ASYMPTOTIC FORMULAS FOR SIGNIFICANCE LEVELS OF CERTAIN DISTRIBUTIONS

By Alfred M. Peiser

Cornell University

1. Introduction. The purpose of this paper is to derive asymptotic formulas for the significance levels, or per cent points, of certain well-known statistical distributions.[1] Although we restrict ourselves here to two distributions, those of Chi-Square and of Student's t, it will be apparent that the methods used are applicable to many other distributions as well.

The following results are obtained. Let y_p be the p per cent point of the normal distribution, that is, the distribution defined by

$$\Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}v^2}\, dv, \tag{1.1}$$

so that

$$\Phi(y_p) = 1 - p. \tag{1.2}$$

If $\chi^2_{p,n}$ and $t_{p,n}$ denote the p per cent points of the Chi-Square and Student's t distributions with n degrees of freedom respectively, then

$$\chi^2_{p,n} = n + y_p\sqrt{2n} + \frac{2}{3}(y_p^2 - 1) + \frac{y_p^3 - 7y_p}{9\sqrt{2n}} + o\left(\frac{1}{\sqrt{n}}\right), \tag{1.3}$$

and

$$t_{p,n} = y_p + \frac{y_p^3 + y_p}{4n} + o\left(\frac{1}{n}\right). \tag{1.4}$$

These formulas approximate the true values of $\chi^2_{p,n}$ and $t_{p,n}$ to a high degree of accuracy. Tables of comparative values for several values of p and n are given in Section 4.

We shall need the following theorem due to Cramér [3, p. 81; see also pp. 86–87].

Theorem 1: *Let $X_1, X_2, \cdots$ be a sequence of independent, identically distributed random variables having an absolutely continuous distribution function with mean value zero, dispersion σ and finite fifth absolute moment. Let $H_n(x)$ be the distribution function of $(X_1 + \cdots + X_n)/(\sigma\sqrt{n})$, and let $n^{-\frac{1}{2}(r-2)}\lambda_r$ denote the r-th semi-invariant of $H_n(x)$. Then*

$$\Phi(x) - H_n(x) = \frac{\lambda_3}{3!\sqrt{n}}\Phi^{(3)}(x) - \frac{\lambda_4}{4!\,n}\Phi^{(4)}(x) - \frac{10\lambda_3^2}{6!\,n}\Phi^{(6)}(x) + 0(n^{-3/2}). \tag{1.5}$$

[1] This problem was proposed to the author by J. H. Curtiss.

2. The Chi-Square distribution. A random variable X is said to be distributed according to Chi-Square with n degrees of freedom ($X = \chi_n^2$) if its distribution function is

$$(2.1)\qquad F_n(x) = \begin{cases} \displaystyle\int_0^x \frac{t^{\frac{1}{2}n-1} e^{-\frac{1}{2}t}}{2^{\frac{n}{2}}\Gamma(\frac{n}{2})}\,dt, & x \geqslant 0, \\ 0 \quad , & x < 0. \end{cases}$$

The variable $(\chi_n^2 - n)/\sqrt{2n}$ then has the distribution function

$$(2.2)\qquad H_n(x) = F_n(n + x\sqrt{2n}).$$

If we write

$$(2.3)\qquad \chi_{p,n}^2 = n + y_p\sqrt{2n} + a_n\,,$$

so that

$$(2.4)\qquad F_n(\chi_{p,n}^2) = 1 - p,$$

and let $z_{pn} = y_p + a_n/\sqrt{2n}$, then $H_n(z_{pn}) = 1 - p$, and it follows from (1.1) and (1.2) that

$$(2.5)\quad \Phi(z_{pn}) - H_n(z_{pn}) = \Phi(z_{pn}) - \Phi(y_p) = \frac{1}{\sqrt{2\pi}}\frac{a_n}{\sqrt{2n}}\,e^{-\frac{1}{2}(y_p+\theta a_n/\sqrt{2n})^2},$$

where $0 < \theta < 1$. Then by a theorem of Liapounoff's [3, p. 77],

$$\frac{|\,a_n\,|}{\sqrt{2n}}\,e^{-\frac{1}{2}(y_p+\theta a_n/\sqrt{2n})^2} \leq \frac{K \log n}{\sqrt{n}},$$

where K denotes a constant independent of n. But if $\lim_{n\to\infty} |\,a_n\,|/\sqrt{2n} = \infty$, then $\lim_{n\to\infty} H_n(z_{pn})$ is either 0 or 1. Hence $a_n = o(\sqrt{n})$.

Fisher [1, p. 81] has suggested the use of

$$\chi_{p,n}^2 \doteq \tfrac{1}{2}[y_p + \sqrt{2n-1}]^2.$$

A closer approximation,

$$\chi_{p,n}^2 \doteq n\left[1 - \frac{2}{9n} + y_p\sqrt{\frac{2}{9n}}\right]^3,$$

has been obtained by Wilson and Hilferty [2].. It is interesting to note that, according to (1.3), this last approximation is correct to terms of the zero-th order in n.

