# A higher order approximation to the distribution of a non-central t-statistic under non-normality

Masafumi, Akahira
*Institute of Mathematics, University of Tsukuba*
*1-1-1 Tennoudai*
*Ibaraki (305-8571), Japan*
*E-mail: akahira@math.tsukuba.ac.jp*

Nao, Ohyauchi
*Institute of Mathematics, University of Tsukuba*
*1-1-1 Tennoudai*
*Ibaraki (305-8571), Japan*
*E-mail: naopu@math.tsukuba.ac.jp*

Shin-ichi, Kawai
*National Research Institute for Earth Science and Disaster Prevention*
*3-1 Tennoudai*
*Ibaraki (305-0006), Japan*
*E-mail: kawai@bosai.go.jp*

## 1. Introduction

Under normality, we derive a higher order approximation to the $100\alpha$ percentile using the Cornish-Fisher expansion (see Akahira (1995)), and show it to be fairly accurate. Recently, the limiting distribution of the non-central t-statistic under non-normality has been discussed by Bentkus et al. (2007), but its speed of covergence does not seem to be high. Here we derive a higher order approximation to the upper $100\alpha$ percentile, without the normality assumption, in a similar way to Akahira (1995), and also obtain the confidence limit and the confidence interval of a non-centrality parameter (see Akahira et al. (2009)).

## 2. The non-central t-statistic and the approximation of its distribution

Suppose that $X_1, X_2, \cdots, X_n, \cdots$ is a sequence of independent and identically distributed (i.i.d.) non-degenerate continuous random variables with mean $\mu$, variance 1 and finite sixth moment. Let $\mu_j := E[(X_1-\mu)^j]$ $(j = 2, \ldots, 6)$, $\bar{X} := (1/n)\sum_{i=1}^n X_i$, $S_n^2 := \sum_{i=1}^n (X_i - \bar{X})^2/(n-1)$. Define $T_n := \sqrt{n}\bar{X}/S_n$ as the non-central t-statistic when $\mu \neq 0$ where $S_n = \sqrt{S_n^2}$. If the underlying distribution is $N(\mu, 1)$, then $T_n$ follows the non-central t-distribution with $n-1$ degrees of freedom and a non-centrality parameter $\mu\sqrt{n}$. Here we put $\sigma_n := E(S_n)$. For any $\alpha$ with $0 < \alpha < 1$, there exists $t_\alpha$ such that $P\{T_n < t_\alpha\} = 1-\alpha$. The $t_\alpha$ is called the upper $100\alpha$ percentile of the distribution of the non-central t-statistic $T_n$. In order to obtain the approximation to a percentage point of the distribution of the non-central t-statistic, we calculate the mean, variance and covariance of $\bar{X}$ and $S_n$. First we have for any $t \in \mathbb{R}^1$

$$(1) \quad P_\mu\{T_n \leq t\} = P_\mu\left\{\sqrt{n}\bar{X} - tS_n \leq 0\right\} = P_\mu\left\{\sqrt{n}(\bar{X}-\mu) - t(S_n - \sigma_n) \leq -\mu\sqrt{n} + t\sigma_n\right\}.$$

Here, for each $i = 1, 2, \cdots$, let $X_i' := X_i - \mu$. Note that $\mu_j = E(X_i'^j)$ $(j = 2, \ldots .6)$. Then (1) is written as

$$(2) \quad P_\mu\{T_n \leq t\} = P_0\left\{\sqrt{n}\bar{X}' - t(S_n - \sigma_n) \leq a_n(t)\right\} = P_0\left\{Z - t(S_n - \sigma_n) \leq a_n(t)\right\},$$

where $a_n(t) := -\mu\sqrt{n}+t\sigma_n$ and $Z := \sqrt{n}\bar{X}'$ with $\bar{X}' := (1/n)\sum_{i=1}^{n} X_i'$. Since $E(X_i') = 0$ $(i = 1, 2, \cdots)$, putting $Y_n := Z - t(S_n - \sigma_n)$, we obtain the mean $E_t(Y_n) = 0$ and the variance

$$(3) \quad V_t(Y_n) = nV(\bar{X}') + t^2 V(S_n) - 2\sqrt{n}t\mathrm{Cov}(\bar{X}', S_n),$$

where $\mathrm{Cov}(\bar{X}', S_n)$ denotes the covariance between $\bar{X}'$ and $S_n$. In the right-hand side of (3), we have

$$(4) \quad nV(\bar{X}') = 1, \quad V(S_n) = E(S_n^2) - \sigma_n^2 = 1 - \sigma_n^2, \quad \mathrm{Cov}(\bar{X}', S_n) = E(\bar{X}'S_n).$$

