TRU Mathematics 17-1 (1981)

ESTIMATION IN MIXTURE OF TWO GAMMA DISTRIBUTIONS WITH EQUAL SCALE PARAMETERS

RYUZO KANNO

(Received April 20, 1981)

1. INTRODUCTION. The mixed gamma distribution appears frequently in medical science, for example, ages of onset of autoimmune diseases. Masuyama [3] has analysed ages of onset of rheumatoid arthritis, which is one of the autoimmune diseases, assuming that it follows a mixed gamma distribution. In his paper, the parameters of gamma mixture were estimated by the method of moments based on two kinds of moment, that is, ordinary moment and the first moment of the logarithmically transformed variate. However, it seems that the procedure requires an especially highly skilled technique in practice. Thus there is room to find a simpler method of estimation.

In this paper we consider the estimation of parameters for a mixed gamma distribution $wG(a,P)+(1-w)G(b,q)$, $0<w<1$, assuming that it has equal scale parameters, namely $a = b$ holds. This assumption may be correspond with hit model in the radiation biology such as each member of both subpopulations has the same chance of hit but the minimum numbers of hits which impair the normal function are different. We here construct the moment estimators by using the first four k-statistics and show that the moment estimates are uniquely ditermined from the root of cubic equation having a unique solution. Next the asymptotic variances of the estimators are calculated for various values of the parameters. In order to investigate the sampling properties of the moment estimators, a simulation study was made. A practical example is taken up in the last section.

2. CONSTRUCTION OF MOMENT ESTIMATORS. The probability density function of the mixed gamma distribution with equal scale parameters is, for $x \geqq 0$,

$$
\begin{aligned}
&f(x) = wf_1(x) + (1-w)f_2(x), \qquad 0 < w < 1 \\
&f_1(x) = \frac{1}{a\Gamma(p)} \left(\frac{x}{a}\right)^{P-1} exp\left(-\frac{x}{a}\right), \quad f_2(x) = \frac{1}{a\Gamma(q)} \left(\frac{x}{a}\right)^{q-1} exp\left(-\frac{x}{a}\right),
\end{aligned}
\tag{2.1}
$$

where scale parameter a is positive constant and indices p and q are assumed positive integers relating $p < q$. We here construct the moment estimators of the parameters w, a, p, q by using Fisher's k-statistics whose expectations are the cumulants of the distribution.

Let κ_r be the r-th cumulant of (2.1) and k_r the r-th k-statistic from a sample of size n. For example the first eight cumulants of this mixed gamma distribution are expressed as follows:

$$
\begin{aligned}
\kappa_1 &= a(c_1\delta+q),\\
\kappa_2 &= a^2(c_2\delta^2+c_1\delta+q),\\
\kappa_3 &= a^3(c_3\delta^3+3c_2\delta^2+2c_1\delta+2q),\\
\kappa_4 &= a^4(c_4\delta^4+6c_3\delta^3+11c_2\delta^2+6c_1\delta+6q),\\
\kappa_5 &= a^5(c_5\delta^5+10c_4\delta^4+35c_3\delta^3+50c_2\delta^2+24c_1\delta+24q),\\
\kappa_6 &= a^6(c_6\delta^6+15c_5\delta^5+85c_4\delta^4+225c_3\delta^3+274c_2\delta^2+120c_1\delta+120q),\\
\kappa_7 &= a^7(c_7\delta^7+21c_6\delta^6+175c_5\delta^5+735c_4\delta^4+1624c_3\delta^3+1764c_2\delta^2+720c_1\delta+720q),\\
\kappa_8 &= a^8(c_8\delta^8+28c_7\delta^7+322c_6\delta^6+1960c_5\delta^5+6769c_4\delta^4+13132c_3\delta^3\\
&\qquad +13068c_2\delta^2+5040c_1\delta+5040q),
\end{aligned} \tag{2.2}
$$

where $\delta = p - q$ and c_r indicates a polynominal equation of the r-th degree in w as the following.