We apply Theorem 1 to the variables $X_j = (\chi_1^2 - 1)/\sqrt{2}$, $j = 1, 2, \cdots$. Then $\sigma = 1$, and, by the additive property of the Chi-Square distribution [3, p. 45], $H_n(x)$, the distribution function of the variable $(X_1 + \cdots + X_n)/\sqrt{n}$, is related to $F_n(x)$ by (2.2). Thus, $\lambda_3 = 2\sqrt{2}$. It follows from (1.5) and (2.3) that

$$\lim_{n\to\infty} \sqrt{2n}\,[\Phi(z_{pn}) - H_n(z_{pn})] = \lim_{n\to\infty} \frac{2}{3\sqrt{2\pi}}(z_{pn}^2 - 1)e^{-\frac{1}{2}z_{pn}^2} \tag{2.6}$$
$$= \frac{2}{3\sqrt{2\pi}}(y_p^2 - 1)e^{-\frac{1}{2}y_p^2},$$

since $a_n = o(\sqrt{n})$. Then by (2.5) and (2.6)

$$\lim_{n\to\infty} a_n = \tfrac{2}{3}(y_p^2 - 1).$$

According to (2.3) we may now write

$$\chi^2_{p,2n} = 2n + 2y_p\sqrt{n} + 2r_p + 2b_n,$$

where

$$r_p = \tfrac{1}{3}(y_p^2 - 1), \tag{2.7}$$

and $b_n = o(1)$. A simple change of variables in (2.1) yields

$$F_{2n}(\chi^2_{p,2n}) = \int_{\frac{n+r_p}{\sqrt{n}}}^{y_p+\frac{b_n}{\sqrt{n}}} \frac{\sqrt{n}e^{-n}n^n}{\Gamma(n+1)} e^{-v\sqrt{n}-r_p}\left[1 + \frac{v}{\sqrt{n}} + \frac{r_p}{n}\right]^{n-1} dv. \tag{2.8}$$

If we let

$$J_n = \int_{y_p}^{y_p+\frac{b_n}{\sqrt{n}}} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}v^2}\,dv, \tag{2.9}$$

then

$$nJ_n = \frac{b_n\sqrt{n}}{\sqrt{2\pi}} e^{-\frac{1}{2}(y_p+\delta_n)^2}, \tag{2.10}$$

where $\delta_n = o(1)$. By (1.2) and (2.4),

$$J_n = \Phi\left(y_p + \frac{b_n}{\sqrt{n}}\right) - F_{2n}(\chi^2_{p,2n}). \tag{2.11}$$

Using Stirling's formula for $\Gamma(n+1)$ in (2.8), (2.11) becomes

$$J_n = \int_{y_p+\frac{b_n}{\sqrt{n}}}^{\infty} \frac{e^{-\frac{1}{2}v^2}}{\sqrt{2\pi}}\left[\exp\left\{-\frac{1}{12n} + 0(n^{-3})\right.\right.$$
$$\left.\left. + (n-1)\log\left(1 + \frac{v}{\sqrt{n}} + \frac{r_p}{n}\right) + \frac{v^2}{2} - v\sqrt{n} - r_p\right\} - 1\right] dv$$
$$= \int_{y+\frac{b_n}{\sqrt{n}}}^{\infty} \frac{e^{-\frac{1}{2}v^2}}{\sqrt{2\pi}}[e^{A_n} - 1]\,dv,$$

where $A_n = \frac{1}{n}\left(-\frac{1}{12} - r_p - \frac{r_p^2}{2} + \frac{v^2}{2} + v^2 r_p - \frac{v^4}{4}\right) + \frac{1}{\sqrt{n}}\left(\frac{v^3}{3} - v - vr_p\right) + f_p(v)$, $nf_n(v)$ being dominated by $P(|v|)$, where P is a polynomial in v independent of n, and $f_n(v) = 0(n^{-3/2})$. Then

$$J_n = \int_{y_p+\frac{b_n}{\sqrt{n}}}^{\infty} \frac{e^{-\frac{1}{2}v^2}}{n\sqrt{2\pi}}\left[-\frac{1}{12} - r_p - \frac{r_p^2}{2} + v^2\left(1 + 2r_p + \frac{r_p^2}{2}\right)\right.$$

$$\tag{2.12} \left. -v^4\left(\frac{7}{12} + \frac{r_p}{3}\right) + \frac{v^6}{18}\right] dv + \int_{y_p+\frac{b_n}{\sqrt{n}}}^{\infty} \frac{e^{-\frac{1}{2}v^2}}{\sqrt{2\pi n}}\left[\frac{v^3}{3} - v - vr_p\right] dv$$

