The Poisson Probability Distribution

Another discrete random variable that has numerous practical applications is Poisson Random variable. Its probability distribution provides a good model for data that represent the number of occurrences of a specified event in a given unit of time or space.

- **The number of calls received by a switchboard during a given period of time,**
- **The number of bacteria per small volume of fluid,**
- **The number of customer arrivals at a checkout counter during a given minute,**
- **The number of machine breakdown during a given day,**
- **The number of traffic accidents at a given intersection during a given time period**

Occurrence (Wikipedia)

The Poisson distribution arises in connection with **Poisson processes**. It applies to various phenomena of discrete properties (that is, those that may happen 0, 1, 2, 3, ... times during a given period of time or in a given area) whenever the probability of the phenomenon happening is constant in time or space. Examples of events that may be modeled as a Poisson distribution include:

- The number of phone calls at a **call centre** per minute.
- Under an assumption of **homogeneity**, the number of times a **web server** is accessed per minute.
- The number of **mutations** in a given stretch of **DNA** after a certain amount of radiation.
- The proportion of **cells** that will be infected at a given multiplicity of infection.

Definition:

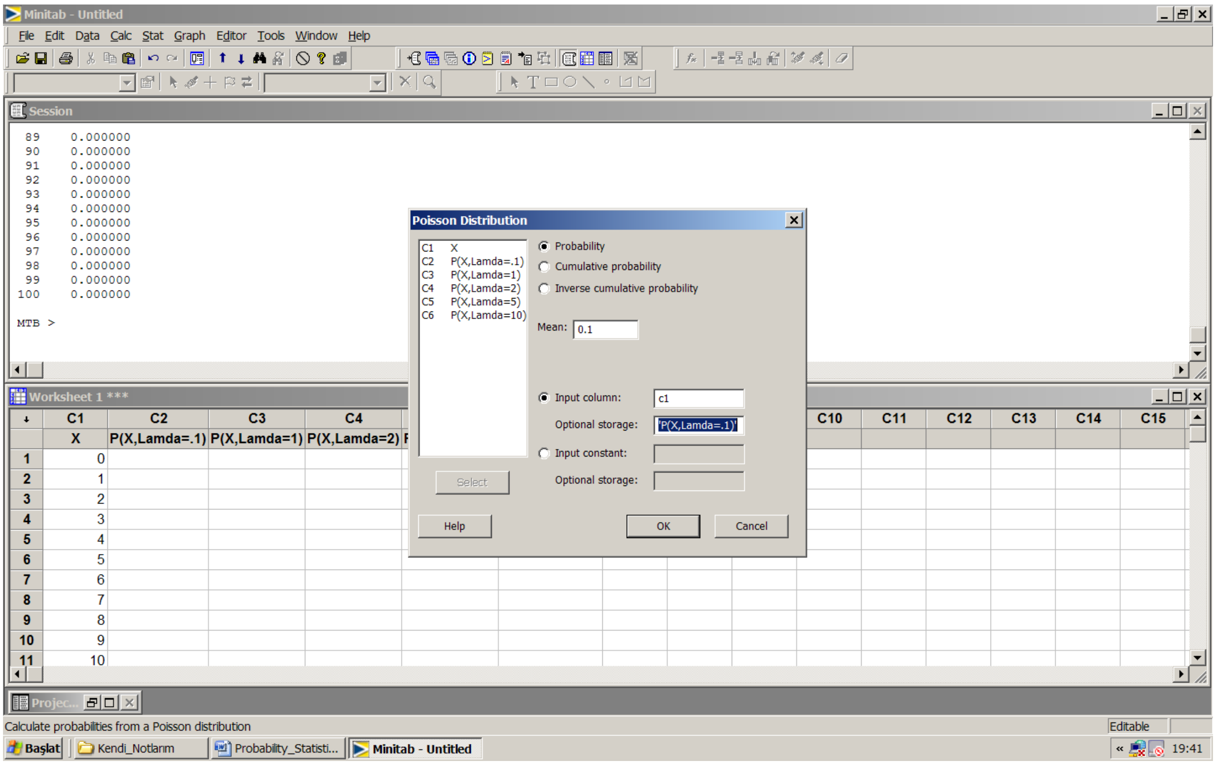
Given an interval of real numbers, assume events occur at random throughout the interval. If the interval can be partitioned into subintervals of small enough length such that