Here, we consider the case when $t = O(\sqrt{n})$. Then we have

$$(5) \quad \sigma_n = 1 - \frac{1}{8n}(\mu_4 - 1) - \frac{1}{128n^2}\left\{8(6\mu_3^2 - \mu_6 + 3\mu_4 + 2) + 15(\mu_4 - 1)^2\right\} + O\left(\frac{1}{n^3}\right),$$

$$(6) \quad E(\bar{X}'S_n) = \frac{\mu_3}{2n} + \frac{1}{16n^2}(9\mu_3 - 2\mu_5 + 3\mu_3\mu_4) + O\left(\frac{1}{n^3}\right).$$

From (3) to (6) we have

$$(7) \quad V_t(Y_n) = 1 - \frac{t\mu_3}{\sqrt{n}} + \frac{t^2}{4n}(\mu_4 - 1) - \frac{t}{8n\sqrt{n}}(9\mu_3 - 2\mu_5 + 3\mu_3\mu_4) + O\left(\frac{1}{n\sqrt{n}}\right).$$

Letting $W_n := Y_n/\sqrt{V_t(Y_n)}$, we obtain $E_t(W_n) = 0$ and $V_t(W_n) = 1$. On the third cumulant of $Y_n$, we have from (5)

$$(8) \quad \kappa_{3,t}(Y_n) = \frac{\mu_3}{\sqrt{n}} - \frac{3t}{4n}\left\{2(\mu_4 - 3) - \mu_3^2\right\} + \frac{3t^2}{4n\sqrt{n}}\left\{\mu_5 - \mu_3(\mu_4 + 5)\right\}$$
$$- \frac{t^3}{16n^2}(1 + 2\mu_6 - 12\mu_3^2 - 3\mu_4^2) + O\left(\frac{t^3}{n^3}\right).$$

Letting $t = c\sqrt{n} + d$ with some constants $c$ and $d$, from (7) and (8) we have

$$(9) \quad V_{c,d}(Y_n) := 1 - c\mu_3 + \frac{c^2}{4}(\mu_4 - 1) + \frac{d}{2\sqrt{n}}\left\{c(\mu_4 - 1) - 2\mu_3\right\} + \frac{d^2}{4n}(\mu_4 - 1)$$
$$- \frac{c}{8n}(9\mu_3 - 2\mu_5 + 3\mu_3\mu_4) + O\left(\frac{1}{n\sqrt{n}}\right),$$

$$(10) \quad \kappa_{3,c,d}(Y_n)$$
$$:= \frac{1}{\sqrt{n}}\left[\mu_3 - \frac{3c}{4}\left\{2(\mu_4 - 3) - \mu_3^2\right\} + \frac{3c^2}{4}\left\{\mu_5 - \mu_3(\mu_4 + 5)\right\} - \frac{c^3}{16}(1 + 2\mu_6 - 12\mu_3^2 - 3\mu_4^2)\right]$$
$$- \frac{3d}{16n}\left[4\left\{2(\mu_4 - 3) - \mu_3^2\right\} - 8c\left\{\mu_5 - \mu_3(\mu_4 + 5)\right\} + c^2(1 + 2\mu_6 - 12\mu_3^2 - 3\mu_4^2)\right]$$
$$+ O\left(\frac{1}{n\sqrt{n}}\right).$$

## 3. A higher order approximation to a percentage point of the distribution of the non-central t-statistic and its application

From (2) we obtain

$$(11) \quad P_\mu\{T_n \le t\} = P_0\left\{Z - t(S_n - \sigma_n) \le a_n(t)\right\} = P_0\{Y_n \le a_n(t)\} = P_0\left\{W_n \le \frac{t\sigma_n - \mu\sqrt{n}}{\sqrt{V_t(Y_n)}}\right\}.$$

Using the Cornish-Fisher expansion, we can obtain a higher order approximation formula of a percentage point of the distribution of $T_n$.

