# The Distribution of McKay's Approximation for the Coefficient of Variation

Johannes Forkman[a,1], Steve Verrill[b]

[a]Department of Biometry and Engineering, Swedish University of Agricultural Sciences,

P.O. Box 7032, SE-750 07 Uppsala, Sweden

[b]U.S.D.A., Forest Products Laboratory,

1 Gifford Pinchot Drive, Madison, WI 53726, USA

**Abstract**

McKay's chi-square approximation for the coefficient of variation is type II noncentral beta distributed and asymptotically normal with mean $n - 1$ and variance smaller than $2(n - 1)$.

Key words: Coefficient of variation, McKay's approximation, Noncentral beta distribution.

[1]corresponding author. *E-mail address*: johannes.forkman@bt.slu.se

# 1 Introduction

The coefficient of variation is defined as the standard deviation divided by the mean. This measure, which is commonly expressed as a percentage, is widely used since it is often necessary to relate the size of the variation to the level of the observations. McKay (1932) introduced a $\chi^2$ approximation for the coefficient of variation calculated on normally distributed observations. It can be defined in the following way.

**Definition 1.** *Let $y_j$, $j = 1, ..., n$, be $n$ independent observations from a normal distribution with expected value $\mu$ and variance $\sigma^2$. Let $\gamma$ denote the population coefficient of variation, i.e. $\gamma = \sigma/\mu$, and let $c$ denote the sample coefficient of variation, i.e.*

$$c = \frac{1}{m}\sqrt{\frac{1}{n-1}\sum_{j=1}^{n}(y_j - m)^2}, \quad m = \frac{1}{n}\sum_{j=1}^{n} y_j.$$

*McKay's approximation $K_n$ is defined as*

$$K_n = \left(1 + \frac{1}{\gamma^2}\right)\frac{(n-1)\,c^2}{1 + \frac{(n-1)c^2}{n}}. \tag{1}$$

As pointed out by Umphrey (1983) formula (1) appears slightly different in the original paper by McKay (1932) since McKay used the maximum likelihood estimator of $\sigma^2$, with denominator $n$, instead of the unbiased estimator with denominator $n-1$.

McKay (1932) claimed that $K_n$ is approximately central $\chi^2$ distributed with $n-1$ degrees of freedom provided that $\gamma$ is small ($\gamma < 1/3$). This result was established by expressing the probability density function of $c$ as a contour integral and making an approximation in the complex plane. McKay did not theoretically express the size of the error of the approximation. For this reason Fieller (1932), in immediate connection to McKay's paper, investigated McKay's approximation $K_n$ numerically and concluded that it is "quite adequate for any practical purpose." Also Pearson (1932) examined the new approximation and found it "very satisfactory." Later Iglewicz & Myers (1970) studied the usefulness of McKay's approximation for calculating quantiles of the distribution of the sample coefficient of variation $c$ when the underlying distribution is normal. They compared results according to the approximation with exact results and found that the approximation is accurate. Umphrey (1983) corrected a similar study made by Warren (1982) and concluded that McKay's approximation is adequate. Vangel (1996) analytically compared the cumulative density function of McKay's approximation with the cumulative density function of the "naïve" $\chi^2$ approximation

$$N_n = \frac{(n-1)\,c^2}{\gamma^2}$$

and showed that McKay's approximation is substantially more accurate. Vangel also proposed a small modification of McKay's approximation useful for calculating approximate confidence intervals for the coefficient of variation. Forkman (2006) suggested McKay's approximation for testing the hypothesis that two coefficients of variation are equal. Another test for the hypothesis of equal coefficients of variation, also based on McKay's approximation, was proposed by Bennett (1976).

It is thus well documented that McKay's approximation is approximately central $\chi^2$ distributed with $n-1$ degrees of freedom, and useful applications have been suggested. In this paper it is shown that McKay's approximation is type II noncentral beta distributed, and its asymptotic behavior is investigated.