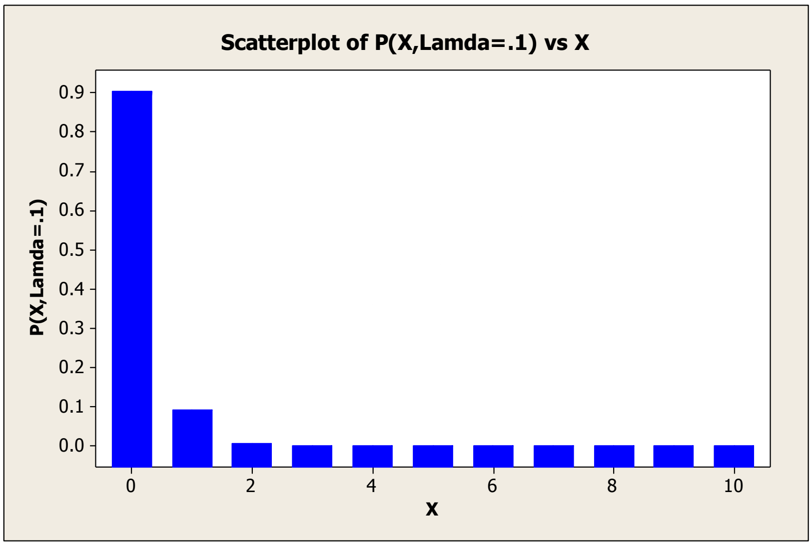
1. ***The probability of more than one event in a subinterval is zero,***
2. ***The probability of one event in a subinterval is the same for all subintervals and proportional to the length of the subinterval, and***
3. ***The event in each subinterval is independent of other subintervals, the random experiment is called a Poisson process.***

The random variable X that equals the number of events in the interval is a Poisson random variable with parameter λ >0 and the probability distribution of X is

