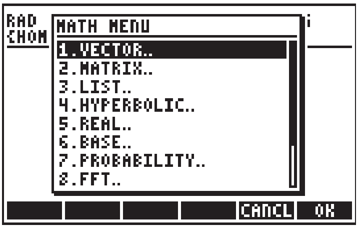
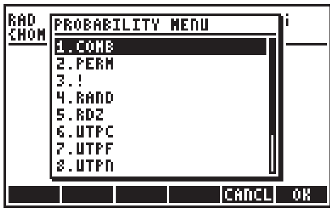
Chapter 17
Probability Applications

In this Chapter we provide examples of applications of calculator's functions to probability distributions.

The MTH/PROBABILITY.. sub-menu - part 1

The MTH/PROBABILITY.. sub-menu is accessible through the keystroke sequence ← MTH . With system flag 117 set to CHOOSE boxes, the following list of MTH options is provided (see left-hand side figure below). We have selected the PROBABILITY.. option (option 7), to show the following functions (see right-hand side figure below):





In this section we discuss functions COMB, PERM, ! (factorial), RAND, and RDZ.

Factorials, combinations, and permutations

The factorial of an integer n is defined as: n! = n· (n-1) · (n-2)…3·2·1. By definition, 0! = 1. Factorials are used in the calculation of the number of permutations and combinations of objects. For example, the number of permutations of r objects from a set of n distinct objects is

$$ {}_nP_r = n(n-1)(n-1)...(n-r+1) = n!/(n-r)! $$

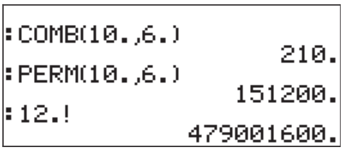
Also, the number of combinations of n objects taken r at a time is

$$ \binom{n}{r} = \frac{n(n-1)(n-2)...(n-r+1)}{r!} = \frac{n!}{r!(n-r)!} $$

To simplify notation, use P(n,r) for permutations, and C(n,r) for combinations. We can calculate combinations, permutations, and factorials with functions COMB, PERM, and ! from the MTH/PROBABILITY.. sub-menu. The operation of those functions is described next:

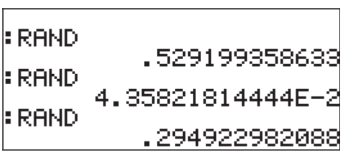
- COMB(n,r): Combinations of n items taken r at a time
- PERM(n,r): Permutations of n items taken r at a time
- n!: Factorial of a positive integer. For a non-integer, x! returns $\Gamma(x+1)$, where $\Gamma(x)$ is the Gamma function (see Chapter 3). The factorial symbol (!) can be entered also as the keystroke combination ALPHA ↱ 2.

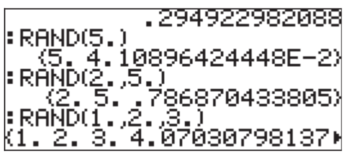
Example of applications of these functions are shown next:



Random numbers

The calculator provides a random number generator that returns a uniformly distributed, random real number between 0 and 1. The generator is able to produce sequences of random numbers. However, after a certain number of times (a very large number indeed), the sequence tends to repeat itself. For that reason, the random number generator is more properly referred to as a pseudo-random number generator. To generate a random number with your calculator use function RAND from the MTH/PROBABILITY sub-menu. The following screen shows a number of random numbers produced using RAND. The numbers in the left-hand side figure are produced with calling function RAND without an argument. If you place an argument list in function RAND, you get back the list of numbers plus an additional random number attached to it as illustrated in the right-hand side figure.