# 2 The distribution of McKay's approximation

If $U$ and $V$ are independent central $\chi^2$ distributed random variables with $u$ and $v$ degrees of freedom respectively, the ratio $R = V/\left(U+V\right)$ is beta distributed with $v/2$ and $u/2$ degrees of freedom respectively. If $V$ is instead a noncentral $\chi^2$ distributed random variable the ratio $R$ is noncentral beta distributed (Johnson & Kotz, 1970). In this case Chattamvelli (1995) calls the distribution of $R$ the type I noncentral beta distribution and the distribution of $1-R$ the type II noncentral beta distribution. We shall in agreement with Chattamvelli (1995) use the following definition.

**Definition 2.** *Let $U$ be a central $\chi^2$ distributed random variable with $u$ degrees of freedom, and let $V$ be a noncentral $\chi^2$ distributed random variable, independent of $U$, with $v$ degrees of freedom and noncentrality parameter $\lambda$. The type II noncentral beta distribution with parameters $u/2$, $v/2$ and $\lambda$, denoted Beta II $(u/2,\ v/2,\ \lambda)$ is defined as the distribution of $U/\left(U+V\right)$.*

The following theorem states that the random variable $K_n$, claimed by McKay (1932) to be approximately $\chi^2$ distributed, is type II noncentral beta distributed.

**Theorem 3.** *The distribution of McKay's approximation $K_n$, as defined in Definition 1, is*

$$n\left(1+\frac{1}{\gamma^2}\right) \text{Beta II}\left(\frac{n-1}{2},\ \frac{1}{2},\ \frac{n}{\gamma^2}\right). \tag{2}$$

*Proof.* Let $s$ denote the standard deviation, i.e. $s = cm$. Then the second factor in (1) can be written

$$\frac{(n-1)\,c^2}{1+\frac{(n-1)c^2}{n}} = \frac{\sum_{j=1}^{n}\left(y_j-m\right)^2}{m^2+\frac{1}{n}\sum_{j=1}^{n}\left(y_j-m\right)^2}$$

$$= \frac{n\sum_{j=1}^{n}\left(y_j-m\right)^2}{\sum_{j=1}^{n}\left(y_j-m\right)^2+\sum_{j=1}^{n}m^2} = \frac{nU}{U+V},$$

where $U = \sum_{j=1}^{n}\left(y_j-m\right)^2/\sigma^2$ and $V = \sum_{j=1}^{n} m^2/\sigma^2$. Here $U$ is central $\chi^2$ distributed with $n-1$ degrees of freedom. The average $m$ is normally distributed with expected value $\mu$ and variance $\sigma^2/n$. Consequently $nm^2/\sigma^2$, i.e. $V$, is $\chi^2$ distributed with 1 degree of freedom and noncentrality parameter $n\mu^2/\sigma^2 = n/\gamma^2$. Since the sums of squares $\sum_{j=1}^{n}\left(y_j-m\right)^2$ and $\sum_{j=1}^{n} m^2$ are independent the theorem follows. □

It is well known that $\sqrt{n}/c$ is noncentral $t$ distributed with $n-1$ degrees of freedom and noncentrality parameter $\sqrt{n}/\gamma$ (e.g. Owen, 1968). Theorem 3 is easily proven from this starting point as well.

We also note that the factor $n(1+1/\gamma^2)$ in (2) is the expected value of $U+V$ as defined in the proof of Theorem 3. This observation suggests application of the law of large numbers when investigating the convergence of McKay's approximation.

**Theorem 4.** *The distribution of McKay's approximation $K_n$ as defined in Definition 1, equals the distribution of $U_nW_n$, where $U_n$ is a central $\chi^2$ distributed random variable with $n-1$ degrees of freedom and $W_n$ is a random variable that converges in probability to 1.*

*Proof.* Let $Z_k$, $k = 1, 2, \ldots, n-1$, be independent standardized normal random variables. Then

$$\frac{U_n}{n-1} \stackrel{d}{=} \frac{1}{n-1} \sum_{k=1}^{n-1} Z_k^2,$$

which converges almost surely to 1. Let also $Z$ denote a standardized normal random variable, and let

$$\frac{V_n}{n} = \frac{1}{n}\left(Z + \frac{\sqrt{n}}{\gamma}\right)^2 = \frac{Z^2}{n} + \frac{2Z}{\gamma\sqrt{n}} + \frac{1}{\gamma^2}.$$