$$P(X=k)=\frac{e^{-\lambda}\lambda^{k}}{k!} \qquad k=0,1,2,....$$

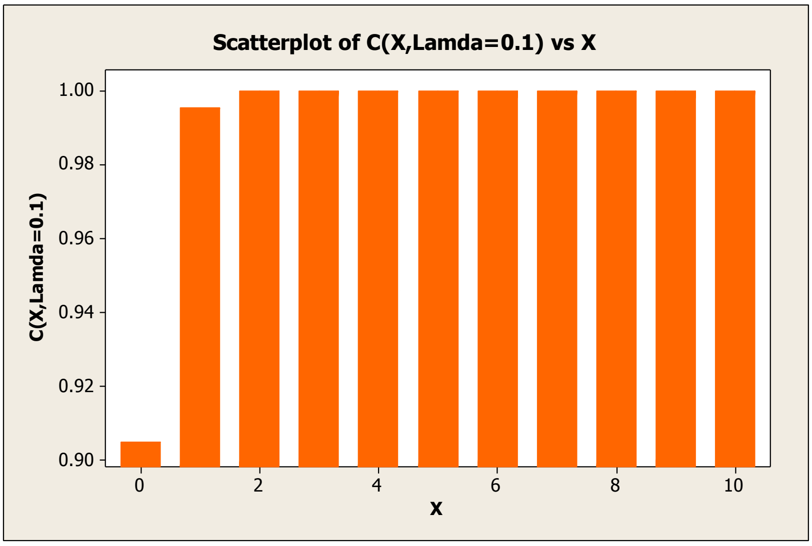
$$e=2.71828...$$



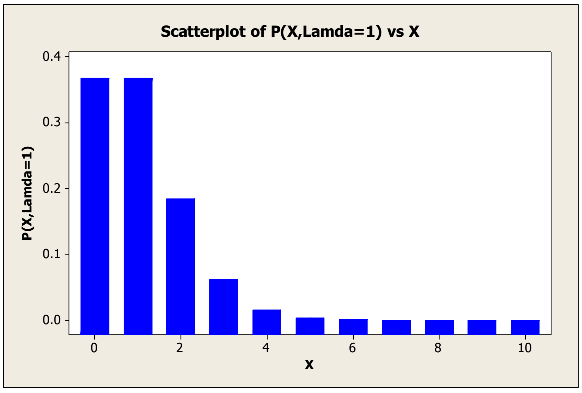


Poisson Probability Distribution with λ=0.1

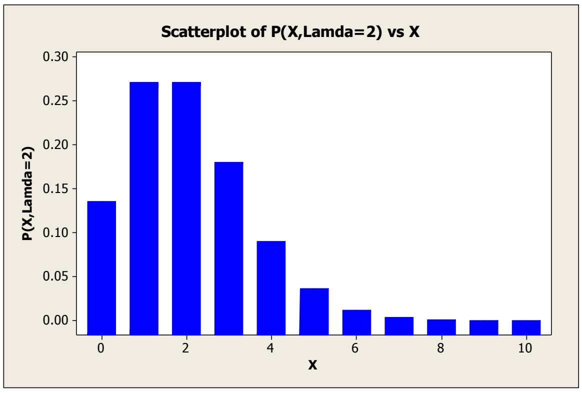
The sum of probabilities is 1 because $\sum_{k=0}^{\infty}\frac{e^{-\lambda}\lambda^{k}}{k!}=e^{-\lambda}\sum_{k=0}^{\infty}\frac{\lambda^{k}}{k!}$ **and the summation on the right-hand side is recognized to be Taylor expansion of e^x evaluated at λ. Therefore summation equals e^{λ} and the right-hand side equals $e^{-\lambda}e^{\lambda}=1$.**



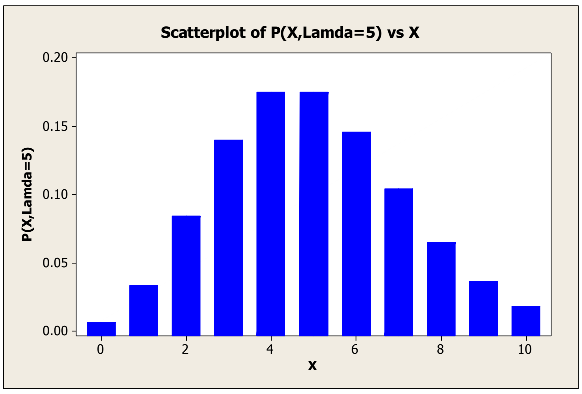
Poisson Cumulative Distribution with λ=0.1



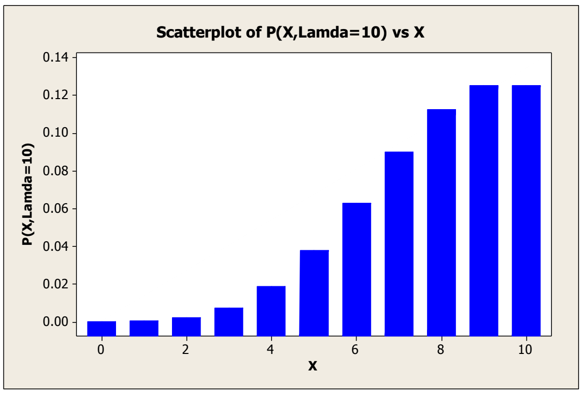
Poisson Prob. Dist. with λ=1



Poisson Prob. Dist. with λ=2



Poisson Prob. Dist. with λ=5



Poisson Prob. Dist. with λ=10

Example: Flaws occur along the length of a thin copper wire. Suppose that the number of flaws follows a Poisson distribution with a mean of 2.3 flaws per millimeter.

i. Determine the probability of exactly 2 flaws in 1 millimeter of wire.
ii. Determine the probability of 10 flaws in 5 millimeter of wire.
iii. Determine the probability of at least 1 flaw in 2 millimeter of wire.

Let X denote the number of flaws in 1 millimeter of wire.

$$P(X=2)=\frac{e^{-2.3}2.3^{2}}{2!}=0.265$$

Let X denote the number of flaws in 5 millimeter of wire.

E(X) = (5 mm)(2.3 flaws/mm)= 11.5 flaws

$$P(X=10)=\frac{e^{-11.5}11.5^{10}}{10!}=0.113$$

Let X denote the number of flaws in 2 millimeter of wire.