Random number generators, in general, operate by taking a value, called the "seed" of the generator, and performing some mathematical algorithm on that "seed" that generates a new (pseudo)random number. If you want to generate a sequence of number and be able to repeat the same sequence later, you can change the "seed" of the generator by using function RDZ(n), where n is the "seed," before generating the sequence. Random number generators operate by starting with a "seed" number that is transformed into the first random number of the series. The current number then serves as the "seed" for the next number and so on. By "re-seeding" the sequence with the same number you can reproduce the same sequence more than once. For example, try the following:

RDZ(0.25) ENTER	Use 0.25 as the "seed."
RAND() ENTER	First random number = 0.75285…
RAND() ENTER	Second random number = 0.51109…
RAND() ENTER	Third random number= 0.085429….
Re-start the sequence:	
RDZ(0.25) ENTER	Use 0.25 as the "seed."
RAND() ENTER	First random number = 0.75285…
RAND() ENTER	Second random number = 0.51109…
RAND() ENTER	Third random number= 0.085429….

To generate a sequence of random numbers use function SEQ. For example, to generate a list of 5 random numbers you can use, in ALG mode: `SEQ(RAND(),j,1,5,1)`. In RPN mode, use the following program:

« → n « 1 n FOR j RND NEXT n →LIST » »

Store it into variable RLST (Random LiST), and use VAR 5 RLST to produce a list of 5 random numbers.

Function RNDM(n,m) can be used to generate a matrix of n rows and m columns whose elements are random integers between -1 and 1(see Chapter 10).

Discrete probability distributions

A random variable is said to be discrete when it can only take a finite number of values. For example, the number of rainy days in a given location can be considered a discrete random variable because we count them as integer numbers only. Let X represent a discrete random variable, its probability mass function (pmf) is represented by f(x) = P[X=x], i.e., the probability that the random variable X takes the value x.

The mass distribution function must satisfy the conditions that

$$f(x) > 0, \text{ for all } x,$$

and

$$\sum_{all\ x} f(x) = 1.0$$

A cumulative distribution function (cdf) is defined as

$$F(x) = P[X \le x] = \sum_{k \le x} f(k)$$

Next, we will define a number of functions to calculate discrete probability distributions. We suggest that you create a sub-directory, say, HOME\STATS\DFUN (Discrete FUNctions) where we will define the probability mass function and cumulative distribution function for the binomial and Poisson distributions.

Binomial distribution

The probability mass function of the binomial distribution is given by

$$f(n,p,x) = \binom{n}{x} \cdot p^x \cdot (1-p)^{n-x}, \quad x = 0,1,2,...,n$$

where $\binom{n}{x}$ = C(n,x) is the combination of n elements taken x at a time. The values n and p are the parameters of the distribution. The value n represents the number of repetitions of an experiment or observation that can have one of two outcomes, e.g., success and failure. If the random variable X represents the number of successes in the n repetitions, then p represents the probability of getting a success in any given repetition. The cumulative distribution function for the binomial distribution is given by

$$F(n,p,x) = \sum_{k=0}^{x} f(n,p,x), \quad x = 0,1,2,\ldots,n$$

Poisson distribution

The probability mass function of the Poisson distribution is given by

$$f(\lambda, x) = \frac{e^{-\lambda} \cdot \lambda^x}{x!}, \quad x = 0,1,2,\ldots,\infty$$

In this expression, if the random variable X represents the number of occurrences of an event or observation per unit time, length, area, volume, etc., then the parameter l represents the average number of occurrences per unit time, length, area, volume, etc. The cumulative distribution function for the Poisson distribution is given by

$$F(\lambda, x) = \sum_{k=0}^{x} f(\lambda, x), \quad x = 0,1,2,\ldots,\infty$$

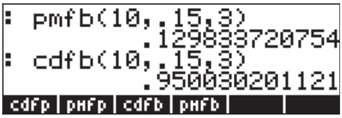
Next, use function DEFINE (← DEF) to define the following probability mass functions (pmf) and cumulative distribution functions (cdf):

```
DEFINE(pmfb(n,p,x) = COMB(n,x)*p^x*(1-p)^(n-x))
DEFINE(cdfb(n,p,x) = Σ(k=0,x,pmfb(n,p,k)))
DEFINE(pmfp(λ,x) = EXP(-λ)*λ^x/x!)
DEFINE(cdfp(λ,x) = Σ(k=0,x,pmfp(λ,x)))
```

