SAS Program Notes
Biostatistics: A Guide to Design, Analysis, and Discovery
Chapter 5: Probability Distributions

Note 5.1 – Finding binomial and Poisson probabilities

1. Finding binomial probabilities:

In SAS, the command **PROBBNML(n,k,p)** can be used to find the cumulative probability (cdf) of *k or less* successes in *n* trials when the probability of success is *p*. Assuming n=4 and p= 0.25, the cumulative probability for k≤1, also expressed as Prob(X≤1) = Prob(X=0) + Prob(X=1), can be found by using the SAS commands below.

SAS commands:

```
DATA BINOMIAL;
PROB = PROBBNML(0.25,4,1);
PROC PRINT;
RUN;
```

SAS output:

```
                    The SAS System

Obs      PROB

 1      0.73828
```

The probability for k=1, also expressed as Prob(X=1) = Prob(X≤1)-Prob(X≤0), can be found using the SAS commands below.

SAS commands:

```
DATA BINOMIAL;
PROB = PROBBNML(0.25,4,1)-PROBBNML(0.25,4,0);
PROC PRINT;
RUN;
```

SAS output:

```
                                The SAS System

Obs      PROB

 1      0.42188
```

To illustrate further, we will use **PROBBNML(n,k,p)** to obtain values of the binomial cumulative distribution function, and by subtracting the cdf evaluated at x from the cdf evaluated at x+1, we will obtain the pdf of the binomial distribution. These calculations are performed using the SAS commands below. We show the calculations for *n* equal to 4 and *p* equal to 0.25. Notice that the values in the output following the SAS commands below are found in Figure 5.1.

SAS commands:

```
DATA BNMLPDF;
  ARRAY P(5) P1-P5;
  P(1) = PROBBNML(.25,4,0);
  DO X = 0 TO 3;
    P(X+2) = PROBBNML(.25,4,X+1) - PROBBNML(.25,4,X);
  END;
PROC PRINT;
  VAR P1-P5;
RUN;
```

SAS output:

```
                                The SAS System

Obs       P1         P2         P3         P4          P5

 1     0.31641    0.42188    0.21094    0.046875    .00390625
```

Let's assume that we wish to obtain the values of binomial variables in 10 samples from a binomial distribution with a sample size of 4 and a population proportion of 0.25. RANBIN generates random numbers from a binomial distribution and stores the value in X. Instead of using PROC PRINT to print the results, we can also use a PUT statement to print the results. The PUT statement prints the results in the LOG window, not in the OUTPUT window.

SAS commands:

```
DATA BINOM;
  RETAIN SEED;
  SEED=5;
```

```
  DO I = 1 TO 10;
  CALL RANBIN(SEED,4,0.25,X);
  PUT ' The value of a binomial variable from B(4,.25) is  ' X;
  END;
RUN;
```

2. Finding Poisson probabilities:

To calculate Poisson probabilities, we can use the SAS command **POISSON(μ,k)** to find the cumulative probability (cdf) of *k or less* events. For example in Table 5.3, we display Poisson probabilities for μ = 1 and μ = 2. We can reproduce the first four values of Table 5.3 where μ = 1 using the SAS commands below.

SAS commands:

```
DATA POISSON;
*Prob(X=0);
PROB  = POISSON(1,0);
*Prob(X=1);
PROB1 = POISSON(1,1)-POISSON(1,0);
*Prob(X=2);
PROB2 = POISSON(1,2)-POISSON(1,1);
*Prob(X=3);
PROB3 = POISSON(1,3)-POISSON(1,2);
PROC PRINT;
RUN;
```

The pdf and cdf for the Poisson distribution and their plots with a mean of 2 are obtained using the SAS commands below. The SAS output displays the graphs. The pdf is illustrated in black and the cdf is illustrated in red.

