Specifications for AP Statistics Calculations

Daniel D. Warner

July 23, 1999

1 Introduction.

This report provides recommendations on the approximations for the AP Stat functions. The goal of these approximations is to provide robust, efficient approximations which are accurate to 15 digits. We interpret that to mean that over the specified domain of the independent variable the maximum relative error is less than 0.5×10^{-15}. We need to be able to compute the following 16 functions or approximations to them:

1.1 The Normal Probability Density [1, 2].

$$\mathrm{Npd}(x, \sigma, \mu) \equiv \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\left(\frac{x-\mu}{\sigma\sqrt{2}}\right)^2\right). \tag{1}$$

1.2 The Normal Cumulative Probability Distribution [1, 2].

$$\mathrm{Ncd}(a, b, \sigma, \mu) \equiv \frac{1}{\sigma\sqrt{2\pi}} \int_a^b \exp\left(-\left(\frac{x-\mu}{\sigma\sqrt{2}}\right)^2\right) dx. \tag{2}$$

1.3 The Inverse Normal Cumulative Probability Distribution [2].

$\mathrm{InvN}(y, \sigma, \mu) = x$, where x satisfies the equation

$$y = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^x \exp\left(-\left(\frac{x-\mu}{\sigma\sqrt{2}}\right)^2\right) dz. \tag{3}$$

1.4 The Student-*t* Probability Density [1, 2].

$$\mathrm{Tpd}(x, df) \equiv \frac{\Gamma\left(\frac{df+1}{2}\right)}{\Gamma\left(\frac{df}{2}\right)\sqrt{\pi df}} \left(1 + \frac{x^2}{df}\right)^{-\frac{df+1}{2}}. \tag{4}$$

On page 308 of [2] there is a typographical error in the statement of this formula.

1.5 The Student-t Cumulative Probability Distribution [1, 2].

$$\operatorname{Tcd}(a,b,df) \equiv \frac{\Gamma\left(\frac{df+1}{2}\right)}{\Gamma\left(\frac{df}{2}\right)\sqrt{\pi df}} \int_a^b \left(1+\frac{x^2}{df}\right)^{-\frac{df+1}{2}} dx. \tag{5}$$

On page 309 of [2] there is a typographical error in the statement of this formula.

1.6 The Inverse Student-t Cumulative Probability Distribution.

This is not an explicitly defined function in AP Statistics menu, but it is needed to provide the Student-t confidence interval.

$\operatorname{InvT}(y, df) = x$, where x satisfies the equation

$$y = \frac{\Gamma\left(\frac{df+1}{2}\right)}{\Gamma\left(\frac{df}{2}\right)\sqrt{\pi df}} \int_{-\infty}^x \left(1+\frac{t^2}{df}\right)^{-\frac{df+1}{2}} dt. \tag{6}$$

1.7 The χ^2 Density [1, 2].

$$\operatorname{Cpd}(x,df) \equiv \frac{\left(\frac{1}{2}\right)^{\frac{df}{2}}}{\Gamma\left(\frac{df}{2}\right)} x^{\left(\frac{df}{2}-1\right)} \exp\left(-\frac{x}{2}\right) \tag{7}$$

1.8 The χ^2 Cumulative Probability Distribution [1, 2].

$$\operatorname{Ccd}(a,b,df) \equiv \frac{\left(\frac{1}{2}\right)^{\frac{df}{2}}}{\Gamma\left(\frac{df}{2}\right)} \int_a^b x^{\left(\frac{df}{2}-1\right)} \exp\left(-\frac{x}{2}\right) dx. \tag{8}$$

1.9 The F Probability Density [1, 2].

$$\operatorname{Fpd}(x,n,d) \equiv \frac{\Gamma\left(\frac{n+d}{2}\right)}{\Gamma\left(\frac{n}{2}\right)\Gamma\left(\frac{d}{2}\right)} \left(\frac{n}{d}\right)^{\frac{n}{2}} x^{\left(\frac{n}{2}-1\right)} \left(1+\frac{nx}{d}\right)^{-\frac{n+d}{2}} \tag{9}$$