The function names stand for:

- pmfb: probability mass function for the binomial distribution
- cdfb: cumulative distribution function for the binomial distribution
- pmfp: probability mass function for the Poisson distribution
- cdfp: cumulative distribution function for the Poisson distribution

Examples of calculations using these functions are shown next:



Continuous probability distributions

The probability distribution for a continuous random variable, X, is characterized by a function f(x) known as the probability density function (pdf). The pdf has the following properties: f(x) > 0, for all x, and

$$P[X < x] = F(x) = \int_{-\infty}^{x} f(\xi)d\xi.$$

$$\int_{-\infty}^{+\infty} f(x)dx = 1.$$

Probabilities are calculated using the cumulative distribution function (cdf), F(x), defined by $P[X < x] = F(x) = \int_{-\infty}^{x} f(\xi)d\xi$, where P[X<x] stands for "the probability that the random variable X is less than the value x".

In this section we describe several continuous probability distributions including the gamma, exponential, beta, and Weibull distributions. These distributions are described in any statistics textbook. Some of these distributions make use of a the Gamma function defined earlier, which is calculated in the calculator by using the factorial function as Γ(x) = (x-1)!, for any real number x.

The gamma distribution

The probability distribution function (pdf) for the gamma distribution is given by

$$f(x) = \frac{1}{\beta^{\alpha}\Gamma(\alpha)} \cdot x^{\alpha-1} \cdot \exp(-\frac{x}{\beta}), for \quad x > 0, \alpha > 0, \beta > 0;$$

The corresponding (cumulative) distribution function (cdf) would be given by an integral that has no closed-form solution.

The exponential distribution

The exponential distribution is the gamma distribution with a = 1. Its pdf is given by

$$f(x) = \frac{1}{\beta} \cdot \exp(-\frac{x}{\beta}), for \quad x > 0, \beta > 0,$$

while its cdf is given by F(x) = 1 - exp(-x/β), for x>0, β >0.

The beta distribution

The pdf for the gamma distribution is given by

$$f(x) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\cdot\Gamma(\beta)} \cdot x^{\alpha-1} \cdot (1-x)^{\beta-1}, for \quad 0 < x < 1, \alpha > 0, \beta > 0$$

As in the case of the gamma distribution, the corresponding cdf for the beta distribution is also given by an integral with no closed-form solution.

The Weibull distribution

The pdf for the Weibull distribution is given by

$$f(x) = \alpha \cdot \beta \cdot x^{\beta-1} \cdot \exp(-\alpha \cdot x^{\beta}), \quad for\ x > 0, \alpha > 0, \beta > 0$$

While the corresponding cdf is given by

$$F(x) = 1 - \exp(-\alpha \cdot x^{\beta}), \quad for\ x > 0, \alpha > 0, \beta > 0$$

Functions for continuous distributions

To define a collection of functions corresponding to the gamma, exponential, beta, and Weibull distributions, first create a sub-directory called CFUN (Continuous FUNctions) and define the following functions (change to Approx mode):

Gamma pdf: `'gpdf(x) = x^(α-1)*EXP(-x/β)/(β^α*GAMMA(α))'`

Gamma cdf: `'gcdf(x) = ∫(0,x,gpdf(t),t)'`

Beta pdf:

`'βpdf(x)= GAMMA(α+β)*x^(α-1)*(1-x)^(β-1)/(GAMMA(α)*GAMMA(β))'`

Beta cdf: `'βcdf(x) = ∫(0,x, βpdf(t),t)'`

Exponential pdf: `'epdf(x) = EXP(-x/β)/β'`

Exponential cdf: `'ecdf(x) = 1 - EXP(-x/β)'`

Weibull pdf: `'Wpdf(x) = α*β*x^(β-1)*EXP(-α*x^β)'`

Weibull cdf: `'Wcdf(x) = 1 - EXP(-α*x^β)'`

Use function DEFINE to define all these functions. Next, enter the values of α and β, e.g., [1] [STO▶] [ALPHA] [→] [A] [ENTER] [2] [STO▶] [ALPHA] [→] [B] [ENTER]