$$
\begin{aligned}
c_1 &= w,\\
c_2 &= w(1 - w),\\
c_3 &= w(1 - w)(1 - 2w),\\
c_4 &= w(1 - w)[1 - 6w(1 - w)],\\
c_5 &= w(1 - w)(1 - 2w)[1 - 12w(1 - w)],\\
c_6 &= w(1 - w)[1 - 30w(1 - w) + 120w^2(1 - w)^2],\\
c_7 &= w(1 - w)(1 - 2w)[1 - 60w(1 - w)+360w^2(1 - w)^2],\\
c_8 &= w(1 - w)[1 - 126w(1 - w) + 1680w^2(1 - w)^2 - 5040w^3(1 - w)^3].
\end{aligned} \tag{2.3}
$$

Note. The coefficient of δ^i in κ_r is expressed as $ka^r c_i$, where constant k is the same as the coefficient of the term δ^i in the expansion of $(\delta + 1)(\delta + 2)\ldots(\delta + r - 1)$. And the coefficient of q in κ_r is equal to $(r - 1)!$.

Equating the first four cumulants in (2.2) to the first four k-statistics k_1, k_2, k_3, k_4, we may obtain the estimating equations as follows:

$$
\begin{aligned}
k_1/\hat{a} &= \hat{q} + t,\\
k_2/\hat{a}^2 &= (\hat{q} + t) + ut^2,\\
k_3/\hat{a}^3 &= 2(\hat{q} + t) + 3ut^2 + u(u - 1)t^3,\\
k_4/\hat{a}^4 &= 6(\hat{q} + t) + 11ut^2 + 6u(u - 1)t^3 + u(u^2 - 4u + 1)t^4,
\end{aligned} \tag{2.4}
$$

where it denotes that $u = (1 - \hat{w})/\hat{w}$, $t = \hat{w}(\hat{p} - \hat{q})$ and the cap over a parameter indicates the moment estimator. If we let

$$(2.5)\qquad \begin{aligned} x &= ut^2 \\ y &= u(u-1)t^3 \\ z &= u(u^2-4u+1)t^4, \end{aligned}$$

then the following equation holds among the x, y and z.

$$(2.6)\qquad x^3 + \frac{1}{2}xz - \frac{1}{2}y^2 = 0.$$

On the other hand, from (2.4) we can obtain the following equations.

$$(2.7)\qquad \begin{aligned} x &= (k_2 - k_1\hat{a})/\hat{a}^2, \\ y &= (k_3 - 3k_2\hat{a} + k_1\hat{a}^2)/\hat{a}^3, \\ z &= (k_4 - 6k_3\hat{a} + 7k_2\hat{a}^2 - k_1\hat{a}^3)/\hat{a}^4. \end{aligned}$$

Substituting x, y and z in (2.7) into the equation of (2.6) yields the following cubic equation for $\hat{a}$.

$$(2.8)\qquad 2(k_1^3 + k_1k_2)\hat{a}^3 + 2(k_2^2 - 3k_1^2k_2 - 2k_1k_3)\hat{a}^2 + (k_1k_4 + 6k_1k_2^2)\hat{a} + (k_3^2 - 2k_2^3 - k_2k_4) = 0.$$

It can be seen that from the definition of x and the first equation of (2.7), the value of $\hat{a}$ is restricted so that $0 < \hat{a} \leqq k_2/k_1$ and further the above cubic equation has always a solution in this interval. From the first and second equations of (2.5) and (2.7), it follows that

$$(2.9)\qquad \begin{aligned} ut^2 &= (k_2 - k_1\hat{a})/\hat{a}^2, \\ u(u-1)t^3 &= (k_3 - 3k_2\hat{a} + k_1\hat{a}^2)/\hat{a}^3. \end{aligned}$$

Eliminating u from the two equations yields a quadratic equation for t.