SAS commands:

```
DATA POISSON;
  RETAIN X J PDF CDF;
  J=0;
  CDF = POISSON(2.0,0);
  PDF = CDF;
  PUT J '  PDF =  ' PDF  '  CDF =  ' CDF;
  OUTPUT;
  DO X=0 TO 9;
  PDF = POISSON(2.0,X+1) - CDF;
  CDF = POISSON(2.0,X+1);
  J = J + 1;
  PUT J '  PDF =  ' PDF  '  CDF =  ' CDF;
```

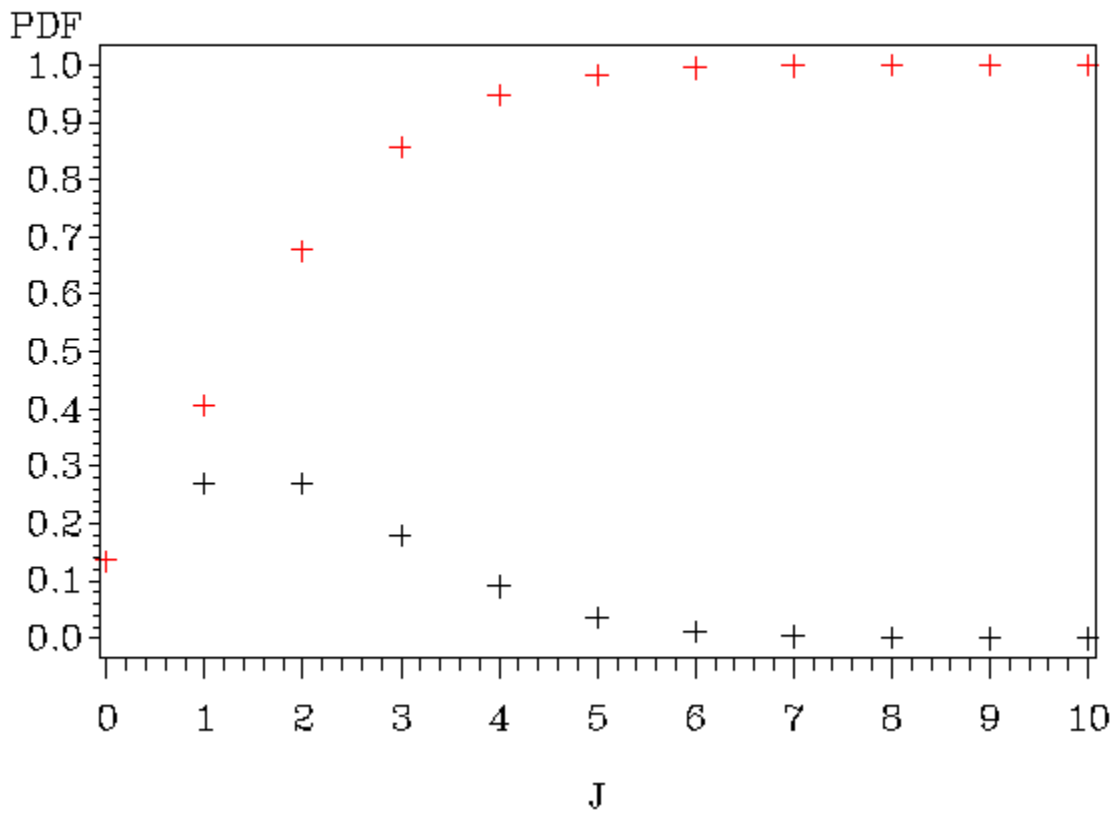
```
  OUTPUT;
  END;

GOPTIONS DEVICE= GIF VPOS= 24 HPOS= 75 VSIZE= 5 HSIZE= 6 FTEXT=COMPLEX;
ODS HTML;
ODS GRAPHICS ON;

PROC GPLOT DATA=POISSON;
  PLOT PDF*J CDF*J/OVERLAY;
RUN;
QUIT;

ODS GRAPHICS OFF;
ODS HTML CLOSE;
```

SAS output:



Note 5.2 – Creating the Poissonness plot

Using the frequency values from Table 5.4, we can create a *Poissonness plot*. Basically we are plotting { *ln(freq(x)) + ln(x!) versus x* }. The SAS functions of interest here are **LOG(X)** and **FACT(X)**. The **LOG(X)** function is the Natural log or (log base *e*) of X. The **LOG10(X)** function is the log base 10 of X. The **FACT(X)** function returns X! or X factorial. The Poissonness plot is a simple scatter plot that can be created using the SAS commands below.