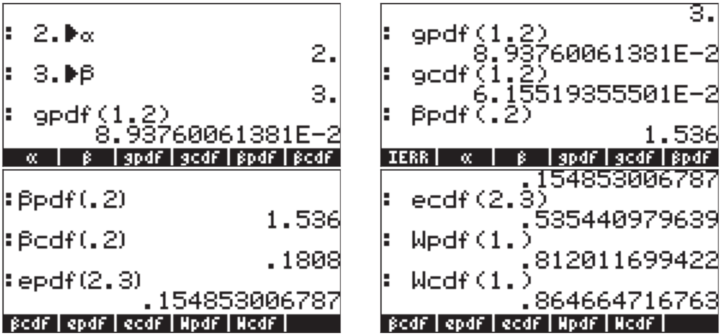
Finally, for the cdf for Gamma and Beta cdf's, you need to edit the program definitions to add →NUM to the programs produced by function DEFINE. For example, the Gamma cdf, i.e., the function gcdf, should be modified to read: « → `x '`→NUM(∫ `(0,x,gpdf(t),t))'` » and stored back into gcdf. Repeat the procedure for βcdf. Use RPN mode to perform these changes.

Unlike the discrete functions defined earlier, the continuous functions defined in this section do not include their parameters (α and/or β) in their definitions. Therefore, you don't need to enter them in the display to calculate the functions. However, those parameters must be previously defined by storing the corresponding values in the variables α and β. Once all functions and the values α and β have been stored, you can order the menu labels by using function ORDER. The call to the function will be the following:

ORDER({'α','β','gpdf','gcdf','βpdf','βcdf','epdf','ecdf','Wpdf','Wcdf'})

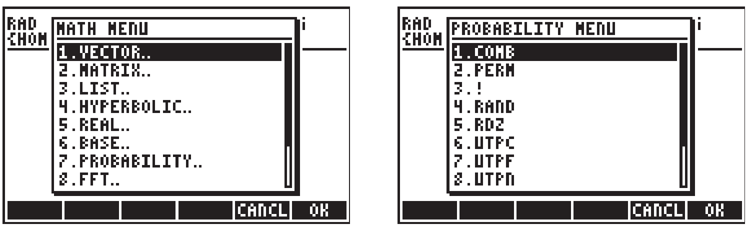
Following this command the menu labels will show as follows (Press [NXT] to move to the second list. Press [NXT] once more to move to the first list):

Some examples of application of these functions, for values of α = 2, β = 3, are shown below. Notice the variable IERR that shows up in the second screen shot. This results from a numerical integration for function gcdf.



Continuous distributions for statistical inference

In this section we discuss four continuous probability distributions that are commonly used for problems related to statistical inference. These distributions are the normal distribution, the Student's t distribution, the Chi-square (χ^2) distribution, and the F-distribution. The functions provided by the calculator to evaluate probabilities for these distributions are contained in the MTH/ PROBABILITY menu introduced earlier in this chapter. The functions are NDIST, UTPN, UTPT, UTPC, and UTPF. Their application is described in the following sections. To see these functions activate the MTH menu: ← MTH and select the PROBABILITY option:



Normal distribution pdf

The expression for the normal distribution pdf is:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}}\exp[-\frac{(x-\mu)^2}{2\sigma^2}],$$

where μ is the mean, and σ^2 is the variance of the distribution. To calculate the value of $f(\mu,\sigma^2,x)$ for the normal distribution, use function NDIST with the following arguments: the mean, μ, the variance, σ^2, and, the value x , i.e., NDIST(μ,σ^2,x). For example, check that for a normal distribution, f(1.0,0.5,2.0) = 0.20755374.