$V_n/n$ converges in probability to $1/\gamma^2$. Thus

$$\left(\frac{U_n + V_n}{n}\right) \xrightarrow{p} 1 + \frac{1}{\gamma^2}. \tag{3}$$

By Theorem 3

$$K_n \stackrel{d}{=} n\left(1 + \frac{1}{\gamma^2}\right)\frac{U_n}{U_n + V_n} = W_nU_n,$$

where $W_n = n(1 + 1/\gamma^2)/(U_n + V_n)$, by (3), converges in probability to 1. □

Given Theorem 4 one might assume that McKay's approximation $K_n$ is asymptotically normal with mean $n-1$ and variance $2(n-1)$. Instead the following result holds.

**Theorem 5.** *Let $K_n$ be McKay's approximation and $\gamma$ the coefficient of variation as defined in Definition 1. Then*

$$\frac{K_n - (n-1)}{\sqrt{2(n-1)}} \xrightarrow{d} \mathrm{N}\left(0, \frac{1 + 2\gamma^2}{1 + 2\gamma^2 + \gamma^4}\right). \tag{4}$$

*Proof.* Let $Z$ denote a standardized normal random variable, and let $V_n = (Z + \sqrt{n}/\gamma)^2$. Let $U_n$ be a central $\chi^2$ distributed random variable with $n-1$ degrees of freedom, independent of $V_n$. Then, by Theorem 3,

$$\frac{K_n - (n-1)}{\sqrt{2(n-1)}} \stackrel{d}{=} \frac{1}{\sqrt{2(n-1)}}\left(\frac{n(1 + 1/\gamma^2)U_n}{U_n + V_n} - (n-1)\right) = A_nB_n, \tag{5}$$

where, by (3),

$$A_n = \frac{n}{U_n + V_n} \xrightarrow{p} \frac{\gamma^2}{1+\gamma^2} \tag{6}$$

and

$$B_n = \frac{1}{\sqrt{2(n-1)}}\left(\frac{U_n(\gamma^2+1)}{\gamma^2} - \frac{(n-1)(U_n+V_n)}{n}\right).$$

We obtain

$$B_n = C_n + D_n + E_n + F_n \tag{7}$$

where

$$C_n = \frac{U_n - (n-1)}{\gamma^2\sqrt{2(n-1)}} \xrightarrow{d} \mathrm{N}\left(0, \frac{1}{\gamma^4}\right), \tag{8}$$

$$D_n = \frac{-\sqrt{2}(n-1)Z}{\gamma\sqrt{n(n-1)}} \xrightarrow{d} \mathrm{N}\left(0, \frac{2}{\gamma^2}\right), \tag{9}$$

$$E_n = \frac{U_n}{n\sqrt{2(n-1)}} \xrightarrow{p} 0, \tag{10}$$

$$F_n = \frac{-(n-1)Z^2}{n\sqrt{2(n-1)}} \xrightarrow{p} 0. \tag{11}$$

Since $C_n$ is independent of $D_n$, results (7) – (11) imply that

$$B_n \xrightarrow{d} \mathrm{N}\left(0, \frac{1}{\gamma^4} + \frac{2}{\gamma^2}\right). \tag{12}$$