$$(2.10)\qquad \left(\frac{k_2}{\hat{a}^2} - \frac{k_1}{\hat{a}}\right)t^2 + \left(\frac{k_3}{\hat{a}^3} - \frac{3k_2}{\hat{a}^2} + \frac{k_1}{\hat{a}}\right)t - \left(\frac{k_2}{\hat{a}^2} - \frac{k_1}{\hat{a}}\right)^2 = 0,$$

where $\hat{a}$ denotes the solution of cubic equation in (2.8). It is clear that this quadratic equation has always two real roots with opposite sign. We here choose negative one, since we assume that $p < q$. Thus the value of u is determined from the first equation in (2.9) as follows:

$$(2.11)\qquad u = (k_2 - k_1\hat{a}) \,/\, (t^2\hat{a}^2).$$

where $\hat{a}$ and t denote the estimates obtained from (2.8) and (2.10), respectively. At last from the definitions of u and t and the first equation in (2.4), it follows that

$$\hat{w} = t^2 / \left(t^2 + \frac{k_2}{\hat{a}^2} - \frac{k_1}{\hat{a}}\right),$$

$$(2.12) \qquad \hat{p} = \frac{1}{t}\left(\frac{k_2}{\hat{a}^2} - \frac{k_1}{\hat{a}}\right) + \frac{k_1}{\hat{a}},$$

$$\hat{q} = \frac{k_1}{\hat{a}} - t.$$

The estimation procedures are summarized as follows:

1) The estimate of parameter a is determined to solve the cubic equation in (2.8) under the restriction $0 < \hat{a} \leqq k_2/k_1$.

2) After substituting the estimate $\hat{a}$ into the equation of (2.10), we solve the quadratic equation for t. Since we assume that $p < q$, the negative root is admissible one as the estimate of t.

3) At last, substituting of $\hat{a}$ and t into the equations of (2.12), we may obtain the estimates of p, q and w.

3. ASYMPTOTIC VARIANCES OF THE MOMENT ESTIMATORS. We here consider the asymptotic variances of the moment estimators given in the previous section. Let $\hat{\alpha}_1 = \hat{w}$, $\hat{\alpha}_2 = \hat{a}$, $\hat{\alpha}_3 = \hat{p}$ and $\hat{\alpha}_4 = \hat{q}$, then the asymptotic variance of order $1/n$ for the $\hat{\alpha}_i$ $(i = 1, 2, 3, 4)$ may be found the following formula as mentioned in Robertson-Fryer [4].

$$(3.1) \qquad Var(\hat{\alpha}_i) = \sum_{j=1}^{4} (D^{ij})^2\, Var(k_j) + \sum_{j \neq k}^{4} D^{ij} D^{ik}\, Cov(k_j, k_k),$$

where D^{ij} is the (i, j) element of the inverse of the matrix D whose (i, j) element is $D_{ij} = \partial k_i / \partial \hat{\alpha}_j$, where the partial derivatives are to be evaluated at the point $\hat{\alpha}_j = \alpha_j$. The partial derivatives may be easily found from the first four equations of (2.2). If the practical value of D_{ij} is known, we may calculate the inverse matrix $D^{-1} = (D^{ij})$ with the aid of computer. The asymptotic variances and covariances of the k-statistics have been given to order $1/n$ as expression of cumulants by Kendall-Stuart [2]. It should be noted that the first eight cumulants are needed to calculate the asymptotic variances and covariances of the first four k-statistics.

The values of asymptotic variances of the moment estimators for the case $n = 1000$ are plotted as a function of w, with $a = 1$, $p = 5$ and some fixed value of q. The results can easily be rescaled for any value of n, by multiplying by $1000/n$.