Normal distribution cdf

The calculator has a function UTPN that calculates the upper-tail normal distribution, i.e., UTPN(x) = P(X>x) = 1 - P(X<x). To obtain the value of the upper-tail normal distribution UTPN we need to enter the following values: the mean, μ; the variance, σ^2; and, the value x, e.g., UTPN((μ,σ^2,x)

For example, check that for a normal distribution, with μ = 1.0, σ^2 = 0.5, UTPN(0.75) = 0.638163. Use UTPN(1.0,0.5,0.75) = 0.638163.

Different probability calculations for normal distributions [X is N(μ,σ^2)] can be defined using the function UTPN, as follows:

- P(X<a) = 1 - UTPN(μ, σ^2,a)
- P(a<X<b) = P(X<b) - P(X<a) = 1 - UTPN(μ, σ^2,b) - (1 - UTPN(μ, σ^2,a)) = UTPN(μ, σ^2,a) - UTPN(μ, σ^2,b)
- P(X>c) = UTPN(μ, σ^2,c)

Examples: Using μ = 1.5, and σ^2 = 0.5, find:

P(X<1.0) = 1 - P(X>1.0) = 1 - UTPN(1.5, 0.5, 1.0) = 0.239750.

P(X>2.0) = UTPN(1.5, 0.5, 2.0) = 0.239750.

P(1.0<X<2.0) = F(1.0) - F(2.0) = UTPN(1.5,0.5,1.0) - UTPN(1.5,0.5,2.0) = 0.7602499 - 0.2397500 = 0.524998.

The Student-t distribution

The Student-t, or simply, the t-, distribution has one parameter ν, known as the degrees of freedom of the distribution. The probability distribution function (pdf) is given by

$$f(t) = \frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{\nu}{2}) \cdot \sqrt{\pi\nu}} \cdot (1+\frac{t^2}{\nu})^{-\frac{\nu+1}{2}}, -\infty < t < \infty$$

where $\Gamma(\alpha) = (\alpha\text{-}1)!$ is the GAMMA function defined in Chapter 3.

The calculator provides for values of the upper-tail (cumulative) distribution function for the t-distribution, function UTPT, given the parameter ν and the value of t, i.e., UTPT(ν,t). The definition of this function is, therefore,

$$UTPT(\nu,t) = \int_t^{\infty} f(t)dt = 1 - \int_{-\infty}^{t} f(t)dt = 1 - P(T \le t)$$

For example, UTPT(5,2.5) = 2.7245…E-2. Other probability calculations for the t-distribution can be defined using the function UTPT, as follows:

- P(T<a) = 1 - UTPT(ν,a)
- P(a<T<b) = P(T<b) - P(T<a) = 1 - UTPT(ν,b) - (1 - UTPT(ν,a)) = UTPT(ν,a) - UTPT(ν,b)
- P(T>c) = UTPT(ν,c)

Examples: Given ν = 12, determine:

P(T<0.5) = 1-UTPT(12,0.5) = 0.68694..

P(-0.5<T<0.5) = UTPT(12,-0.5)-UTPT(12,0.5) = 0.3738…

P(T> -1.2) = UTPT(12,-1.2) = 0.8733…

The Chi-square distribution

The Chi-square (χ^2) distribution has one parameter ν, known as the degrees of freedom. The probability distribution function (pdf) is given by

$$f(x) = \frac{1}{2^{\frac{\nu}{2}} \cdot \Gamma(\frac{\nu}{2})} \cdot x^{\frac{\nu}{2}-1} \cdot e^{-\frac{x}{2}}, \nu > 0, x > 0$$

The calculator provides for values of the upper-tail (cumulative) distribution function for the χ^2-distribution using [UTPC] given the value of x and the parameter ν. The definition of this function is, therefore,

$$UTPC(\nu, x) = \int_t^{\infty} f(x)dx = 1 - \int_{-\infty}^{t} f(x)dx = 1 - P(X \le x)$$