E(X) =(2 mm)(2.3 flaws/mm)= 4.6 flaws

$$P(X\geq 1)=1-P(X=0)=1-e^{-4.6}=0.9899$$

Mean and Variance of Poisson Distribution

If X is a Poisson random variable with parameter λ, then

$$E(X)=\lambda \quad \text{and}$$

$$V(X)=\lambda$$

$E(X)=\lambda$ $\quad E(X^2)=\lambda^2+\lambda$ $\quad V(X)=\lambda^2+\lambda-\lambda^2=\lambda$

The mean and variance of a Poisson random variable are equal. Consequently, information on the variability is easily obtained. Conversely, if the variance count data is much greater than the mean of the same data, the Poisson distribution is not a good model for the distribution of the random variable.

Wikipedia

Poisson noise and characterizing small occurrences

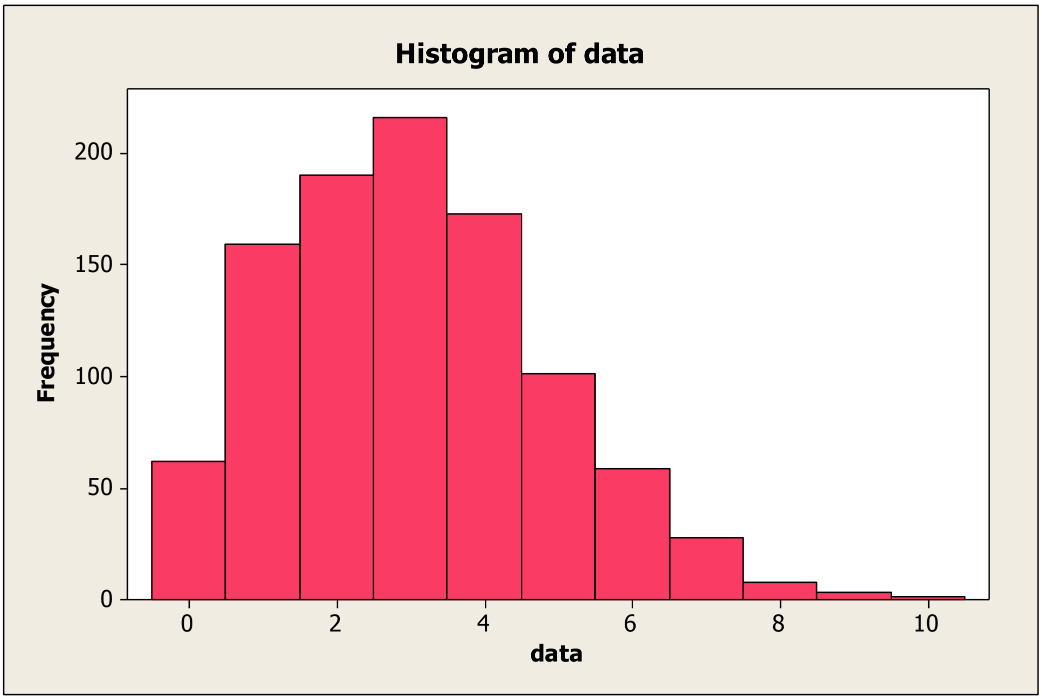
The parameter λ is not only the *mean* number of occurrences $\langle k \rangle$, but also its variance $\sigma_k^2 = \langle k^2 \rangle - \langle k \rangle^2$. Thus, the number of observed occurrences fluctuates about its mean λ with a standard deviation $\sigma_k = \sqrt{\lambda}$. These fluctuations are denoted as **Poisson noise** or (particularly in electronics) as **shot noise**.

```
MTB > Random 1000 c1;
SUBC>   Poisson 3.
```

