〈Research Paper〉

Journal of the Korean Statistical Society
Vol. 13, No. 1, 1984

Probability Integral of the Inverted Dirichlet Distribution with Application

Keeyoung Kim*

ABSTRACT

A technique which has been used for the evaluation of certain kinds of multiple integrals, viz., the technique of incomplete gamma function operators, is employed and extended to the case where the parameters and arguments are non-equal and non-integer for the probability integral of the inverted Dirichlet distribution. Several types of recurrence formulas have been developed for the tail probabilities and a subset selection procedure in ranking variances is discussed as an application.

1. Introduction

Let $D_p'(a_1, a_2, \ldots, a_p; \nu_1, \nu_2, \ldots, \nu_p; \nu_0)$ denote the cumulative distribution function (cdf) of the inverted Dirichlet distribution with parameters $(\nu_1, \nu_2, \ldots, \nu_p; \nu_0)$ evaluated at the upper bound of $(a_1, a_2, \ldots, a_p)$. That is,

$$D_p'(a_1, a_2, \ldots, a_p; \nu_1, \nu_2, \ldots, \nu_p; \nu_0)$$

$$=\frac{\Gamma(\nu)}{\prod_{i=0}^{p} \Gamma(\nu_i)} \int\int_{\mathscr{A}}\cdots\int \prod_{i=1}^{p} x_i^{\nu_i-1}\Big(1+\sum_{1}^{p} x_j\Big)^{-\nu} \prod_{1}^{p} dx_i, \tag{1.1}$$

where $\mathscr{A}=\{0\leqslant x_i\leqslant a_i,\ i=1, 2, \ldots, p\}$ and $\nu=\sum_{0}^{p} \nu_i$.

Several authors have considered this type of probability integral. Among them, Tiao and Guttman (1965) obtained a recursive relation (with only integer parameters) for the inverted Dirichlet integral by a direct integration technique. Krishnaiah(1965) used Gauss-Hermite quadrature to evaluate the multivariate F probability integral and suggested some approximations due to Poincaré. Gupta and Sobel (1962a) obtained a formula for the cdf of the minimum of a set of random variables having a multivariate F distribution. In 1941, Finney investigated a method using differential operator to solve the multivariate

* Department of Statistics, Korea University, Seoul 132, Korea

F probability integral for equal and integer exponents. In this paper the incomplete gamma function operators have been employed to evaluate the incomplete inverted Dirichlet distribution. Several types of recurrence formulas are obtained and the results are applied to the subset selection procedure in ranking variances.

2. Main Lemmas

As preliminary results we consider some properties of the incomplete gamma function operator. Let $G(x,m)$ be the incomplete gamma function for any positive m, that is,

$$G(x,\ m)=\frac{1}{\Gamma(m)}\int_0^x u^{m-1}e^{-u}\,du.$$

Then $G(x,m)=e^{-x}\sum_0^\infty \frac{x^{j+m}}{\Gamma(j+m+1)}$; if we replace x by $-aD$, where D is the differential operator, we have, formally,

$$G(-aD,\ m)=e^{aD}\sum_0^\infty \frac{(-aD)^{j+m}}{\Gamma(j+m+1)}\ .$$

This operator is called $M_{(a,m)}$ and its complemental operator, $1-M_{(a,m)}$ will be denoted by $L_{(a,m)}$. In this notation one should understand "1" as an identity operator. For an integer value of m, the complement, $1-G(x,m)$ can be expressed as a finite series. Thus, for integer m,

$$L_{(a,m)}=e^{aD}\sum_0^{m-1}\frac{(-aD)^j}{\Gamma(j+1)}$$

It can be easily verified that the t'th derivative of $(c+\lambda)^{-\nu}$ is

$$D^t(c+\lambda)^{-\nu}=(-1)^t\frac{\Gamma(\nu+t)}{\Gamma(\nu)}(c+\lambda)^{-(\nu+t)}$$

for all positive ν, c being a quantity independent of λ, and $D=\partial/\partial\lambda$.

The Taylor's shift operation states that

$$e^{\alpha D}f(\lambda)=\sum_0^\infty\frac{(\alpha D)^j}{\Gamma(j+1)}f(\lambda)=f(\lambda+\alpha).$$

Using the above properties of the incomplete gamma function operators and integration by parts successively, the following lemmas are obtained.