To use this function, we need the degrees of freedom, ν, and the value of the chi-square variable, x, i.e., UTPC(ν,x). For example, UTPC(5, 2.5) = 0.776495…

Different probability calculations for the Chi-squared distribution can be defined using the function UTPC, as follows:

- P(X<a) = 1 - UTPC(ν,a)
- P(a<X<b) = P(X<b) - P(X<a) = 1 - UTPC(ν,b) - (1 - UTPC(ν,a)) = UTPC(ν,a) - UTPC(ν,b)
- P(X>c) = UTPC(ν,c)

Examples: Given ν = 6, determine:

P(X<5.32) = 1-UTPC(6,5.32) = 0.4965..

P(1.2<X<10.5) = UTPC(6,1.2)-UTPC(6,10.5) = 0.8717…

P(X> 20) = UTPC(6,20) = 2.769..E-3

The F distribution

The F distribution has two parameters νN = numerator degrees of freedom, and νD = denominator degrees of freedom. The probability distribution function (pdf) is given by

$$f(x) = \frac{\Gamma(\frac{\nu N + \nu D}{2}) \cdot (\frac{\nu N}{\nu D})^{\frac{\nu N}{2}} \cdot F^{\frac{\nu N}{2} - 1}}{\Gamma(\frac{\nu N}{2}) \cdot \Gamma(\frac{\nu D}{2}) \cdot (1 - \frac{\nu N \cdot F}{\nu D})^{(\frac{\nu N + \nu D}{2})}}$$

The calculator provides for values of the upper-tail (cumulative) distribution function for the F distribution, function UTPF, given the parameters νN and νD, and the value of F. The definition of this function is, therefore,

$$UTPF(\nu N,\nu D,F) = \int_t^{\infty} f(F)dF = 1-\int_{-\infty}^{t} f(F)dF = 1-P(\Im \leq F)$$

For example, to calculate UTPF(10,5, 2.5) = 0.161834…

Different probability calculations for the F distribution can be defined using the function UTPF, as follows:

- P(F<a) = 1 - UTPF(νN, νD,a)
- P(a<F<b) = P(F<b) - P(F<a) = 1 -UTPF(νN, νD,b)- (1 - UTPF(νN, νD,a))
  = UTPF(νN, νD,a) - UTPF(νN, νD,b)
- P(F>c) = UTPF(νN, νD,a)

Example: Given νN = 10, νD = 5, find:

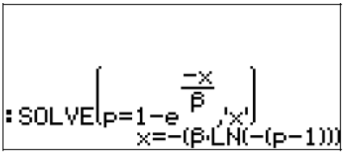
P(F<2) = 1-UTPF(10,5,2) = 0.7700…
P(5<F<10) = UTPF(10,5,5) – UTPF(10,5,10) = 3.4693..E-2
P(F>5) = UTPF(10,5,5) = 4.4808..E-2

Inverse cumulative distribution functions

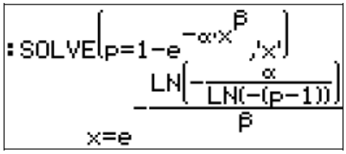
For a continuous random variable X with cumulative density function (cdf) F(x) = P(X<x) = p, to calculate the inverse cumulative distribution function we need to find the value of x, such that x = $F^{-1}(p)$. This value is relatively simple to find for the cases of the exponential and Weibull distributions since their cdf's have a closed form expression:

- Exponential, F(x) = 1 - exp(-x/β)
- Weibull, F(x) = 1-exp(-αx^{β})

(Before continuing, make sure to purge variables α and β). To find the inverse cdf's for these two distributions we need just solve for x from these expressions, i.e., Exponential:



Weibull:



For the <u>Gamma and Beta distributions</u> the expressions to solve will be more complicated due to the presence of integrals, i.e.,

- Gamma, $p = \int_0^x \frac{1}{\beta^\alpha \Gamma(\alpha)} \cdot z^{\alpha-1} \cdot \exp(-\frac{z}{\beta}) dz$
- Beta, $p = \int_0^x \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\cdot\Gamma(\beta)} \cdot z^{\alpha-1} \cdot (1-z)^{\beta-1} dz$