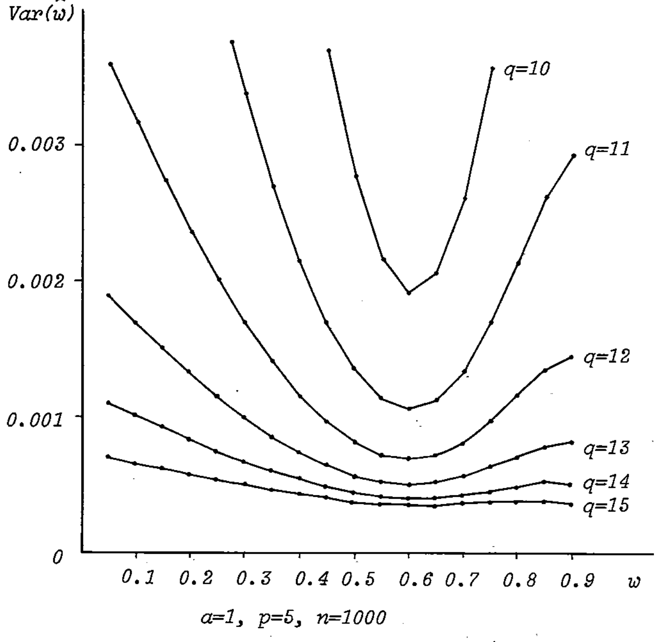

Fig. 1 Asymptotic variance of $\hat{w}$

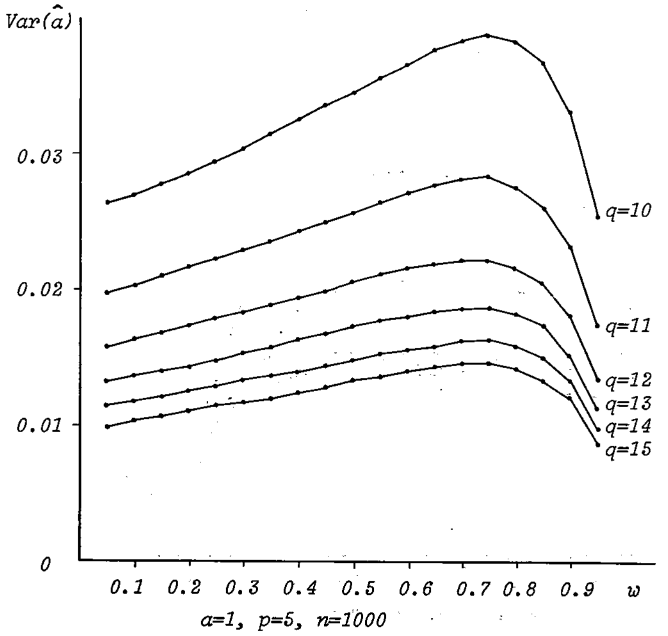

Fig. 2 Asymptotic variance of $\hat{a}$

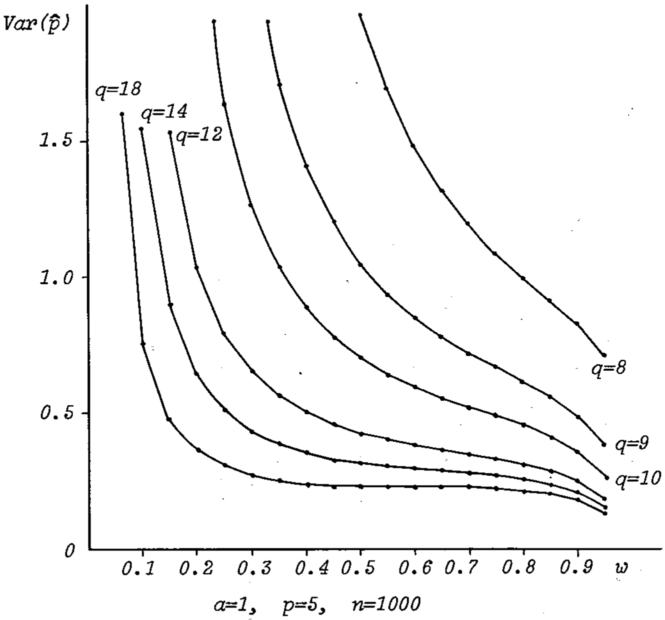

Fig. 3 Asymptotic variance of $\hat{p}$

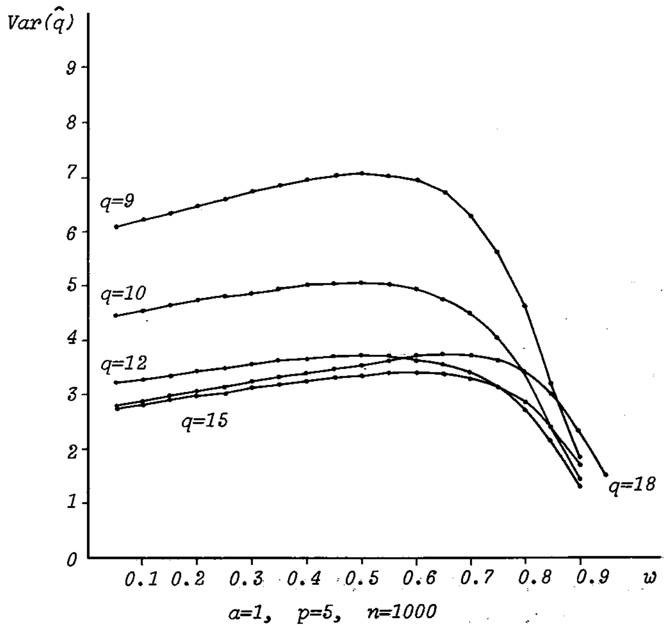

Fig. 4 Asymptotic variance of $\hat{q}$

As expected, the variances generally have a tendency to decrease as the component distributions are well separated, that is, as the value of q becomes larger. However, as Fig. 4 shows, when the value of q is sufficiently large the variance of $\hat{q}$ does not decrease even if q increases. From Fig. 1 and Fig.2, it can be seen that the variance of $\hat{w}$ becomes minimum for some value of w and the variance of $\hat{a}$ becomes a maximum for some value of w. On the other hand, as Fig. 3 shows, the variance of $\hat{p}$ decreases uniformly as w increases.

4. THE SIMULATION RESULTS. In order to investigate the sampling properties of the moment estimators, a Monte Carlo study was made. The computations were run on an IBM 3031 using programs written in FORTRAN IV. By empirical sampling, 30 sets of n observations were generated from a mixed gamma distribution with known parameters values, and the moment estimates were found from each set of

Table 1. Sampling distribution of the moment estimates based on 30 simulations for several sample size.