**Theorem 1** The upper $100\alpha$ percentile $t_\alpha$ of the distribution of $T_n$ can be derived from the formula

$$\text{(12)}\quad \frac{t_\alpha \sigma_n - \mu\sqrt{n}}{\sqrt{V_{t_\alpha}(Y_n)}} = u_\alpha + \frac{1}{6}\kappa_{3,t_\alpha}(W_n)(u_\alpha^2 - 1) + O\left(\frac{1}{n}\right),$$

where $u_\alpha$ is the upper $100\alpha$ percentile of the standard normal distribution, $\sigma_n$ and $V_t(Y_n)$ are given by (5) and (7), respectively, and $\kappa_{3,t}(W_n) = \kappa_{3,t}(Y_n)\{V_t(Y_n)\}^{-3/2}$ with (8).

**Corollary** Let $T_n$ be the non-central t-statistic. Then the lower confidence limit $\hat{\delta}$ of the non-centrality parameter $\delta := \mu\sqrt{n}$ of level $1-\alpha$ and the confidence interval $[\underline{\delta}, \bar{\delta}]$ of $\delta$ of level $1-\alpha$ are given by

$$\hat{\delta} = \sigma_n T_n - \sqrt{V_{T_n}(Y_n)}\left\{u_\alpha + \frac{1}{6}\kappa_{3,T_n}(W_n)(u_\alpha^2 - 1)\right\} + O_p\left(\frac{1}{n}\right),$$

$$\underline{\delta} = \sigma_n T_n - \sqrt{V_{T_n}(Y_n)}\left\{u_{\alpha/2} + \frac{1}{6}\kappa_{3,T_n}(W_n)(u_{\alpha/2}^2 - 1)\right\} + O_p\left(\frac{1}{n}\right),$$

$$\bar{\delta} = \sigma_n T_n + \sqrt{V_{T_n}(Y_n)}\left\{u_{\alpha/2} + \frac{1}{6}\kappa_{3,T_n}(W_n)(u_{\alpha/2}^2 - 1)\right\} + O_p\left(\frac{1}{n}\right).$$

Since $t_\alpha$ is the upper $100\alpha$ percentile of the distribution of the non-central t-statistic $T_n$, it follows from (9), (11) and the first order approximation that $t_\alpha = (1/\sigma_n)\{\mu\sqrt{n} + u_\alpha\sqrt{V_{c,d}(Y_n)}\} + o(1)$. From (5) and (9) we have $t_\alpha = \mu\sqrt{n} + u_\alpha\{1 - c\mu_3 + (c^2/4)(\mu_4 - 1)\}^{1/2} + o(1)$. Here, letting $c = \mu$, we obtain $d = \sigma_0 u_\alpha$, where $\sigma_0 := \{1 - \mu\mu_3 + (\mu^2/4)(\mu_4 - 1)\}^{1/2}$. From (9) and (10) we have

$$\begin{aligned}\text{(13)}\quad V_{\mu,\sigma_0 u_\alpha}(Y_n) =& \sigma_0^2 + \frac{\sigma_0 u_\alpha}{2\sqrt{n}}\{\mu(\mu_4 - 1) - 2\mu_3\} \\ &+ \frac{\sigma_0^2 u_\alpha^2}{4n}(\mu_4 - 1) - \frac{\mu}{8n}(9\mu_3 - 2\mu_5 + 3\mu_3\mu_4) + O\left(\frac{1}{n\sqrt{n}}\right), \\ \kappa_{3,\mu,\sigma_0 u_\alpha}(Y_n) =& \frac{1}{\sqrt{n}}\left[\mu_3 - \frac{\mu}{16}\{12(2(\mu_4 - 3) - \mu_3^2) - 12\mu(\mu_5 - \mu_3(\mu_4 + 5))\right. \\ &\left. + \mu^2(1 + 2\mu_6 - 12\mu_3^2 - 3\mu_4^2)\}\right] \\ &- \frac{3\sigma_0 u_\alpha}{16n}\left[4\{2(\mu_4 - 3) - \mu_3^2\} - 8\mu\{\mu_5 - \mu_3(\mu_4 + 5)\}\right. \\ &\left. + \mu^2(1 + 2\mu_6 - 12\mu_3^2 - 3\mu_4^2)\right] + O\left(\frac{1}{n\sqrt{n}}\right) \\ =:& \frac{A}{\sqrt{n}} + \frac{B}{n} + O\left(\frac{1}{n\sqrt{n}}\right) \qquad \text{(say)}.\end{aligned}$$