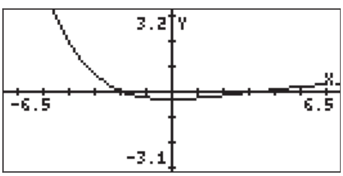
A numerical solution with the numerical solver will not be feasible because of the integral sign involved in the expression. However, a graphical solution is possible. Details on how to find the root of a graph are presented in Chapter 12. To ensure numerical results, change the CAS setting to Approx. The function to plot for the Gamma distribution is

Y(X) = ∫(0,X,z^(α-1)*exp(-z/β)/(β^α*GAMMA(α)),z)-p

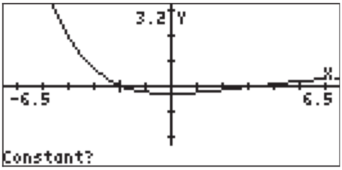
For the Beta distribution, the function to plot is

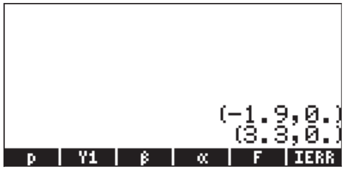
Y(X) =
∫(0,X,z^(α-1)*(1-z)^(β-1)*GAMMA(α+β)/(GAMMA(α)*GAMMA(β)),z)-p

To produce the plot, it is necessary to store values of α, β, and p, before attempting the plot. For example, for α = 2, β = 3, and p = 0.3, the plot of Y(X) for the Gamma distribution is shown below. (Please notice that, because of the complicated nature of function Y(X), it will take some time before the graph is produced. Be patient.)

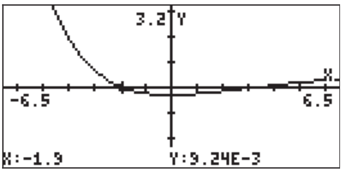


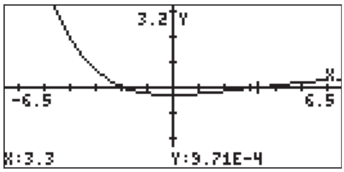
There are two roots of this function found by using function ROOT within the plot environment. Because of the integral in the equation, the root is approximated and will not be shown in the plot screen. You will only get the message Constant? Shown in the screen. However, if you press ENTER at this point, the approximate root will be listed in the display. Two roots are shown in the right-hand figure below.



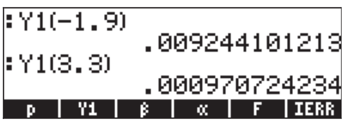


Alternatively, you can use function TRACE (X,Y) to estimate the roots by tracing the curve near its intercepts with the x-axis. Two estimates are shown below:





These estimates suggest solutions x = -1.9 and x = 3.3. You can verify these "solutions" by evaluating function Y1(X) for X = -1.9 and X = 3.3, i.e.,

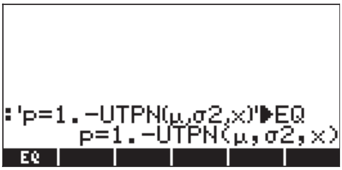


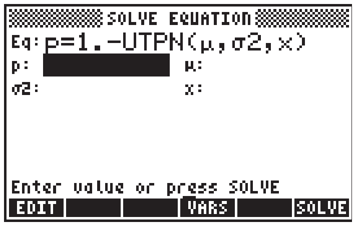
For the normal, Student's t, Chi-square (χ^2), and F distributions, which are represented by functions UTPN, UTPT, UPTC, and UTPF in the calculator, the inverse cuff can be found by solving one of the following equations:

- Normal, p = 1 – UTPN(μ,σ2,x)
- Student's t, p = 1 – UTPT(ν,t)
- Chi-square, p = 1 – UTPC(ν,x)
- F distribution: p = 1 – UTPF(νN,νD,F)

