Hypergeometric Functions and CDFs in J

By Ewart Shaw

Introduction

A valuable but little-mentioned feature of the J language is the conjunction `H.` with which *hypergeometric series* and hence many important mathematical functions can be constructed. A hypergeometric series has the general form

$$ {}_pF_q(a_1 \ldots a_p ; b_1 \ldots b_q ; z) = \sum_{n=0}^{\infty} \frac{(a_1)_n \cdots (a_p)_n}{(b_1)_n \cdots (b_q)_n} \frac{z^n}{n!} , $$

where $(a)_n = a(a+1)\ldots(a+n-1)$ for n>1, and $(a)_0=1$. The equivalent J representation of ${}_pF_q(a_1 \ldots a_p ; b_1 \ldots b_q ; z)$ is $(a_1 \ldots a_p$ `H.` $b_1 \ldots b_q)\ z$.

For example, ${}_1F_1(1; 1; \bullet)$ reduces to the exponential function, as shown in J by:

```
   17j14 ": ^ i.3   NB. exp(0), exp(1), exp(2) to 15 digits
 1.00000000000000 2.71828182845905 7.38905609893065
   17j14 ": 1 H. 1 i.3   NB. or simply  '' H.'' i.3
 1.00000000000000 2.71828182845905 7.38905609893065
```

This article describes how the cumulative distribution functions (CDFs) of many common probability distributions can be expressed via hypergeometric series, making it easy to produce statistical tables and (for example) tail areas arising from t, chi-squared and F tests. Other examples using the hypergeometric conjunction `H.`are given in Iverson (1995), with particular reference to Graham, Knuth & Patashnik (1988) and Abramowitz & Stegun (1970). Another useful source for hypergeometric identities is Spanier & Oldham (1987).

The error function and the Normal CDF

The standard Normal distribution N(0,1), with mean 0 and variance 1, has a CDF $\Phi(\bullet)$ easily expressed in terms of the *error function* erf($\bullet$):

```
   erf=: (*&(%:4p_1) % ^@:*:) * [: 1 H. 1.5 *:
   n01cdf=: [: -: 1: + [: erf %&(%:2)   NB. CDF of N(0,1)
```

On a PC, the absolute error of the resulting approximation is 10^{-14} or less, though the relative error might be judged unacceptable more than (say) seven standard deviations from the mean.

```
   n01cdf _15 _8 _7 0 1.96 7   NB. accurate in range (-7,7)
8.88178e_16 2.16493e_15 1.28109e_12 0.5 0.975002 1
   ncdf=: n01cdf : ([: n01cdf (] - {.@[) % %:@{:@[)
   5 4 ncdf 1               NB. Pr(X<1) where X~N(5,4)
0.0227501
```

For extreme values of z, the tail areas Φ(z) or 1-Φ(z) are well approximated using continued fractions, see Abramowitz & Stegun. A simple such approximation is:

```
   n01pdf=: (%: 0.5p_1)"_ * [: ^ *: % _2:
   n01cdf2=: n01pdf * (1: - [: % *:+3:) % ]
   n01cdf2 _20 _8 _7   NB. testing
_2.75362e_89 _6.22108e_16 _1.27986e_12
```

Note that other hypergeometric series formulae for the error function exist but suffer serious convergence problems. For example:

```
   erfbad=: *&(%:4p_1) * [: 0.5 H. 1.5 -@:*:
   14j11 ": erfbad 0.5 + 2*i.4
0.52049987781 0.99959304798 0.99999999883 3.53955177465
   14j11 ": erf 0.5 + 2*i.4
0.52049987781 0.99959304798 0.99999999980 1.00000000000
```

The incomplete gamma and beta functions

The CDFs of gamma (including chi-squared) and Poisson distributions can be obtained from the *incomplete gamma function*, and CDFs of beta, t, F and binomial distributions from the *incomplete beta function.*

```
   gamma=: ! & <:            NB. gamma function
   ig0=: 4 : '(1 H. (1+x.) % x.&((* ^) * (^ -)~)) y.'
   incgam=: ig0 % gamma@[     NB. incomplete gamma
   beta=:  *&gamma % gamma@+   NB. beta function
   ib0=: 4 : '(((,-.)/x.) H. (1+{.x.) * (^ % ])&({.x.)) y.'
   incbet=: ib0 % [: beta/ [   NB. incomplete beta

   gamma 5 6 7 1.5           NB. testing gamma...
24 120 720 0.886227
   6 ig0 0 5 10 20 30        NB. unnormalised
```

```
0 46.0847 111.95 119.991 120
   6 incgam 0 5 10 20 30
0 0.384039 0.932914 0.999928 1
   4 beta 2 3 4 1.5           NB. testing beta...
0.05 0.0166667 0.00714286 0.101587
   4 4 2 4 beta 2 3 4 1.5
0.05 0.0166667 0.05 0.101587
   4 2 ib0 0 0.1 0.6 0.7 1     NB. unnormalised
0 2.3e_5 0.016848 0.026411 0.05
   4 2 incbet 0 0.1 0.6 0.7 1
0 0.00046 0.33696 0.52822 1
```