SAS commands:

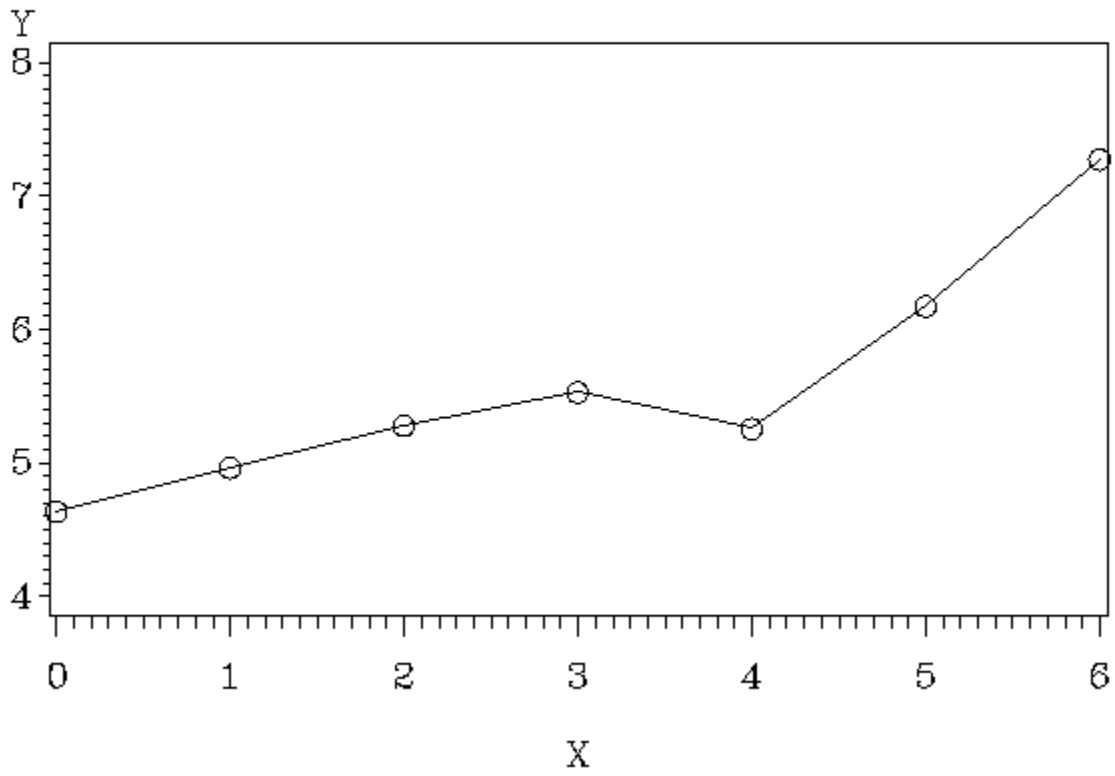
```
DATA POISSONPLOT;
INPUT X FREQ;
Y = LOG(FREQ) + LOG(FACT(X));
DATALINES;
0 103
1 143
2  98
3  42
4   8
5   4
6   2
;

GOPTIONS DEVICE= GIF VPOS= 24 HPOS= 75 VSIZE= 5 HSIZE= 6 FTEXT=COMPLEX;
ODS HTML;
ODS GRAPHICS ON;

PROC GPLOT DATA=POISSONPLOT;
  PLOT Y*X;
TITLE 'Poissonness Plot';
SYMBOL INTERPOL=JOIN VALUE=CIRCLE;
RUN;
QUIT;

ODS GRAPHICS OFF;
ODS HTML CLOSE;
```

SAS output:



Note 5.3 – Finding normal probabilities

The SAS function **PROBNORM(Z)** returns the cumulative probability associated with the standard normal variable. For example if we would like to find the area under the standard normal distribution's probability density function less than or equal to 1.0, we could use the command **PROBNORM(1.0)**.

SAS commands:

```
DATA NORMAL;
PROB = PROBNORM(1);
PROC PRINT;
RUN;
```

SAS output:

```
                                The SAS System

                                   PROB

                                  0.84134
```

The **PROBNORM** statement is used in the **DATA** step to obtain the values of the standard normal cdf for a number of values. **PROC GPLOT** is then used to plot these values.