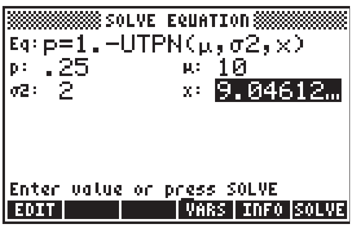
Notice that the second parameter in the UTPN function is σ2, not σ^2, representing the variance of the distribution. Also, the symbol ν (the lower-case Greek letter no) is not available in the calculator. You can use, for example, γ (gamma) instead of ν. The letter γ is available thought the character set (→ CHARS).

For example, to obtain the value of x for a normal distribution, with μ = 10, σ^2 = 2, with p = 0.25, store the equation 'p=1-UTPN(μ, σ2, x)' into variable EQ (figure in the left-hand side below). Then, launch the numerical solver, to get the input form in the right-hand side figure:



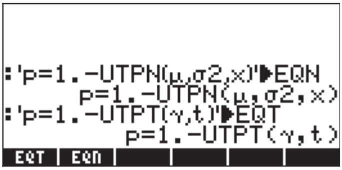


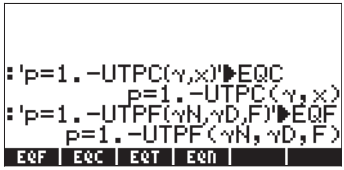
The next step is to enter the values of μ, σ^2, and p, and solve for x:



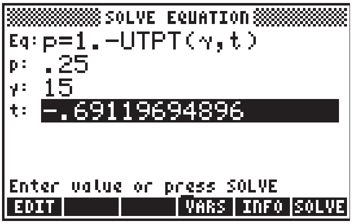
This input form can be used to solve for any of the four variables involved in the equation for the normal distribution.

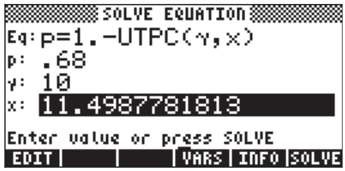
To facilitate solution of equations involving functions UTPN, UTPT, UTPC, and UTPF, you may want to create a sub-directory UTPEQ were you will store the equations listed above:

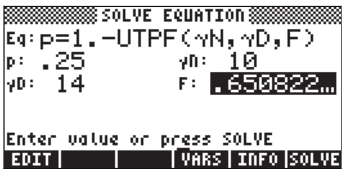




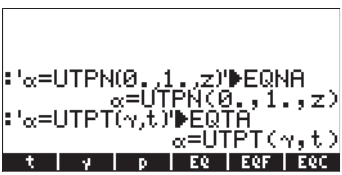
Thus, at this point, you will have the four equations available for solution. You needs just load one of the equations into the EQ field in the numerical solver and proceed with solving for one of the variables. Examples of the UTPT, UTPC, and UPTF are shown below:

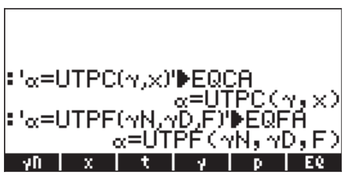




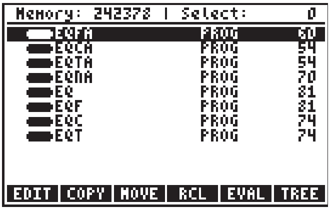


Notice that in all the examples shown above, we are working with p = P(X<x). In many statistical inference problems we will actually try to find the value of x for which P(X>x) = α. Furthermore, for the normal distribution, we most likely will be working with the standard normal distribution in which μ =0, and σ^2 = 1. The standard normal variable is typically referred to as Z, so that the problem to solve will be P(Z>z) = α. For these cases of statistical inference problems, we could store the following equations:





With these four equations, whenever you launch the numerical solver you have the following choices:



Examples of solution of equations EQNA, EQTA, EQCA, and EQFA are shown below: