Fisher's Second k Statistics

Hark Hwang*

1. Introduction

The practical importance of determining the sampling distribution of moments calculated from a sample was early recognized in the development of statistical theory.

Further, because the theoretical distributions were found often to have intractable functional forms, the moments (or cumulants) of sample moments (or cumulants) provide a means of deriving functions approximating to such theoretical distribution.

Thus, expansions in terms of Gram Charlier or Edgeworth series with parameters derived from sampling moments have frequently been used to provide such approximations.

The problem of finding the sampling moments took a leap forward with Fisher's [2] introduction of k statistics and their sampling cumulants which are defined as follows.

A set of independent random variables $X_1, X_2, \ldots, X_n$ is assumed with identical probability density function for the X's. This probability density function has as many cumulants K_r, $r=1, 2\ldots$ as desired.

Functions k_i of the random variables $X_1, X_2, \ldots, X_n$ are defined such that they are unbiased estimates of K_i i.e., $E(k_i)=K_i$

For the second k statistic (sample variance), the first six cumulants have been calculated by Fisher [2].

In this paper the first seven cumulants of the sample variance in terms

*Assistant Professor, Department of Industrial Science, Korea Advanced Institute of Science.

of the population cumulants are established.

2. Cumulants of the Sample Variance

Let X, $X_i (i=1, 2, \ldots, N)$ be independent identically distributed random variables with cumulant K_i and $S1=\Sigma X_i$ and $S2=\Sigma X_i^2$.

Then the sample mean is defined as $\bar{X}=S1/N$ and the sample variance, V is

$$V=\frac{1}{N-1}\left(S2-\frac{S1^2}{N}\right)$$

The jth moment of V, $MV(j)$ is

$$MV(j)=\sum_{p=0}^{j}\binom{j}{p}\frac{(-1)^p}{(N-1)^j N^p}MS12\ (2p, j-p) \qquad (2.1)$$

where $MS12(i,j)=E(\Sigma X)^i(\Sigma X^2)^j$

A computer program is established to calculate $MS12(i,j)$ [3] and with those result the first seven moments of the sample variance are obtained from equation (2. 1).

Those moments are transformed to cumulants in the usual manner and then modified such that all terms of ith cumulant have a common denominator, $(N(N-1))^{i-1}$.

For instance, the second cumulant of the sample variance, $K(2^2)$ is

$$K(2^2)=\frac{K_4}{N}+\frac{2K_2^2}{N-1}=\frac{1}{N(N-1)}(K_4(N-1)+2K_2^2N)$$

and $K_4(N-1)+2K_2^2N$ is the output for $K(2^2)$.

3. Results and Discussion

These results are consistent with those in Table (3. 1) given by David and Kendall [1].

The seventh cumulant of the sample variance consists of 149 terms and is shown on Table (3. 1).

Table 3.1. *K*(2^7) with Dividing Factor (*N*(*N*−1))6**

	N	*N***2	*N***3	*N***4	*N***5	*N***6	*N***7
*K*14	1	−6	15	−20	15	−6	1
*K*12**K*2		−84	420	−840	840	−420	84
*K*11**K*3		−560	2520	−4480	3920	−1680	280
*K*10**K*4		−2100	8820	−14840	12600	−5460	980
*K*9**K*5		−4928	19488	−31360	25760	−10752	1792
*K*8**K*6		−8120	31080	−49280	40320	−16968	2968
*K*7**2		−4752	17928	−28160	22800	−9456	1576
*K*10**K*2**2			2520	−10080	15120	−10080	2520
*K*9**K*3**K*2			26880	−94080	120960	−67200	13440
*K*8**K*3**2			64960	−198240	219520	−104160	17920
*K*8**K*4**K*2			80640	−262080	322560	−181440	40320
*K*7**K*5**K*2			145152	−443520	524160	−282240	56448
*K*7**K*4**K*3			336000	−947520	1003520	−477120	85120
*K*6**2**K*2			89040	−268800	319200	−174720	37968
*K*6**K*4**2			372960	−987840	1021440	−497280	99680
*K*6**K*5**K*3			519680	−1397760	1447040	−672000	116480
*K*5**2**K*4			470400	−1202880	1209600	−564480	100800
*K*8**K*2**3				−33600	100800	−100800	33600
*K*7**K*3**K*2**2				−403200	1008000	−806400	201600
*K*5**2**K*2**2				−564480	1249920	−967680	241920
*K*6**K*4**K*2**2				−907200	2116800	−1713600	504000
*K*6**K*3**2**K*2				−1424640	2970240	−1962240	416640
*K*5**K*4**K*3**K*2				−5160960	9999360	−6612480	1451520
*K*5**K*3**3				−1397760	2419200	−1344000	241920
*K*4**3**K*2				−1249920	2378890	−1653120	443520
*K*4**2**K*3**2				−2983680	4959360	−2835840	551040
*K*6**K*2**4					201600	−403200	201600
*K*5**K*3**K*2**3					2150400	−3225600	1075200
*K*4**2**K*2**3					1612800	−2419200	1075200
*K*4**K*3**2**K*2**2					7257600	−8467200	2419200
*K*3**4**K*2					1935360	−1774080	403200
*K*4**K*2**5						−483840	483840
*K*3**2**K*2**4						−1612800	806400
*K*2**7							46080

The table reads as follows:

$$K(2^7)=\frac{K_2^{\,7}}{[N(N-1)]^6}(46080N^7)+\frac{K_3^{\,2}K_2^{\,4}}{[N(N-1)]^6}(-1612800N^6+806400N^7)+\cdots$$

Fisher established the first six cumulants of the sample variance after many times of corrections and it is significant to confirm the Fisher's result with the computer output and extend his work.

Also the result obtained provides a basis for developing (i,j)th cumulant of sample mean and sample variance, $K(1^i\ 2^j)$ using "rule 10" from Kendall [4].

ABSTRACT

The cumulants of the sample variance are obtained in terms of the population cumulants using ALMAP (Algebraic Manipulation Package) and these results extend those of Fisher who gave the first six cumulants of the sample variance.

REFERENCES

[1] David, F. N., Kendall, M. G. and Barton, D. E., *Symmetric Functions and Allied Tables*, Cambridge University Press, 1966

[2] Fisher, R. A., "Moments and Product Moments of Sampling Distributions," *Proceedings of the London Mathematical Society*, Series 2, Vol. 30, 1928

[3] Hwang, H., "The Cumulants of the Non-normal t Distribution," *The Journal of the Korean Statistical Society*, Vol. 5, Number 2, 1976

[4] Kendall, M. G. and Stuart, A., *The Advanced Theory of Statistics*, Vol. 1, Hafner Publishing Company, New York, 1958.