TILASTOLLISET MENETELMÄT TÄHTITIETEESSÄ
IDL-HARJOITUS I, Heikki Salo 21.2.2008

1) Binomijakauma

1.1) Tee IDL-ohjelma joka plottaa binomijakauman

$$\begin{aligned} \mathrm{prob}(V) &= \binom{N}{V} p^V (1-p)^{N-V} \qquad V = 0, 1, ..., N \\ \binom{N}{V} &= \frac{N!}{V!(N-V)!} \end{aligned} \tag{1}$$

arvoille $N = 10, p = 0.5$, käyttäen suoraan kertomien avulla laskettuja binomitekijöitä. Lisää samaan kuvaan $p = 0.1$ ja $p = 0.9$.

1.2) Tee sama arvoille $p = 0.5$, $N = 10, 20, 40, 80, 160$ Käytä hyväksi relaatiota

$$\begin{aligned} \log \frac{N!}{v!(N-v)!} &= \log N! - \log n! - \log(N-n)! \\ &= \log\Gamma(N+1) - \log\Gamma(v+1) - \log\Gamma(n-v+1) \end{aligned}$$

IDL-funktio `lngamma(x)` palauttaa lausekkeen $\log\Gamma(x)$

1.3) Laske keskeisen raja-arvolauseen perusteella tn V=1500 kun N=3000, p=0.5 (vastaus 0.014566503)

2. Poisson jakauma
- tuota luentojen sivulla 31 oleva kuva
- Osoita että binimijakauma lähestyy Poisson jakaumaa kun $N- > \infty$ ja nP äärellinen (luentojen sivun 29 kuva).

Useful statistical procedures in IDL:

```
--------------------------------------------
PROBABILITY FUNCTIONS
--------------------------------------------

BINOMIAL-function
----------------------
Cumulative binomial probability distribution

Y=Binomial(V,N,P)

    V= minimum number of event occurring in
    N independent trials.
    P=probability in single event


GAUSS_PDF(V) function:
----------------------
  gives the probability Prob(x le V)
  for a Gaussian probability distribution
  with a mean of 0.0 and a variance of 1.0,

IDL> print,gauss_pdf(1)
     0.841345
IDL> print,gauss_pdf(2)
     0.977250
IDL> print,gauss_pdf(3)
     0.998650

GAUSS_CVF(P) function:
----------------------
  gives the cut-off value V so that Prob(x gt V)=P
  for a Gaussian probability  distribution
  with a mean of 0.0 and a variance of 1.0,

IDL> print,gauss_cvf(0.1)
      1.28155
IDL> print,gauss_cvf(0.01)
      2.32635
IDL> print,gauss_cvf(0.001)
      3.09025
```

```
CHISQR_PDF(V,df):
----------------------
  gives the probability Prob(x le V)
  for a Chi-square distribution with DF degrees of freedom.

CHISQR_CVF(P,df)
----------------------
  gives the cut-off value V so that Prob(x gt V)=P
  for a Chi-square distribution with DF degrees of freedom.


T_PDF(V,df):
----------------------
  gives the probability Prob(x le V)
  for a Student's t distribution with DF degrees of freedom.

T_CVF(P,df)
----------------------
  gives the cut-off value V so that Prob(x gt V)=P
  for a Student's t distribution with DF degrees of freedom.


F_PDF(V, DFn, DFd):
----------------------
  gives the probability Prob(x le V)
  for an F distribution with DFn,DFd degrees of freedom.

F_CVF(P, DFn, DFd)
----------------------
  gives the cut-off value V so that Prob(x gt V)=P
  for an F distribution with DFn,DFd degrees of freedom.
```

```
------------------------------------------
CALCULATION OF SAMPLE STATISTICS
------------------------------------------

  MEAN(X)             mean
  MEDIAN(X)           median
  MIN(X)              smallest element
  MAX(X)              largest  element
  STDDEV(X)           standard deviation (sqrt of sample variance)
  MEANABSDEV(X)       mean absolute deviation
  SKEWNESS(X)         skewness
  KURTOSIS(X)         kurtosis

MOMENT-function
----------------
RESULT = MOMENT(X, MDEV=variable, SDEV=variable)
  X=vector of sample values
  Result=[mean,sample variance, skewness, kurtosis]
  MDEV = mean absolute deviation
  SDEV = standard deviation = sqrt of sample variance

SORT-function
-------------
INDEX=SORT(X)
  returns indices so that X(index) in ascending order

-----------------------------------------
DISPLAYING DISTRIBUTIONS
-----------------------------------------

HISTOGRAM-function
-------------------

Y=HISTOGRAM(X)
  Y(i)=number of array elements X(j) with value i

Y=HISTOGRAM(X, BINSIZE=value,MIN=value,MAX=value)
  Y(I)=number of array elements in the specified
  (MAX-MIN)/BINSIZE intervals


HISTO_F-procedure (Auxillary procedure written in IDL)
```

```
------------------------------------------------
pro histo_f,x,x1,x2,dx,xx,yy,gg
  x=input values
  histogram from x1 to x2 with step dx
  xx,yy return calculated values
  gg returns corresponding gaussian fit (if /gauss)
    /plot      -> plot data
    /oplot     -> oplot data
    /auto      -> automatic scaling with 50 part/bin
    auto=nbins -> automatic scaling with nbin bins
    /noscale   -> do not scale distribution (default: normalized area)
    color=col  -> plot with given color
    psym =sym  -> plot with given symbol type (try psym=10)
```

```
--------------------------------------------
GENERATTION OF RANDOM NUMBERS
--------------------------------------------

----------------------
* Uniform distribution
-----------------------

The RANDOMU function returns one or more uniformly-distributed,
floating-point, pseudo-random numbers in the range 0 < Y < 1.0.

  N=1000
  Y=RANDOMU(SEED,N)   ;N=1000 uniform random numbers

If the variable SEED in undefined before the call, IDL initializes
the random sequence by a value formed from time and date of call.
After the call SEED is defined, and is used at subsequent calls to
continue the random sequence.

If SEED is set to scalar value before the first call, yields a fixed
set of random numbers.


----------------------
* Normal distribution:
-----------------------

  N=1000
  Y=RANDOMN(seed, N)    ;N=1000 random numbers following Gaussian distribution
         ;with zero mean and unit variance

To create Gaussian numbers with mean MU and variance VARI, use
  Y=sqrt(VARI)*RANDOMN(seed,N)+MU

-----------------------
* Other distributions:
-----------------------

Either RANDOMU or RANDOMN accepts keywords:

BINOMIAL=[n,p] -->
  generate random numbers from a binomial distribution.
```

```
  If an event occurs with probability p, with n trials, then the
  number of times it occurs has a binomial distribution.

  N=1000
  Y=RANDOMN(seed,N,binomial=[10,.25])

POISSON=A  -->
  generate random numbers from a Poisson distribution with the mean A
  (mean number of rare events occurring during a unit time)

GAMMA=i    -->
  random number from gamma-distribution = waiting time to the i-th event
  in Poisson random process of unit mean.
  i=1 corresponds to exponential distribution
```