Data Display

C1

6	3	4	3	1	5	4	6	2	4	3	2	0	0	2
2	1	4	4	4	3	1	7	4	3	1	2	2	2	5
5	4	8	3	2	0	1	6	1	3	2	2	1	4	3
2	3	2	4	2	4	3	4	3	3	1	1	3	3	7
2	2	6	6	4	6	0	2	2	3	3	8	6	3	0
5	2	4	6	3	4	2	4	3	5	5	4	3	3	1
3	3	1	3	3	5	3	3	6	4	5	4	4	1	4
2	1	0	1	2	0	6	5	3	1	0	5	2	1	2
6	4	1	4	0	7	5	1	2	2	3	2	3	3	2
2	3	4	1	2	1	1	3	3	2	3	4	7	3	1
3	1	2	2	2	3	6	0	0	3	3	3	1	4	4
3	2	3	5	3	10	2	3	4	3	5	3	4	2	1
1	2	2	5	2	3	1	2	2	4	1	3	5	4	4
2	4	3	2	2	3	4	6	3	3	7	2	3	3	3
2	2	2	2	4	2	5	7	0	1	2	2	1	3	2
0	3	4	2	4	4	6	7	4	2	1	1	5	1	3
1	2	1	4	1	7	1	4	1	6	6	2	4	2	5
4	2	2	4	4	2	6	0	2	2	5	4	7	4	5
6	4	2	5	5	0	4	3	3	1	2	4	7	4	2
8	1	6	0	2	1	5	2	7	2	6	4	4	3	4
2	5	1	4	0	0	5	2	5	7	6	4	1	5	4
1	2	3	8	2	7	4	4	6	3	2	3	4	0	4
1	3	5	3	3	3	3	0	0	2	3	2	5	2	3
1	3	3	0	2	1	4	4	2	4	3	3	4	1	3
2	5	3	1	6	0	5	1	4	5	6	3	5	0	3
3	2	3	4	0	8	0	3	3	4	5	4	0	5	3
5	4	4	3	3	2	4	1	2	4	2	5	5	1	4
4	3	1	2	1	4	4	1	3	2	4	2	3	1	2
2	3	2	3	2	3	0	7	4	1	4	2	2	1	0
0	5	1	3	1	6	6	5	3	1	3	4	4	4	0
5	6	1	1	3	1	4	2	3	3	3	4	0	2	3
4	1	1	2	2	1	1	5	3	3	6	3	1	2	1
3	2	4	6	5	2	1	5	5	7	7	5	0	4	4
4	5	4	1	6	1	4	4	2	1	3	4	3	3	4
3	4	1	3	1	1	4	1	1	4	1	2	3	4	1
5	5	6	2	2	1	1	4	1	3	1	5	1	0	3
3	3	6	1	2	3	4	2	7	4	2	1	3	5	5
3	3	3	7	5	2	1	5	2	3	3	3	6	2	8
2	3	1	3	9	7	2	4	4	6	4	4	7	2	8
2	1	6	2	3	4	2	1	1	2	2	3	4	4	2
1	1	2	1	3	3	5	4	1	4	2	0	2	1	1
3	6	5	6	0	1	0	2	2	4	2	1	4	4	4
3	1	2	5	3	4	2	1	3	5	3	5	3	1	4
5	0	1	4	3	3	6	1	5	5	3	1	4	1	1
3	1	1	5	3	1	1	2	2	1	5	4	2	6	4
5	4	3	1	3	4	5	3	0	2	4	4	3	3	2
5	6	3	2	4	1	4	5	6	3	4	4	1	1	1
7	1	2	4	1	3	1	6	3	0	3	4	5	2	4
4	2	0	3	2	1	4	1	1	5	9	6	3	5	0
2	5	5	0	3	0	1	3	1	4	5	1	4	4	5
6	3	3	3	1	4	2	3	5	5	4	3	1	1	0
3	0	3	1	6	5	4	3	5	1	3	3	2	2	3
3	0	5	5	2	1	2	3	4	4	2	2	2	5	9
4	0	3	2	3	4	1	6	2	4	6	3	4	3	3
4	7	1	2	0	3	0	4	2	6	2	2	5	4	4
6	5	3	2	3	4	3	6	2	2	2	0	3	2	1
7	3	6	1	2	1	2	0	1	2	2	2	3	2	5
4	3	1	2	3	3	3	1	2	6	2	3	3	0	3
3	5	2	3	4	2	5	0	3	2	2	0	6	2	2
6	3	3	4	7	5	0	4	2	2	2	1	3	2	3
1	6	2	2	4	4	3	2	4	2	3	4	1	3	3
0	4	0	5	1	3	0	3	6	3	1	8	7	4	5
3	1	3	1	1	3	1	1	4	3	4	7	1	3	5
3	3	5	3	4	1	4	4	5	1	3	1	0	4	2
2	2	4	7	5	3	5	4	4	7	5	2	6	2	6
2	3	1	3	2	3	0	6	1	1	2	5	3	3	5
2	5	2	5	4	3	0	5	4	4					




```
MTB > desc data
```

Descriptive Statistics: data

```
Variable     N  N*    Mean  SE Mean   StDev  Minimum      Q1  Median      Q3
data      1000   0  3.0350   0.0574  1.8163   0.0000  2.0000  3.0000  4.0000

Variable  Maximum
data      10.0000
```

```
MTB > PDF c1 c2;
SUBC>   Poisson 1.5.
MTB > CDF c1 c3;
SUBC>   Poisson 1.5.
MTB > print c1 c2 c3
```

Data Display

```
Row    x       P(X)      C(X)
  1    0   0.223130   0.22313
  2    1   0.334695   0.55783
  3    2   0.251021   0.80885
  4    3   0.125511   0.93436
  5    4   0.047067   0.98142
  6    5   0.014120   0.99554
  7    6   0.003530   0.99907
  8    7   0.000756   0.99983
  9    8   0.000142   0.99997
 10    9   0.000024   1.00000
 11   10   0.000004   1.00000
```

Exercise: Consider a Poisson random variable with λ=1.5. Fill in the blanks in the table below.

The Problem	List the Values x	Write the probability	Rewrite the probability (if needed)	Find the Probability
3 or less				
3 or more				
More than 3	4,5,6…	P(X>3)	1-P(X≤3)	0.066
Fewer than 3				
Between 2 and 4 (inclusive)				
Exactly 3				

The Poisson Approximation to the Binomial Distribution

The Poisson probability distribution provides a simple, easy-to-compute, and accurate approximation to binomial probabilities when n is large and λ=np is small, preferably np<5.

n = 100 p = 0.04 λ = 100 (0.04) =4

```
MTB > PDF 'x' 'P(X)';
SUBC>   Binomial 100 0.04.
MTB > PDF c1 c3;
SUBC>   Poisson 4.
MTB > print c1-c3
```