$$+ \int_{y_p+\frac{b_n}{\sqrt{n}}}^{\infty} \frac{e^{-\frac{1}{2}v^2}}{\sqrt{2\pi}} g_n(v)\, dv + \int_{y_p+\frac{b_n}{\sqrt{n}}}^{\infty} \frac{e^{-\frac{1}{2}v^2}}{\sqrt{2\pi}}\left(\sum_{j=3}^{\infty} \frac{A_n^j}{j!}\right) dv,$$

where $g_n(v)$ has the same properties given above for $f_n(v)$. If we call these last integrals K_1, K_2, K_3 and K_4 respectively, we see that

$$\tag{2.13} \lim_{n\to\infty} nK_3 = \int_{y_p}^{\infty} \frac{e^{-\frac{1}{2}v^2}}{\sqrt{2\pi}} \lim_{n\to\infty} ng_n(v)\, dv = 0.$$

Also, since A_n^j, $j = 1, 2, \cdots$ is dominated by $P_j(|v|)$, $P_j(v)$ being a polynomial in v independent of n, we see that

$$\sum_{j=3}^{\infty} \int_{y_p+\frac{b_n}{\sqrt{n}}}^{\infty} \frac{e^{-\frac{1}{2}v^2}}{\sqrt{2\pi}} \frac{|A_n|^j}{j!}\, dv \le \sum_{j=3}^{\infty} e^{-\frac{1}{2}\left(y_p+\frac{b_n}{\sqrt{n}}\right)^2} \frac{Q_j\left(y_p + \frac{b_n}{\sqrt{n}}\right)}{j!},$$

where Q_j is a polynomial. Since this last sum converges, we have

$$\tag{2.14} nK_4 = \sum_{j=3}^{\infty} \int_{y_p+\frac{n}{\sqrt{n}}}^{\infty} \frac{e^{-\frac{1}{2}v^2}}{\sqrt{2\pi}} \frac{nA_n^j}{j!}\, dv,$$

and by the uniform convergence of (2.14),

$$\tag{2.15} \lim_{n\to\infty} nK_4 = \sum_{j=3}^{\infty} \int_{y_p}^{\infty} \frac{e^{-\frac{1}{2}v^2}}{j!\sqrt{2\pi}} \lim_{n\to\infty} nA_n^j\, dv = 0,$$

since $A_n^j = 0(n^{-j/2})$.

Integrating by parts we obtain

$$nK_2 = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\left(y_p+\frac{b_n}{\sqrt{n}}\right)^2}\left(\frac{2}{3} y_p b_n + \frac{b_n^2}{3\sqrt{n}}\right),$$

and since $b_n = o(1)$,

$$\tag{2.16} \lim_{n\to\infty} nK_2 = 0.$$

Further integration by parts and the use of (2.7) yields

$$\lim_{n\to\infty} nK_1 = \frac{e^{-\frac{1}{2}y_p^2}}{36\sqrt{2\pi}}(y_p^3 - 7y_p). \tag{2.17}$$

Then, by (2.10), (2.12), (2.13), (2.15), (2.16) and (2.17),

$$\lim_{n\to\infty} b_n\sqrt{n} = \frac{1}{36}(y_p^3 - 7y_p),$$

so that

$$\chi^2_{p,2n} = 2n + 2y_p\sqrt{n} + \frac{2}{3}(y_p^2 - 1) + \frac{y_p^3 - 7y_p}{18\sqrt{n}} + o\left(\frac{1}{\sqrt{n}}\right).$$

Equation (1.3) now follows at once.

3. Student's t. If the random variable Y_n has the distribution function $\Phi(x/\sqrt{n})$, then $t_n = Y_n/\chi_n$ is distributed according to Student's distribution for n degrees of freedom and has the distribution function

$$G_n(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{n\pi}} \frac{\Gamma[\frac{1}{2}(n+1)]}{\Gamma[\frac{1}{2}n]}\left(1 + \frac{t^2}{n}\right)^{-\frac{1}{2}(n+1)} dt.$$

If $\sigma = \sqrt{n/(n-2)}$, the variable t_n/σ then has the distribution function

$$H_n(x) = G_n(x\sigma). \tag{3.1}$$

If we write

$$t_{p,n} = y_p + a_n, \tag{3.2}$$

so that

$$G_n(t_{p,n}) = 1 - p, \tag{3.3}$$

and let $z_{pn} = t_{p,n}/\sigma$, then $H_n(z_{pn}) = 1 - p$, and it follows from (1.1) and (1.2) that

$$\begin{aligned} \Phi(z_{pn}) - H_n(z_{pn}) &= \Phi(z_{pn}) - \Phi(y_p) \\ &= \frac{1}{\sqrt{2\pi}}\left[y_p\left(\frac{1}{\sigma} - 1\right) + \frac{a_n}{\sigma}\right] e^{-\frac{1}{2}\left[y_p + \theta\left(y_p\left(\frac{1}{\sigma}-1\right)+a_n\right)\right]^2}, \end{aligned} \tag{3.4}$$

where $0 < \theta < 1$. Then by Liapounoff's Theorem [3, p. 77],

$$\left| y_p\left(\frac{1}{\sigma} - 1\right) + \frac{a_n}{\sigma}\right| e^{-\frac{1}{2}\left[y_p+\theta\left(y_p\left(\frac{1}{\sigma}-1\right)+a_n\right)\right]^2} \leqslant \frac{K\log n}{\sqrt{n}},$$

where K denotes a constant independent of n. But if $\lim_{n\to\infty} |a_n| = \infty$, then $\lim_{n\to\infty} H_n(z_{pn})$ is either 0 or 1.

Hence $a_n = o(1)$.

We apply Theorem 1 to the variables $X_j = Y_n/\chi_n$, $j = 1, 2, \cdots$. Then $\sigma = \sqrt{n/(n-2)}$, and by the additive property of the normal distribution, $H_n(x)$, the distribution function of $(X_1 + \cdots + X_n)/(\sigma\sqrt{n})$, satisfies the relation (3.1). Thus $\lambda_3 = 0$ and $\lambda_4 = 6n/(n-4)$. It follows from (1.5) and (3.2) that

$$\begin{aligned} \lim_{n\to\infty} n[\Phi(z_{pn}) - H_n(z_{pn})] &= \lim_{n\to\infty} \frac{n}{4(n-4)\sqrt{2\pi}} (z_{pn}^3 - 3z_{pn})e^{-\frac{1}{2}z_{pn}^2} \\ &= \frac{1}{4\sqrt{2\pi}} (y_p^3 - 3y_p)e^{-\frac{1}{2}y_p^2}, \end{aligned} \tag{3.5}$$

since $a_n = o(1)$. By (3.4) and (3.5) we have

$$\lim_{n\to\infty} n\left[y_p\left(\frac{1}{\sigma} - 1\right) + \frac{a_n}{\sigma}\right] = \frac{y_p^3 - 3y_p}{4}.$$

But $\lim_{n\to\infty} n(1-\sigma)/\sigma = -1$, so that

$$\lim_{n\to\infty} na_n = \frac{y_p^3 + y_p}{4}.$$

Hence

$$a_n = \frac{y_p^3 + y_p}{4n} + o\left(\frac{1}{n}\right),$$

and equation (1.4) follows at once.

4. Tables. The following tables compare the true values of $\chi^2_{p,n}$ and $t_{p,n}$ with those obtained from (1.3) and (1.4). The true values [4], [5], (to three decimal places) are shown in *italics*.

TABLE 1

$\chi^2_{p,n}$

n \ p	.01	.05	.1	.5	.9
10	23.253 *23.209*	18.318 *18.307*	15.989 *15.987*	9.333 *9.342*	4.875 *4.865*
30	50.908 *50.892*	43.777 *43.773*	40.257 *40.256*	29.333 *29.336*	20.600 *20.599*
50	76.163 *76.154*	67.507 *67.505*	63.168 *63.167*	49.333 *49.335*	37.689 *37.689*
100	135.811 *135.807*	124.343 *124.342*	118.499 *118.498*	99.333 *99.334*	82.358 *82.358*

TABLE II

$t_{p,n}$

n \ p	.0125	.025	.05	.125	.25
10	2.579 *2.634*	2.197 *2.228*	1.797 *1.813*	1.212 *1.221*	0.700 *0.700*
30	2.354 *2.360*	2.039 *2.042*	1.696 *1.697*	1.171 *1.173*	0.683 *0.683*
60	2.298 *2.299*	2.000 *2.000*	1.670 *1.671*	1.161 *1.162*	0.679 *0.679*
120	2.270 *2.270*	1.980 *1.980*	1.658 *1.658*	1.156 *1.156*	0.677 *0.677*

REFERENCES

[1] R. A. FISHER, *Statistical methods for research workers*, Oliver and Boyd, 1925.

[2] E. B. WILSON and M. M. HILFERTY, *National Academy of Sciences Proc.*, Vol. 17 (1931), p. 687.

[3] H. CRAMÉR, *Random variables and probability distributions*, Cambridge University Press, 1937.

[4] M. MERRINGTON, "Tables of percentage points of the *t* distribution," *Biometrika*, Vol. 32 (April 1942), p. 300.

[5] M. MERRINGTON, "Numerical approximation to the percentage points of the χ^2 distribution," *Biometrika*, Vol. 32 (October 1941), p. 200.