$(w = 0.4,\ a = 1.0,\ p = 5,\ q = 15)$

n		mean of estimates	variance of estimates	asymptotic variance
250	w	0.4010	0.002062	0.001768
	a	0.9834	0.040972	0.049409
	p	5.2020	0.814430	1.266738
	q	15.8839	9.687776	13.026493
500	w	0.4011	0.000985	0.000884
	a	0.9833	0.010930	0.024705
	p	5.1102	0.326881	0.633369
	q	15.4760	3.375192	6.513247
750	w	0.4025	0.000510	0.000589
	a	0.9849	0.007205	0.016470
	p	5.0731	0.182647	0.422246
	q	15.3349	1.723997	4.342165
1000	w	0.4000	0.000448	0.000442
	a	1.0119	0.008120	0.012352
	p	4.9633	0.198266	0.316684
	q	14.9178	1.946904	3.256623

Table 2. Sampling distribution of the moment estimates based on *30* simulations for some selected parameter values.

(sample size $n = 500$)

parameters		mean of estimates	variance of estimates	asymptotic variance
w	*0.2*	*0.1983*	*0.000775*	*0.001165*
a	*1.0*	*0.9938*	*0.016635*	*0.021807*
p	*5*	*5.1095*	*0.606105*	*1.075611*
q	*15*	*15.3403*	*4.359503*	*5.953983*
w	*0.4*	*0.4011*	*0.000985*	*0.000884*
a	*1.0*	*0.9833*	*0.010930*	*0.024705*
p	*5*	*5.1102*	*0.326881*	*0.633369*
q	*15*	*15.4760*	*3.375192*	*6.513247*
w	*0.6*	*0.5966*	*0.001003*	*0.000723*
a	*1.0*	*1.0097*	*0.016979*	*0.027772*
p	*5*	*4.9561*	*0.217138*	*0.543835*
q	*15*	*15.0834*	*2.991390*	*6.830301*
w	*0.8*	*0.7965*	*0.000508*	*0.000761*
a	*1.0*	*0.9770*	*0.018717*	*0.028585*
p	*5*	*5.1510*	*0.328299*	*0.486986*
q	*15*	*15.5983*	*4.057123*	*5.743927*

observations. The sample means and variances calculated from each set of *30* estimates along with the asymptotic variances are given in Table 1 for several values of n, and in Table 2 for some selected parameter values. Table 2 shows that the biases of $\hat{w}$ and $\hat{a}$ can be ignored, but a small amount of bias in $\hat{p}$ and $\hat{q}$. However, as shown in Table 1, the bias decreases as sample size n increases and it can be ignored for large value of n. In general, it seems that asymptotic variance overestimates the Monte Carlo variance.

5. APPLICATION TO THE DATA OF RHEUMATOID ARTHRITIS. To illustrate the practical application of computation procedures described in this paper, we take up the data appeared in Inagaki [1] which also has been taken up in Masuyama [3] as an example. The data is ages of onset of the rheumatoid athritis (female) and gives, as shown in Table 3, the frequency distribution showing the bimodality with peak at about 20 and 35 years of age.

Table 3

age of onset	frequency
0 - 4	2
5 - 9	7
10 - 14	34
15 - 19	72
20 - 24	66
25 - 29	72
30 - 34	76
35 - 39	74
40 - 44	68
45 - 49	49
50 - 54	49
55 - 59	36
60 - 64	8
65 - 69	3
70 - 74	2
75 - 79	1
Total	619

The values of the first four k-statistics are

$$k1 = 34.309370,$$
$$k_2 = 194.53642,$$
$$k_3 = 570.20306,$$
$$k_4 = -24963.358.$$

The cubic equation of (2.8) becomes as follows;

$$94122.218\hat{a}^3 - 1376535.7\hat{a}^2 + 6934031.7\hat{a} - 9542821.1 = 0.$$

This cubic equation has a single root $\hat{a} = 2.1770798$ as the solution in the interval $(0, k_2/k_1] = (0, 5.670067]$. By using the value of $\hat{a}$, equation (2.10) becomes

$$25.28488t^2 - 52.113925t - 639.32554 = 0.$$

On solving for the negative root of this equation, we obtain $t = -4.1023872$. When these values of $\hat{a}$ and t are substituted into the equations of (2.12), we have the following estimates:

$$\hat{p} = 9.595896,$$
$$\hat{q} = 19.861741,$$
$$\hat{w} = 0.399615.$$

Since indices p and q are assumed to be positive integers, we should round up the values of $\hat{p}$ and $\hat{q}$. Thus we have the final estimates: $p^* = 10$ and $q^* = 20$. Assuming that $p = 10$ and $q = 20$, we anew estimate w^* and a^* from the first two equations of (2.4). It becomes that $a^* = 2.197$, $w^* = 0.438$. This result is the same as that of Masuyama [3]. Based on these estimates of parameters and by using the method described in section 3, we obtain the asymptotic variances of the estimators as follows:

$$Var(w^*) = 0.000994,$$
$$Var(a^*) = 0.101724,$$
$$Var(p^*) = 1.835789,$$
$$Var(q^*) = 9.738741.$$

REFERENCES

[1] INAGAKI, K. (1967): Kansetsu Ryumati no Ekigaku. Sogo-Rinsyo, 16, 112-123. (in Japanese).

[2] KENDALL, M.G. and STUART, A. (1958): *The advanced theory of statistics. Vol. 1*, London: Charles Griffin.

[3] MASUYAMA, M. (1977): A mixture of two gamma distributions applied to Rheumatoid Arthritis. Rep. Stat. Appl. Res., 24, no.2, 81-85.

[4] ROBERTSON, C.A. and FRYER, J.G. (1970): The bias and accuracy of moment estimators. Biometrika, 57, 57-65.