Results (5), (6) and (12) yield the theorem. □

# 3 Discussion

We have seen that McKay's $\chi^2$ approximation for the coefficient of variation is exactly type II noncentral beta distributed. This observation provides insight into the approximation, originally derived by complex analysis. We showed that McKay's $\chi^2$ approximation in distribution equals the product of a $\chi^2$ distributed random variable and a variable that converges in probability to 1. Nevertheless, according to Theorem 5, McKay's $\chi^2$ approximation is asymptotically normal with mean $n-1$ and variance $2(n-1)(1+2\gamma^2)/(1+\gamma^2)^2$, where $\gamma$ is the coefficient of variation. Since it has previously been assumed that McKay's approximation is "asymptotically exact" (Vangel, 1996) it is surprising that the variance does not equal $2(n-1)$. It should be noted, however, that McKay's $\chi^2$ approximation is intended for the cases in which the coefficient of variation $\gamma$ is smaller than 1/3. This requirement should be fulfilled when analyzing observations from a positive variable that is approximately normally distributed, since otherwise $\sigma > \mu/3$ and the probability of negative observations is not negligable. Provided that $\gamma < 1/3$ the standardized McKay's $\chi^2$ approximation (4) converges in distribution to a normal distribution with expected value 0 and variance larger than 0.99 but smaller than 1. McKay's $\chi^2$ approximation should consequently asymptotically be sufficiently accurate for most applications.

Though the inverse of the coefficient of variation is noncentral $t$ distributed and algorithms for calculating the cumulative density function of this distribution nowadays exist (Lenth, 1989), McKay's approximation is still adequate and may be useful for various composite inferential problems on the coefficient of variation in normally distributed data. Algorithms for computing the cumulative distribution function of the noncentral beta distribution were reviewed by Chattamvelli (1995). The open source software R makes use of algorithms given by Lenth (1987) and Frick (1990).

# 4 Acknowledgements

We thank Prof. Dietrich von Rosen for ideas and discussions. The Centre of Biostochastics, Swedish University of Agricultural Sciences and Pharmacia Diagnostics AB financed the research.

# 5 References

Bennett, B.M. (1976), On an Approximate Test for Homogeneity of Coefficients of Variation, in: W.J. Ziegler, eds. Contributions to Applied Statistics dedicated to A.

Linder, *Experientia Suppl.* **22**, 169–171.

Chattamvelli R. (1995), A Note on the Noncentral Beta Distribution Function, *Amer. Statist.* **49**, 231–234.

Fieller, E.C. (1932), A Numerical Test of the Adequacy of A. T. McKay's Approximation, *J. Roy. Statist. Soc.* **95**, 699–702.

Forkman, J. (2006), Statistical Inference for the Coefficient of Variation in Normally Distributed Data, *Centre of Biostochastics, Swedish University of Agricultural Sciences, Research Report* 2006:2.

Frick, H. (1990), Algorithm AS R84: A Remark on Algorithm AS 226: Computing Non-central Beta Probabilities, *Appl. Statist.* **39**, 311–312.

Iglewicz, B. and Myers, R.H. (1970), Comparison of Approximations to the Percentage Points of the Sample Coefficient of Variation, *Technometrics* **12**, 166–169.

Johnson, N.L. and Kotz, S. (1970), *Distributions in Statistics: Continuous Univariate Distributions–2* (Wiley, New York).

Lenth, R.V. (1987), Algorithm AS 226: Computing Noncentral Beta Probabilities, *Appl. Statist.* **36**, 241–244.

Lenth, R.V. (1989), Algorithm AS 243: Cumulative Distribution Function of the non-central $t$ Distribution, *Appl. Statist.* **38**, 185–189.

McKay, A.T. (1932), Distribution of the Coefficient of Variation and the Extended "t" Distribution, *J. Roy. Statist. Soc.* **95**, 695–698.

Owen, D.B. (1968), A Survey of Properties and Applications of the Noncentral t-Distribution, *Technometrics* **10**, 445–478.

Pearson, E.S. (1932), Comparison of A. T. McKay's Approximation with Experimental Sampling Results, *J. Roy. Statist. Soc.* **95**, 703–704.

Umphrey, G.J. (1983), A Comment on McKay's Approximation for the Coefficient of Variation, *Commun. Stat. – Simul. C.* **12**, 629–635.

Vangel, M.G. (1996), Confidence Intervals for a Normal Coefficient of Variation. *Amer. Statist.* **15**, 21–26.

Warren, W.G. (1982), On the Adequacy of the Chi-Squared Approximation for the Coefficient of Variation. *Commun. Stat. – Simul. C.* **11**, 659–666.