**Theorem 2** The upper $100\alpha$ percentile $t_\alpha$ of the distribution of the non-central t-statistic $T_n$ is given by

$$t_\alpha = \frac{1}{\sigma_n}\left[\mu\sqrt{n} + \sqrt{V_{\mu,\sigma_0 u_\alpha}(Y_n)}\left\{u_\alpha + \frac{1}{6}\kappa_{3,\mu,\sigma_0 u_\alpha}(W_n)(u_\alpha^2 - 1) + O\left(\frac{1}{n}\right)\right\}\right],$$

where $\sigma_n$ and $V_{\mu,\sigma_0 u_\alpha}(Y_n)$ are given by (5) and (13), respectively and

$$\kappa_{3,\mu,\sigma_0 u_\alpha}(W_n) = \sigma_0^{-3}\left[\frac{A}{\sqrt{n}} - \frac{3Au_\alpha}{4n}\{\mu(\mu_4 - 1) - 2\mu_3\} + \frac{B}{n} + O\left(\frac{1}{n\sqrt{n}}\right)\right]$$

with

$$A = \mu_3 - \frac{\mu}{16}\left[12\left\{2(\mu_4 - 3) - \mu_3^2\right\} - 12\mu\left\{\mu_5 - \mu_3(\mu_4 + 5)\right\} + \mu^2(1 + 2\mu_6 - 12\mu_3^2 - 3\mu_4^2)\right],$$

$$B = -\frac{3\sigma_0 u_\alpha}{16}\left[4\left\{2(\mu_4 - 3) - \mu_3^2\right\} - 8\mu\left\{\mu_5 - \mu_3(\mu_4 + 5)\right\} + \mu^2(1 + 2\mu_6 - 12\mu_3^2 - 3\mu_4^2)\right].$$

The limiting distribution of the non-central t-statistic $T_n$ is given by Bentkus et al. (2007), i.e. the statistic $\sigma_0^{-1}(T_n - \mu\sqrt{n})$ converges in law to $N(0, 1)$ as $n \to \infty$. Then the upper $100\alpha$ percentile $t_\alpha$ of the distribution of $T_n$ is asymptotically given by $t_\alpha = \mu\sqrt{n} + \sigma_0 u_\alpha + o(1)$ as $n \to \infty$, which concides with the first order approximation of (12).

## 4. Conclusion

Without the normality assumption, we derive the higher order approximation to the percentage point of the distribution of a non-central t-statistic, using the Cornish-Fisher expansion, from which the approximate value is easily obtained by a calculator. In the cases of gamma, exponential, two-sided exponential, normal, t-, Weibull, logistic and Birnbaum-Saunders distributions, the higher order approximation is seen to be numerically better than the first order one. The errors of the higher order approximation are almost below 10 percent for a sample of size $n$ bigger than 20 in the above cases. Hence it seems to be useful for symmetric and asymmetric distributions.

## REFERENCES (RÉFERENCES)

Akahira, M. (1995). A higher order approximation to a percentage point of the non-central t-distribution. *Commun. Statist. – Simula.*, **24**(3), 595–605.

Akahira, M., Ohyauchi, N. and Kawai, S. (2009). A higher order approximation to a percentage point of the distribution of a non-central t-statistic without the normality assumption. *Mathematical Research Note* 2009–001, Institute of Mathematics, University of Tsukuba.

Bentkus, V., Jing, B.-Y. and Shao, Q.-M. (2007). Limiting distributions of the non-central t-statistic and their applications to the power of t-tests under non-normality. *Bernoulli*, **13**(2), 346–364.

## RÉSUMÉ (ABSTRACT) — optional

*In a usual case, the distribution of the statistic based on a random sample from the normal distribution with mean 0, under non-zero mean, is called the non-central distribution. the non-central distribution is useful for calculating the power in hypothesis testing, obtaining the confidence interval of a non-centrality parameter. In such a case a percentage point of the non-central distribution plays an important part, but it is not so easy to get the percentage point analytically. So, a higher order approximation for a percentage point of the non-central t-distribution under normality is given by Akahira (1995, Commun. Statist. -Simula., 24, pp. 595-605) and is also shown to be numerically better than others. In this article, without the normality assumption, we obtain a higher order approximation to the percentage point of the distribution of a non-central t-statistic, in a similar way to Akahira (1995) where the statistic based on a linear combination of a normal random variable and a chi-statistic takes an important role. Its application to the confidence limit and the confidence interval for a non-centrality parameter are also given. Further, a numerical comparison of the higher order approximation with the limiting normal distribution is done and the former is shown to be more accurate. As a result of the numerical calculation, the higher order approximation seems to be useful in practical situations, when the size of sample is not so small.*