Lemma 1.

For any positive a_i, ν_i for $i=1, 2, \ldots, p$, and ν_0

$$\frac{\Gamma(\nu)}{\prod_0^p\Gamma(\nu_i)_{\mathscr{A}}}\iint\cdots\int\prod_1^p x_i^{\nu_i-1}\Big(\mu+\lambda+\sum_1^p x_i\Big)^{-\nu}\prod_1^p dx_i$$

$$=\prod_1^p M_{(a_i,\nu_i)}\{(\mu+\lambda)^{-\nu_0}\},$$

where μ is a constant independent of λ and $\mathscr{A}=\{0\leq x_i\leq a_i,\ i=1,2,\dots,p\}$.

By taking $\mu=1$ and $\lambda=0$ in Lemma 1, we have the following relation.

$$D_p'(a_1, a_2, \dots, a_p;\ \nu_1, \nu_2, \dots, \nu_p\ ;\ \nu_0)=\prod_{i=1}^{p} M_{(a_i,\nu_i)}(1+\lambda)^{-\nu_0}|_{\lambda=0}.$$

As a particular case, we have

$$D_p'(a, a, \dots, a;\ \nu, \nu, \dots, \nu;\ \nu_0)=M^{p}_{(a,\nu)}(1+\lambda)^{-\nu_0}|_{\lambda=0}.$$

The following relation on the M operator is useful to obtain a recurrence formula for several different types of parameters in the inverted Dirichlet integral.

Lemma 2.

Let a_0, ν_0, a_x, ν_x, for $x=r, r+1, \dots, p$, be positive numbers and r, p be positive integers, $p\leqq r\leqq 1$.

$$(1+a_0)^{\nu_0}\prod_{x=r}^{p} M_{(a_x,\nu_x)}(1+a_0+\lambda)^{-\nu_0}|_{\lambda=0}=\prod_{x=r}^{p} M_{(a_x/(1+a_0),\nu_x)}(1+\lambda)^{-\nu_0}|_{\lambda=0}.$$

It can be verified that the same relation holds true for the L operator in Lemma 2.

The exact evaluation of the inverted Dirichlet integral is complicate in general, especially when p is large. As will be seen later, the recurrence formula in the p-dimensional integral involves finite (for integer ν_i's) or infinite (for non-integer ν_i's) summation of the $(p-1)$-dimensional inverted Dirichlet integral. Thus it may be computationally useful to obtain a recurrence formula in which each term in the summation is expressed in terms of a previous term plus a finite number of correction terms.

Lemma 3.

For $p\geqq 2$

$$\begin{aligned}&D'_{p-1}(b_2, b_3, \dots, b_p;\ \nu_2/2, \nu_3/2, \dots, \nu_p/2;\ \nu_0+j+1)\\&\quad=D'_{p-1}(b_2, b_3, \dots, b_p;\ \nu_2/2, \nu_3/2, \dots, \nu_p/2;\ \nu_0+j)\\&\qquad+\frac{1}{\Gamma(\nu_0+j+1)}\sum_{k=2}^{p}\frac{\Gamma(\nu_k/2+\nu_0+j)}{\Gamma(\nu_k/2)(1+b_k)^{\nu_0+j}}(b_k/(1+b_k))^{\nu_k/2}\cdot\\&\qquad D'_{p-2}\left(\frac{b_i}{1+b_k};\ \nu_i/2,\ i=2,3,\dots,p,\ i\neq k;\ \nu_0+j+\nu_k/2\right),\end{aligned}$$

where $D_p'(.)=1$.

By analogy it may be easily proved that, for the upper tail probability integral in the inverted Dirichlet distribution, the same relationship holds true if we replace the sign of the second term in Lemma 3 by negative.