The interrelationships between these functions and various CDFs, illustrated below, are conveniently summarised in Press *et al*. (1986). As with the Normal distribution, extreme tail areas would require alternative methods.

```
   NB. (n,p) bincdf y.   binomial(n,p)
   NB. n chisqcdf y.     chi-squared on n d.f.
   NB. (n1,n2) fcdf y.   F on (n1,n2) d.f.
   NB. mu poissoncdf y.  Poisson, mean mu
   NB. n tcdf y.         t on n d.f.

   bincdf=: (({.@[ - ]) ,. >:@]) incbet"1 _ -.@{:@[
   chisqcdf=: incgam&-:
   fcdf=: -:@[ incbet ({. % +/)@(* ,:&1)
   poissoncdf=: -.@incgam"0~ >:
   tcdf=: [: -:@>: *@] * 1&,@[ fcdf *:@]

   binpmf=: (!/ {.)~ * [: */ (, -.)@{:@[ ^ ] ,: {.@[ - ]
   +/\ 5 0.2 binpmf i.6     NB. binomial CDF 'by hand'
0.32768 0.73728 0.94208 0.99328 0.99968 1
   5 0.2 bincdf i.6         NB. _%_ for y.=n
0.32768 0.73728 0.94208 0.99328 0.99968 _.
   5 chisqcdf 0.831211 4.35146 11.0705 20.515
0.025 0.5 0.95 0.999
   3 10 fcdf 0.84508 3.70826 6.55231 12.5527
0.5 0.95 0.99 0.999
   poissonpmf=: ^ * ^@-@[ % !@]
   +/\ 2.3 poissonpmf i.5   NB. Poisson CDF 'by hand'
0.100259 0.330854 0.596039 0.799347 0.916249
   2.3 poissoncdf i.5
0.100259 0.330854 0.596039 0.799347 0.916249
```

```
   30 tcdf _1 0 1 2.04227 4
0.162654 0.5 0.837346 0.975 0.999809
   NB. table of t CDFs (using table adverb from stdlib)
   1 2 6 10 20 30 tcdf"0 _ table i.6
+--+------------------------------------------------+
|  |  0      1       2       3       4       5|
+--+------------------------------------------------+
| 1|0.5     0.75 0.852416 0.897584 0.922021 0.937167|
| 2|0.5 0.788675 0.908248 0.952267 0.971405 0.981125|
| 6|0.5 0.822041 0.953787 0.987996 0.996441 0.998774|
|10|0.5 0.829553 0.963306 0.993328 0.998741 0.999731|
|20|0.5 0.835372 0.970367 0.996462 0.999648 0.999966|
|30|0.5 0.837346 0.972687 0.997305 0.999809 0.999988|
+--+------------------------------------------------+
```

Implementation and summary

Classes of probability distributions, including their CDFs, pseudorandom variable generation etc., can be created using J's object-oriented capabilities. Implementation details involve splicing together different approximations over different ranges, and iteration to approximate inverse CDFs such as $\Phi^{-1}(\bullet)$, but such details can mercifully be hidden from the user. The hypergeometric conjunction H. makes approximations over the regions of most interest easy in J. Hypergeometric series are also not difficult to implement in traditional APL and other languages, and the above methods deserve to be more widely known and used.

References

Abramowitz, M. and Stegun, I.A. (1970). *Handbook of Mathematical Functions*, Dover, New York.

Graham, R.L., Knuth, D.E., and Patashnik, O. (1988). *Concrete Mathematics*, Addison-Wesley, Reading, Massachusetts.

Iverson, K.E. (1995). *Concrete Math Companion*, Iverson Software Inc., Toronto.

Press, W.H., Flannery, B.P., Teukolsky, S.A., and Vetterling, W.T. (1986). *Numerical Recipes*, Cambridge University Press, Cambridge.

Spanier, J. and Oldham, K.B. (1987). *An Atlas of Functions*, Springer-Verlag, Berlin.