1.10 The F Cumulative Probability Distribution [1, 2].

$$\operatorname{Fcd}(a,b,n,d) \equiv \frac{\Gamma\left(\frac{n+d}{2}\right)}{\Gamma\left(\frac{n}{2}\right)\Gamma\left(\frac{d}{2}\right)} \left(\frac{n}{d}\right)^{\frac{n}{2}} \int_a^b x^{\left(\frac{n}{2}-1\right)} \left(1+\frac{nx}{d}\right)^{-\frac{n+d}{2}} dx. \tag{10}$$

1.11 The Binomial Probability Density [1, 2].

$$\operatorname{Bpd}(x,n,p) \equiv \binom{n}{k} p^x (1-p)^{(n-x)} \tag{11}$$

1.12 The Binomial Cumulative Probability Distribution [1, 2].

$$\text{Bcd}(x, n, p) \equiv \sum_{k=0}^{x} \binom{n}{k} p^k (1-p)^{(n-k)} \tag{12}$$

where $(x = 0, 1, 2, ..., n)$, p is the probability of success, and n is the number of trials.

1.13 The Poisson Probability Density [1, 2].

$$\text{Ppd}(x, \mu) \equiv \frac{e^{(-\mu)} \mu^x}{x!} \tag{13}$$

where $(x = 1, 2, 3, ...)$, and μ is the population mean.

1.14 The Poisson Cumulative Probability Distribution [1, 2].

$$\text{Pcd}(x, \mu) \equiv \sum_{k=0}^{x} \frac{e^{(-\mu)} \mu^k}{k!} \tag{14}$$

where $(x = 1, 2, 3, ...)$, and μ is the population mean.

1.15 The Geometric Probability Density [2].

$$\text{Gpd}(x, p) \equiv p(1-p)^{(x-1)} \tag{15}$$

where $(x = 1, 2, 3, ...)$ and p is the probability of success.

1.16 The Geometric Cumulative Probability Distribution [2].

$$\text{Gcd}(x, p) \equiv \sum_{k=1}^{x} p(1-p)^{(k-1)} \tag{16}$$

where $(x = 1, 2, 3, ...)$ and p is the probability of success.

1.17

A Maple worksheet, DIST.mws, has been developed which contains the definitions for each of these 16 functions. Moreover, these definitions can be computed in Maple to arbitrary precision. This provides the first step, namely the *a priori* calculation to high precision, of the test bed to evaluate approximations to these functions. The worksheet is available at the restricted access ftp site hosted on henrici.math.clemson.edu, and a noninteractive html copy is available at the web site

www.math.clemson.edu/faculty/Warner/Casio0599.

Two additional Maple worksheets, HypTests.mws and Interval.mws, have also been developed to verify the roles of these functions in calculating confidence intervals and the p-values for a variety of hypothesis tests. These are also available on the restricted access ftp site with noninteractive versions on the web site.

2 Preliminary Overview of the Calculations

2.1 The Normal Probability Density.

The Normal density function, defined in Equation 1, can be computed directly from its definition.

2.2 The Normal Cumulative Probability Distribution.

The Normal cumulative distribution in Equation 2 can be computed from the error function [1] by the following relation

$$\mathrm{Ncd}(-\infty, x, \sigma, \mu) = \frac{1}{2}\left(1 + \mathrm{erf}\left(\frac{x-\mu}{\sigma\sqrt{2}}\right)\right) \tag{17}$$

where $\mathrm{erf}(x)$ is the error function [1] defined in Equation 28.

2.3 The Inverse Normal Cumulative Probability Distribution.

The Inverse Normal cumulative distribution in Equation 3 can be computed by first obtaining a rough approximation to the answer and then refining it with a small number of Newton-Raphson iterations. The iterations only require the ability to compute Equations 1 and 2.

2.4 The Student-t Probability Density.

The Student-t density function in Equation 4 can be computed from its definition given a routine for calculating the Gamma function, $\Gamma(x)$, [1] which is defined in Equation 30. Since $\Gamma(1/2) = \sqrt{\pi}$, we can also calculate the density from the equation

$$\mathrm{Tpd}(x, df) = \frac{1}{\sqrt{df}\,B\left(\frac{1}{2}, \frac{df}{2}\right)}\left(1 + \frac{x^2}{df}\right)^{-\frac{df+1}{2}}, \tag{18}$$

where $B(z, w)$ is the Beta function [1] defined by Equation 36.

2.5 The Student-t Cumulative Probability Distribution.

The Student-t cumulative distribution in Equation 5 can be computed from the Incomplete Beta function [1] by the following relation

$$\mathrm{Tcd}(-t, t, df) = \frac{1}{\sqrt{df}\,B(\frac{1}{2}, \frac{df}{2})}\int_{-t}^{t}\left(1 + \frac{x^2}{df}\right)^{-\frac{df+1}{2}} dx = 1 - I_p\left(\frac{df}{2}, \frac{1}{2}\right) \tag{19}$$

where $p = df/(df + t^2)$ and the Incomplete Beta function, $I_x(a, b)$, is defined by Equation 37.

2.6 The Inverse Student-t Cumulative Probability Distribution.

The Inverse Student-t cumulative distribution in Equation 6 can be computed by first obtaining a rough approximation to the answer and then refining it with a small number of Newton-Raphson iterations. The iterations only require the ability to compute Equations 4 and 5.

2.7 The χ^2 Density.

The χ^2 density function in Equation 7 can be computed from its definition given a routine for calculating the Gamma function, $\Gamma(x)$, [1] which is defined in Equation 30.

2.8 The χ^2 Cumulative Probability Distribution.

The χ^2 cumulative distribution in Equation 8 can be computed from the Incomplete Gamma function [1] by the following relation

$$\mathrm{Ccd}(0, x, df) = \frac{1}{\Gamma(\frac{df}{2})} \int_0^{x/2} t^{\frac{df}{2}-1} e^{-t} dt = x^{\frac{df}{2}} \gamma^*(\frac{df}{2}, \frac{x}{2}). \tag{20}$$

where $\gamma^*(a, x)$ is Tricomi's Incomplete Gamma function defined in Equation 34.

2.9 The F Probability Density.

The F density function in Equation 9 can be computed from its definition given a routine for calculating the Beta function, $B(x)$, [1] which is defined in Equation 36. It is also possible to write the F density function in the following form

$$\mathrm{Fpd}(x, n, d) = \frac{n^{\frac{n}{2}} d^{\frac{d}{2}}}{B\left(\frac{n}{2}, \frac{d}{2}\right)} x^{\left(\frac{n}{2}-1\right)} (d + nx)^{-\frac{n+d}{2}} \tag{21}$$

2.10 The F Cumulative Probability Distribution.

The F cumulative distribution in Equation 10 can be computed from the Incomplete Beta function [1] by the following relation

$$\mathrm{Fcd}(0, t, n, d) = 1 - I_x\left(\frac{d}{2}, \frac{n}{2}\right) \tag{22}$$

where $x = d/(d + nt)$ and the Incomplete Beta function, $I_x(a, b)$, is defined by Equation 37.

2.11 The Binomial Probability Density.

The Binomial density function in Equation 11 can be computed from its definition given a routine for calculating the binomial coefficient, $\binom{n}{k}$, where

$$\left(\begin{array}{c} n \\ k \end{array}\right) \equiv \frac{n!}{k!(n-k)!} \tag{23}$$

For modest values of n, the binomial coefficient can be computed with no danger of overflow and with integers for all intermediate results using the following algorithm.

```
m = min(k,(n-k));
y = 1;
for (j=1; j<=m; j++) {
```

```
 y = y*(n-j+1);
 y = y/j;
}
```

Suitable approximations for large values of n are given by

$$\begin{aligned}\begin{pmatrix} n \\ k \end{pmatrix} &= \frac{\Gamma(n+1)}{\Gamma(k+1)\Gamma(n-k+1)} \\ &= \frac{1}{(n+1)\mathrm{B}(k+1, N-k+1)}\end{aligned}$$

Here $\Gamma(z)$ and $B(z, w)$ are the Gamma and Beta functions defined in Equations 30 and 36.

2.12 The Binomial Cumulative Probability Distribution.

The Binomial cumulative distribution, defined in Equation 12, can be computed from the relation [1]

$$\mathrm{Bcd}(x, n, p) = 1 - I_p(x+1, n-x) = I_{1-p}(n-x, x+1) \tag{24}$$

where $I_x(a, b)$ is the Incomplete Beta Function [1] defined in Equation 37.

2.13 The Poisson Probability Density.

The Poisson density function, defined in Equation 13, can be computed directly from its definition.

2.14 The Poisson Cumulative Probability Distribution.

The Poisson cumulative distribution in Equation 14 can be computed directly as

$$\mathrm{Pcd}(x, \mu) = e^{-\mu}\, e_x(\mu) \tag{25}$$

where $(x = 1, 2, 3, ...)$, and $e_x(\mu)$ is the Taylor polynomial of degree x for the exponential function

$$e_x(\mu) \equiv \sum_{k=0}^{x} \frac{\mu^k}{k!} \tag{26}$$

This Taylor polynomial should be computed using Horner's rule. An elementary majorization argument can be used to determine a bound which guarantees when the remainder term will be less than 0.5×10^{-15}. In such a case the answer is simply 1. Specifically, the remainder term is

$$R_x(\mu) = \sum_{k=x+1}^{\infty} \frac{\mu^k}{k!}$$

Hence, for $0 \le \mu < x+2$ we can write

$$\begin{aligned}R_x(\mu) &\le \frac{\mu^{x+1}}{(x+1)!} \sum_{k=0}^{\infty} \left(\frac{\mu}{x+2}\right)^k \\ &= \frac{\mu^{x+1}}{(x+1)!} \left(\frac{1}{1 - \frac{\mu}{x+2}}\right)\end{aligned}$$

In this bound, the first factor is simply the first neglected term, and the second factor is a correction for the rate of growth. The correction term is critical. It can be very large if μ is close to $x+2$.

2.15 The Geometric Probability Density.

The Geometric density function, defined in Equation 15, can be computed directly from its definition.

2.16 The Geometric Cumulative Probability Distribution.

The Geometric cumulative distribution in Equation 16 is simply a finite geometric series and can be computed directly as

$$\mathrm{Gcd}(x,p) = 1 - (1-p)^x. \tag{27}$$

3 Approximations for the Special Functions

3.1 The Error Function

The Error function, $\mathrm{erf}(x)$, [1] is defined by

$$\mathrm{erf}(x) \equiv \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt. \tag{28}$$

The complementary error function, $\mathrm{erfc}(x)$, is defined by

$$\mathrm{erfc}(x) \equiv \frac{2}{\sqrt{\pi}} \int_x^\infty e^{-t^2} dt. \tag{29}$$

The accurate calculation of the complementary error function is necessary in certain regimes in order to avoid loss of significance due to cancellation.

An excellent approximation for calculating the error function and the complementary error function was developed by W. J. Cody in 1969 [4, 5]. A careful Fortran implmentation has been distributed in the specfun package from the Argonne National Laboratory which was developed by W. J. Cody and colleagues. It is widely used and can be downloaded from Netlib [6].

We have tested the the Cody approximations in the Maple worksheet, Erf.mws. The approximations are highly accurate. The maximum relative error in the approximations is less than 10^{-18}.

As illustrated in the Maple worksheet, Erf.mws, the error function approximation can be used to calculate $\mathrm{Ncd}(a, b, \sigma, \mu)$ with a maximum relative error of about 1.2×10^{-18}.

3.2 The Gamma Function

The Gamma function, $\Gamma(z)$, [1] is defined by

$$\Gamma(z) \equiv \int_0^\infty t^{z-1} e^{-t} dt. \tag{30}$$

Two key properties of the Gamma function are captured in the recurrence formula

$$\Gamma(z+1) = z\Gamma(z) \tag{31}$$

and the reflection formula

$$\Gamma(z)\Gamma(-z) = -\frac{\pi}{zsin(\pi z)} \tag{32}$$

3.3 The Incomplete Gamma Function

The Incomplete Gamma function, $\Gamma(a, x)$, [1] is defined by

$$\Gamma(a, x) \equiv \int_x^\infty e^{-t} t^{a-1} dt. \tag{33}$$

Tricomi's incomplete gamma function, $\gamma^*(a, x)$, [1] is defined by

$$\gamma^*(a, x) \equiv \frac{x^{-a}}{\Gamma(a)} \int_0^x e^{-t} t^{a-1} dt. \tag{34}$$

These two functions are related by the formula

$$\Gamma(a, x) + x^a \Gamma(a) \gamma^*(a, x) = \Gamma(a). \tag{35}$$

3.4 The Beta Function

The Beta function, $B(z, w)$, [1] is defined by

$$B(z, w) \equiv \frac{\Gamma(z)\Gamma(w)}{\Gamma(z+w)} \equiv \int_0^1 t^{z-1}(1-t)^{w-1} dt \tag{36}$$

3.5 The Incomplete Beta Function

The Incomplete Beta function, $I_x(a, b)$, [1] is defined by

$$I_x(a, b) \equiv \frac{B_x(a, b)}{B(a, b)} \equiv \frac{1}{B(a, b)} \int_0^x t^{a-1}(1-t)^{b-1} dt \tag{37}$$

	Emergency Condition (B)					
Display Panel (A)	1	2	3	4	5	Average
1	18	31	22	39	15	25
	16	35	27	36	12	
2	13	33	24	35	10	23.5
	15	30	21	38	16	
3	24	42	40	52	28	37.8
	28	46	37	57	24	
Average	19	36.17	28.5	42.83	17.5	28.77

References

[1] Abramowitz, M. and Stegun, I.A., *Handbook of Mathematical Functions,* Applied Mathematics Series, Vol. 55, National Bureau of Standards, 1964. reprinted by Dover Publications, 1968.

[2] Casio Computer Company, *CFX-99850Ga Plus User's Guide*

[3] Cody, W. J. and Hillstrom, K. E., *Chebyshev Approximations for the Natural Logarithm of the Gamma Function,* Math. Comp. 21, 1967, pp. 198-203.

[4] Cody, W. J., *Rational Chebyshev approximations for the error function,* Math. Comp., 1969, pp. 631-638.

[5] Cody, W. J., *An Overview of Software Development for Special Functions,* Lecture Notes in Mathematics, 506, Numerical Analysis Dundee, 1975, G. A. Watson (ed.), Springer Verlag, Berlin, 1976.

[6] Cody, W. J., *FORTRAN programs Argonne National Laboratory* NETLIB/SPECFUN, March 19, 1990.

[7] Erdelyi, A., *Higher Transcendental Functions,* Volumes 1 and 2, McGraw-Hill, 1953.

[8] Hart, John F., Cheney, E. W., Lawson, Charles L., Maehly, Hans J., Mestenyi, Charles K., Rice, John R., Thacher, Jr., Henry G., and Witzgall, Chritoph. *Computer Approximations,* Published by John Wiley Sons, Inc. 1968.

[9] Hillstrom, K. E., *ANL/AMD Program ANLC366S, DGAMMA/DLGAMA,* May, 1969.

[10] Korn, G. E., and Korn, T. M., *Mathematical Handbook for Scientists and Engineers,* McGraw-Hill, 1961.

[11] Luke, Y. L., *The Special Functions and Their Approximations,* Volumes I and II, Academic Press, 1969.

[12] Pearson, E., and Johnson, N., *Tables of the Incomplete Beta Function,* Cambridge University Press, 1968.

Additional References Nuerical Recipes, Gautschi, Lanczos, and the IBM ref from Tokyo