3. Probability Integral of the Inverted Dirichlet Distribution

The probability integral with integer parameters involves only finite series. Thus for an integer $\nu_1/2$, say, we have, using the Lemma 1, 2, and 3,

$$\prod_{x=1}^{p} M_{(a_x,\nu_x/2)}(1+\lambda)^{-\nu_0}|_{\lambda=0}=(1-L_{(a_1,\nu_1/2)})\prod_{2}^{p} M_{(a_x,\nu_x/2)}(1+\lambda)^{-\nu_0}|_{\lambda=0}$$

$$=\prod_{2}^{p} M_{(a_x,\nu_x/2)}(1+\lambda)^{-\nu_0}|_{\lambda=0}-\sum_{j_1=0}^{\nu_1/2-1}\frac{e^{a_1 D}(-a_1 D)^{j_1}}{\Gamma(j_1+1)}\prod_{x=2}^{p} M_{(a_x,\nu_x/2)}(1+\lambda)^{-\nu_0}|_{\lambda=0}$$

$$=\prod_{x=2}^{p} M_{(a_x,\nu_x/2)}(1+\lambda)^{-\nu_0}|_{\lambda=0}-\sum_{j_1=0}^{\nu_1/2-1}\frac{\Gamma(j_1+\nu_0)a_1^{j_1}}{\Gamma(\nu_0)\Gamma(j_1+1)(1+a_1)}{}^{\nu_0+j_1}\cdot$$

$$\prod_{2}^{p} M_{(\frac{a_x}{1+a_1},\ \frac{\nu_x}{2})}\{(1+\lambda)^{-(\nu_0+j_1)}\}|_{\lambda=0},$$

where the last identity follows from Lemma 2.

Thus for an integer $\nu_1/2$, the recurrence formula is given by

$$D_p'(a_1, a_2, \ldots, a_p;\ \nu_1/2, \nu_2/2, \ldots, \nu_p/2;\ \nu_0)=D'_{p-1}(a_2, \ldots, a_p;\ \nu_2/2, \ldots, \nu_p/2; \nu_0)$$

$$-\frac{1}{\Gamma(\nu_0)(1+a_1)^{\nu_0}}\sum_{j_1=0}^{\nu_1/2-1}\frac{\Gamma(j_1+\nu_0)}{\Gamma(j_1+1)}\left(\frac{a_1}{1+a_1}\right)^{j_1} D'_{p-1}(a_2/(1+a_1), \ldots, a_p/(1+$$

$$a_1)\ ;\nu_2/2, \ldots, \nu_p/2;\ \nu_0+j_1),$$

where $D_0'(.)=1$.

The situation of non-integer parameters$(\nu_1/2, \nu_2/2, \ldots, \nu_p/2)$ has several useful applications; for instance the p-dimensional multivariate T distribution is related to the inverted Dirichlet with all parameters equal to $1/2$.

When we apply the algorithm of the incomplete gamma function operators and the Lemmas for the non-integer case, the corresponding recurrence formula is expressed in the form of non-terminating series.

For a real number, $\nu_1/2$, we have

$$D_p'(a_1, a_2, \ldots, a_p;\ \nu_1/2, \nu_2/2, \ldots, \nu_p/2;\ \nu_0)=M_{(a_1,\nu_1/2)}\prod_{x=2}^{p} M_{(a_x,\nu_x/2)}(1+\lambda)^{-\nu_0}|_{\lambda=0}$$

$$=\frac{1}{\Gamma(\nu_0)}\sum_{j_1=\nu_1/2}^{\infty}\frac{a_1^{j_1}\Gamma(j_1+\nu_0)}{\Gamma(j_1+1)}\prod_{x=2}^{p} M_{(a_x,\nu_x/2)}(1+a_1+\lambda)^{-(\nu_0+j_1)}|_{\lambda=0}$$

$$=\frac{1}{\Gamma(\nu_0)(1+a_1)^{\nu_0}}\sum_{j_1=\nu_1/2}^{\infty}\frac{\Gamma(j_1+\nu_0)}{\Gamma(j_1+1)}(a_1/(1+a_1))^{j_1}\cdot D'_{p-1}(a_2/(1+a_1), \ldots,$$

$$a_p/(1+a_1); \nu_2/2, \ldots, \nu_p/2;\ \nu_0+j_1),$$

where $D_0'(.)=1$.

