\,, \qquad (n\to\infty)\,,$$

and applying Taylor's expansion we have

$$\frac{1}{2}\sum_{j=1}^{k}\left[\frac{b_j^2(d_j+1)}{(n+1)^2\sigma_n^{*2}}t^2-\frac{b_j^4t^4/\{(n+1)\sigma_n^*\}^4\cdot(d_j+1)}{2[1+\theta_jb_j^2t^2/\{(n+1)\sigma_n^*\}^2]^2}\right]\to\frac{t^2}{2}\,, \qquad (n\to\infty)\,,$$

where θ_j is some constant in (0, 1). Thus, for sufficiently small t there exists a positive constant M such that

$$\sum_{j=1}^{k}\frac{b_j^4t^4(d_j+1)}{\{(n+1)\sigma_n^*\}^4+\theta_jb_j^2t^2}>Mt^4\sum_{j=1}^{k}\left\{\frac{b_j^2\sqrt{d_j+1}}{(n+1)^2\sigma_n^{*2}}\right\}^2\to 0\,, \qquad (n\to\infty)$$

for any combinations of $k=k(n)$ and $d_j=d_j(n)$ (cf. Chernoff, Gastwirth and Johns [2], Eicker and Puri [5]). This implies that

$$\max_{1\le j\le k} b_j^2\sqrt{d_j+1}/\{(n+1)^2\sigma_n^{*2}\}\to 0\,, \qquad (n\to\infty)\,, \tag{4.33}$$

which is weaker than (4.28). But, if d_j ($1\leqq j\leqq k$) are bounded from above, (4.28) is a necessary condition for the asymptotic normality of T_n^* and thus T_n.

Moreover, it should be noted that because of (4.29), the Lyapunov condition (4.20) is also a necessary condition for T_n and T_n^* to be asymptotically normally distributed to $N(\mu_n^*, \sigma_n^{*2})$ in the same sense as before. From this fact and the Corollaries 4.3 and 4.4 the condition (4.22) becomes the necessary condition, too. Further, owing to the general theory of central limiting theorems, the Lindeberg condition becomes other necessary and sufficient condition for our problem, although it is not so manageable, too.

Consequently, we can summarize above discussion as follows:

THEOREM 4.4. *If d_j's are bounded, the following are equivalent*:

(i) $\max_{1\leq j\leq k}|b_j|/\{(n+1)\sigma_n^*\}\to 0,\ (n\to\infty)$.

(ii) $(B_{3n}/\sigma_n^*)^3\to 0,\ (n\to\infty)$.

(iii) *For any* $\tau>0$

$$\sum_{j=1}^{k}\int_{|b_j|/(n+1)\cdot|y_j-(d_j+1)|\geqq\tau\sigma_n^*} b_j^2\left\{\frac{y_j-(d_j+1)}{(n+1)\sigma_n^*}\right\}^2 d\,\mathrm{P}(Y_j<y)\to 0\ ,\qquad (n\to\infty)\,.$$

(iv) $\sup_{-\infty<t<\infty}|\mathrm{P}(T_n^*<t)-\Phi_{\mu_n^*,\sigma_n^*}(t)|\to 0,\ (n\to\infty)$.

(v) $\sup_{-\infty<t<\infty}|\mathrm{P}(T_n<t)-\Phi_{\mu_n^*,\sigma_n^*}(t)|\to 0,\ (n\to\infty)$.

Remarks. The implication relation (i)⇒(4.22) can be also derived directly as

$$\begin{aligned}|(\sigma_n/\sigma_n^*)^2-1|&=\left|(n+2)^{-1}-\left\{\sum_{j=1}^{k}b_j(d_j+1)/(n+1)\right\}^2\Big/\{(n+2)\sigma_n^{*2}\}\right|\\&<(n+2)^{-1}+\max_{1\leq j\leq k}\{b_j^2(d_j+1)\}/\{(n+1)\sigma_n^*\}^2\to 0\ ,\qquad(n\to\infty)\,.\end{aligned}$$

In view of the above condition we are permitted to replace σ_n^* with σ_n and vice versa in the all statements of the Theorem 4.4. In case of $k=n$ and hence $d_j=0$ for all j, namely the case of using whole order statistics, (i) coincides with the condition by Hecker [6] where the limiting normality of $Z_n=(T_n-\mu_n)/\sigma_n$ is treated. A more accurate approximation by the Edgeworth expansion for Z_n is possible according to van Zwet [12]. Some extended results to the linear combinations of selected order statistics from certain general continuous distributions are obtainable based on our theorems and the probability integral transformations, which are designed for discussing elsewhere.

Acknowledgement

The author is grateful to the referees for their detailed comments and useful suggestions.

THE INSTITUTE OF STATISTICAL MATHEMATICS

REFERENCES

[1] Andrews, G. E. (1976). *The Theory of Partitions, Encyclopedia of Mathematics and Its Applications* 2, Addison-Wesley.

[2] Chernoff, H., Gastwirth, J. L. and Johns, M. V. (1967). Asymptotic distribution of linear combinations of functions of order statistics with applications to estimations, *Ann. Math. Statist.*, **38**, 52-72.

[3] Dempster, A. P. and Kleyle, R. M. (1968). Distributions determined by cutting a simplex with hyperplanes, *Ann. Math. Statist.*, **39**, 1473-1478.

[4] Doetsch, G. (1961). *Guide to the Applications of Laplace Transforms*, D. Van Nostrand.

[5] Eicker, F. and Puri, M. L. (1976). On linear combination of order statistics, *Essays in Probability and Statistics*, Shinko Tsusho Co., Ltd. (ed. S. Ikeda and others), 433-449.

[6] Hecker, H. (1976). A characterization of the asymptotic normality of linear combinations of order statistics from the uniform distribution, *Ann. Statist.*, **4**, 1244-1246.

[7] Matsunawa, T. (1982). Some strong ε-equivalence of random variables, *Ann. Inst. Statist. Math.*, **34**, A, 209-224.

[8] Pyke, R. (1965). Spacings, *J. Roy. Statist. Soc.*, B, **27**, 395-436.

[9] Shorack, G. (1969). Asymptotic normality of linear combinations of functions of order statistics, *Ann. Math. Statist.*, **40**, 2041-2050.

[10] Stigler, S. M. (1974). Linear functions of order statistics with smooth weight functions, *Ann, Statist.*, **4**, 676-693.

[11] van Beek, P. (1972). An application of Fourier Methods to the problem of sharpning the Berry-Esseen inequality, *Zeit. Wahrsheinlichkeitsth.*, **23**, 187-196.

[12] van Zwet, W. R. (1979). The Edgeworth expansion for linear combinations of uniform order statistics, *Proceedings of the Second Prague Symposium on Asymptotic Statistics* (ed. Mandl, P. and Hušková), 93-101, North-Holland.

[13] Weisberg, H. (1971). The distribution of linear combinations of order statistics from the uniform distribution, *Ann. Math. Statist.*, **42**, 704-709.