SAS commands:

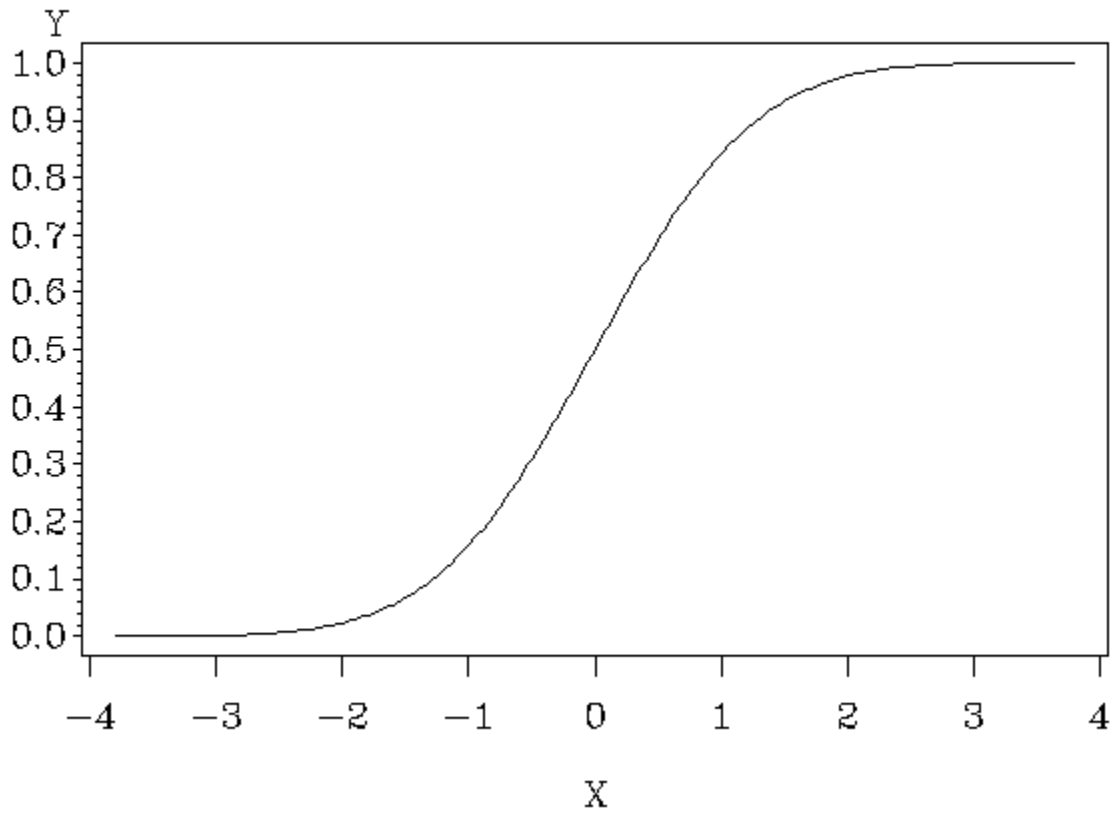
```
DATA NORMCDF;
  DO X = -3.8 TO 3.8 BY .1;
    Y = PROBNORM(X);
    OUTPUT;
  END;

GOPTIONS DEVICE= GIF VPOS= 24 HPOS= 75 VSIZE= 5 HSIZE= 6 FTEXT=COMPLEX;
ODS HTML;
ODS GRAPHICS ON;

PROC GPLOT DATA=NORMCDF;
  PLOT Y*X;
  SYMBOL INTERPOL=SMOOTH VALUE=NONE;
RUN;
QUIT;

ODS GRAPHICS OFF;
ODS HTML CLOSE;
```

SAS output:



Since **PROBNORM** can be used to find the cdf for a N(0,1) variable, we can also use **PROBNORM** to find the probability that a N(80,10) variable is greater than 95 or Prob(X>95). We shall find the probability that the variable is less than or equal to 95 and we can then subtract that probability from 1.

SAS commands:

```
DATA NORMCDF;
  Z=(95-80)/10;
  PLTE=PROBNORM(Z);
  PGT=1-PLTE;
  PUT ' PROBABILITY OF BEING GREATER THAN 95 IS  ' PGT;
RUN;
```

Note 5.4 – Creating a normal probability plot

As was mentioned in the SAS program notes for Chapter 3, if we wish to create a normal probability plot , we saw that this plot is part of the **PROC UNIVARIATE** output when the option PLOT is specified.