Let $(i_1, i_2, \ldots, i_p)$ be any possible permutation of indices$(1, 2, \ldots, p)$. Then we can rearrange a_i's to match with the corresponding ν_i's for the probability integral in the inverted Dirichlet distribution by interchanging integration. That is,

$$D_p'(a_1, a_2, \ldots, a_p;\ \nu_1/2, \nu_2/2, \ldots, \nu_p/2;\ \nu_0)$$

$$=D_p'(a_{i1}, a_{i2}, \ldots, a_{ip};\ \nu_{i1}/2, \nu_{i2}/2, \ldots, \nu_{ip}/2;\ \nu_0).$$

Without loss of generality, we may say the first r of ν_i's are even integers so that ω_i's are integers, where $\omega_i=\nu_i/2$ for $i=1, 2, \ldots, r$. Then we have the following.

$$D_p'(a_1, a_2, \ldots, a_p;\ \nu_1/2, \nu_2/2, \ldots, \nu_p/2;\ \nu_0)$$
$$=D_p'(a_1, a_2, \ldots, a_p;\ \omega_1, \omega_2, \ldots, \omega_r, \nu_{r+1}/2, \ldots, \nu_p/2;\ \nu_0).$$

Therefore the inverted Dirichlet integral with mixed parameters (involving integers, half integers, and real numbers) can be handled by a combination of the formulas developed above. That is, we can apply the recurrence formula for the integer parameters on the first r terms and then use the non-integer formula for D'_{p-r} which has lower dimension $(p-r)$.

4. Application to the Subset Selection Procedure in Ranking Variances

Let $F_j=\nu_0\ X_j/\nu_j,\ j=1,2,\ldots,p$, where $(X_1, X_2, \ldots, X_p)$ are from the inverted Dirichlet distribution with parameters $(\nu_1/2, \nu_2/2, \ldots, \nu_p/2;\ \nu_0/2)$. Then random variables $(F_1, F_2, \ldots, F_p)$ follows the multivariate F distribution with parameters $(\nu_1, \nu_2, \ldots, \nu_p;\ \nu_0)$ through a simple Jacobian transformation. The distribution of the largest F may be applied to selecting a subset of populations.

Let $\pi_1, \pi_2, \pi_3, \ldots, \pi_p$ denote $p(\geqslant 2)$ independent normal populations with means $\mu_1, \mu_2, \ldots, \mu_p$ and variances $\sigma_1^2, \sigma_2^2, \ldots, \sigma_p^2$ respectively and let π_0 denote the control population having mean μ_0 and variance σ_0^2. It is assumed that measurements $(x_{1j}, x_{2j}, \ldots, x_{pj})$, $j=1, 2, \ldots, (\nu+1)$ and x_{0k}, $k=1, 2, \ldots, (\nu_0+1)$ are observed under the populations $\pi_1, \pi_2, \ldots, \pi_p$ and π_0.

Chen and Pickett(1983) proposed several procedures on selecting a subset of treatments better than a control in repeated measurement design. Since the number of better populations is unknown, a lower bound in the minimum of a probability statement was imposed to obtain a feasible but conservative solutions. To improve the conservative procedure they introduced a discrete uniform and binomial prior distributions on the number of the unknown better populations in order to obtain a better selection procedure in the sense that the expected subset size becomes smaller. In this application we consider a (unspecified) prior distribution on the number of better populations and develope a procedure to select all populations better than a control so that the probability of correct selection is maintained at least at an arbitrary fixed level. By "better than the control" we mean the population having variance smaller than or equal to that of the control. If a correct selection ($C.S.$) is defined to be the selection of any subset which contains all those populations better than the control, then the object is to propose a procedure such that $Pr(C.S.|\boldsymbol{P})$, the probability of a correct selection, is at least equal to a preassigned confidence level P^* for all true configurations of the population variances (P^*-require ment).

Let the ordered variances be denoted by $\sigma^2_{[1]} \leqslant \sigma^2_{[2]} \leqslant \ldots \leqslant \sigma^2_{[p]}$, and let S_j^2 denote the sample variance from π_j for $j=1, 2, \ldots, p$, 0 respectively, where $\nu\, S_j^2/\sigma_j^2$ is distributed as chi-square with ν d.f. for $j=1, 2, \ldots, p$, and $\nu_0\, S_0^2/\sigma_0^2$ is a chi-square with ν_0 d.f.. Let $p_1(\leqslant p)$ be the number of populations better than the control π_0 and p_2 be the one which is worse than π_0, so that $p_1+p_2=p$.

The proposed procedure is

$$\boldsymbol{P} : \text{select } \pi_i \text{ if and only if } S_i^2 \leqslant cS_0^2, \quad c \geqslant 1,$$

where c is a constant depending on P^*.

If $S^2_{(i)}$ is associated with $\sigma^2_{[i]}$ which is smaller than or equal to σ_0^2, for some i, $i=1, 2, \ldots, p$, and $P_{p_1}(m)$ denote the prior distribution of the random variate p_1, then we have

$$Pr(C.\,S.|\boldsymbol{P}) = \sum_{m=0}^{p} Pr\{S^2_{(i)} \leqslant cS_0^2,\ i=1,\ 2, \ldots,\ m\} P_{p_1}(m)$$

$$= \sum_{m=0}^{p} Pr\{F_i \leqslant c\sigma_0^2/\sigma^2_{[i]}, i=1,\ 2, \ldots,\ m\} P_{p_1}(m),$$

where $(F_1,\ F_2, \ldots,\ F_m)$ is distributed as a multivariate F distribution with parameters $(\nu,\ \nu, \ldots,\ \nu;\ \nu_0)$. This probability becomes its infimum by setting $\sigma^2_{[1]}=\sigma^2_{[2]}=\ldots=\sigma^2_{[p]}=\sigma_0^2$ Thus we obtain

$$\text{Inf } Pr(C.S.|\boldsymbol{P}) = \sum_{m=0}^{p} Pr\{\underset{i}{\text{Max}}\, F_i(i=1, 2, \ldots,\ m) \leqslant c\} \cdot P_{p_1}(m)$$
$$= P^*.$$

If we have a prior information on p_1, then we can obtain the constant c using the tail probability of the largest F statistic or equivalently the tail probability of the inverted Dirichlet distribution. In the situation where the prior distribution is not available, one may bound the infimum of $Pr(C.S.|\boldsymbol{P})$ from below. That is,

$$\text{Inf } Pr(C.S.|\boldsymbol{P}) = \sum_{m=0}^{p} Pr\{\underset{i}{\text{Max}} F_i(i=1,\ 2, \ldots, p) \leqslant c\} \cdot P_{p_1}(m)$$
$$= Pr\{\underset{i}{\text{Max}}\, F_i(i=1,\ 2, \ldots, p) \leqslant c\}.$$

We note that the upper tail probability integral of the inverted Dirichlet distribution can be evaluated by employing the L operator. Especially for integer parameters it can be expressed in terms of p-fold summation and it is expandable for non-integer parameters case. Gupta and Sobel (1962b) developed a formula for evaluating $Pr\{\min_i F_i \geq c\}$, which involves $(p+1)$—fold summation when all $\nu_j (j=1, 2, ..., p)$ are integers and applied the results to the problem of selecting a subset containing populations with minimum variance.

We remark that at this point it may be considerable to introduce some prior distributions on $\sigma_1^2, \sigma_2^2, ..., \sigma_p^2$ as well as the prior on p_1. However it is left unsolved in this paper.

REFERENCES

(1) Chen, H.J. and Pickett, J.R. (1983). Selecting All Treatments Better Than a Control Under a Multivariate Normal Distribution and a Uniform Prior Distribution, *Comm. Stat.*(Accepted)

(2) Finney, D.J. (1941). The Joint Distribution of Variance Ratio Based on a Common Error Mean Square, *Annals of Eugenics 11*, 136—140.

(3) Gupta, S.S. and Sobel, M. (1962a). On the Smallest of Several Correlated F Statistics, *Biometrika 49*, 509—523.

(4) Gupta, S.S. and Sobel, M. (1962b). On Selecting a Subset containing the Population with the Smallest Variance, *Biometrika 49*, 495—507.

(5) Krishnaiah, P.R. (1965). Multiple Comparison Tests in Multi-Response Experiments, *Sankya 27*, 65—72

(6) Tiao, G.G. and Guttman, I. (1965). The Inverted Dirichlet Distribution with Applications, *J. Amer. Statist. Asso. 60*, 793—805.