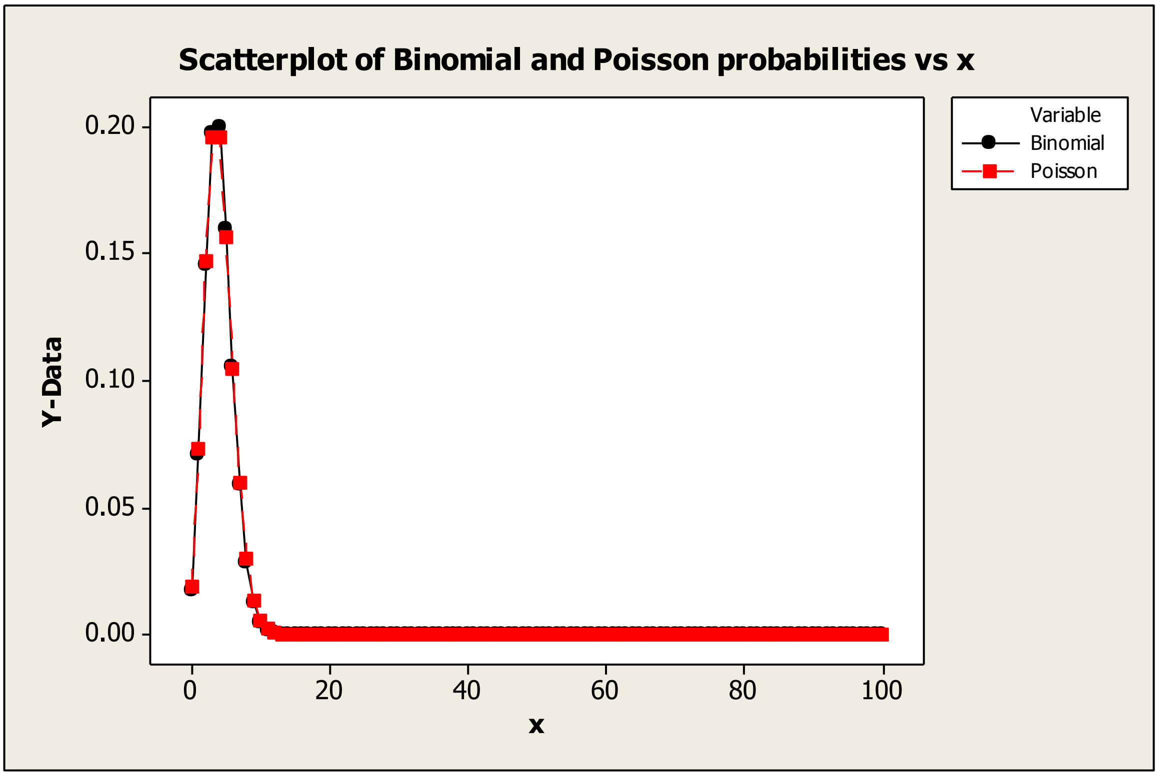
Data Display

Row	x	Binomial	Poisson
1	0	0.016870	0.018316
2	1	0.070293	0.073263
3	2	0.144979	0.146525
4	3	0.197333	0.195367
5	4	0.199388	0.195367
6	5	0.159511	0.156293

7	6	0.105233	0.104196
8	7	0.058880	0.059540
9	8	0.028520	0.029770
10	9	0.012147	0.013231
11	10	0.004606	0.005292
12	11	0.001570	0.001925
13	12	0.000485	0.000642
14	13	0.000137	0.000197
15	14	0.000035	0.000056
16	15	0.000008	0.000015
17	16	0.000002	0.000004
18	17	0.000000	0.000001
19	18	0.000000	0.000000
20	19	0.000000	0.000000
21	20	0.000000	0.000000
22	21	0.000000	0.000000
23	22	0.000000	0.000000
24	23	0.000000	0.000000
25	24	0.000000	0.000000
26	25	0.000000	0.000000
27	26	0.000000	0.000000
28	27	0.000000	0.000000
29	28	0.000000	0.000000
30	29	0.000000	0.000000
31	30	0.000000	0.000000
32	31	0.000000	0.000000
33	32	0.000000	0.000000
34	33	0.000000	0.000000
35	34	0.000000	0.000000
36	35	0.000000	0.000000
37	36	0.000000	0.000000
38	37	0.000000	0.000000
39	38	0.000000	0.000000
40	39	0.000000	0.000000

41	40	0.000000	0.000000
42	41	0.000000	0.000000
43	42	0.000000	0.000000
44	43	0.000000	0.000000
45	44	0.000000	0.000000
46	45	0.000000	0.000000
47	46	0.000000	0.000000
48	47	0.000000	0.000000
49	48	0.000000	0.000000
50	49	0.000000	0.000000
51	50	0.000000	0.000000
52	51	0.000000	0.000000
53	52	0.000000	0.000000
54	53	0.000000	0.000000
55	54	0.000000	0.000000
56	55	0.000000	0.000000
57	56	0.000000	0.000000
58	57	0.000000	0.000000
59	58	0.000000	0.000000
60	59	0.000000	0.000000
61	60	0.000000	0.000000
62	61	0.000000	0.000000
63	62	0.000000	0.000000
64	63	0.000000	0.000000
65	64	0.000000	0.000000
66	65	0.000000	0.000000
67	66	0.000000	0.000000
68	67	0.000000	0.000000
69	68	0.000000	0.000000
70	69	0.000000	0.000000
71	70	0.000000	0.000000
72	71	0.000000	0.000000
73	72	0.000000	0.000000
74	73	0.000000	0.000000

75	74	0.000000	0.000000
76	75	0.000000	0.000000
77	76	0.000000	0.000000
78	77	0.000000	0.000000
79	78	0.000000	0.000000
80	79	0.000000	0.000000
81	80	0.000000	0.000000
82	81	0.000000	0.000000
83	82	0.000000	0.000000
84	83	0.000000	0.000000
85	84	0.000000	0.000000
86	85	0.000000	0.000000
87	86	0.000000	0.000000
88	87	0.000000	0.000000
89	88	0.000000	0.000000
90	89	0.000000	0.000000
91	90	0.000000	0.000000
92	91	0.000000	0.000000
93	92	0.000000	0.000000
94	93	0.000000	0.000000
95	94	0.000000	0.000000
96	95	0.000000	0.000000
97	96	0.000000	0.000000
98	97	0.000000	0.000000
99	98	0.000000	0.000000
100	99	0.000000	0.000000
101	100	0.000000	0.000000



Hypergeometric > Binomial > Poisson

Hyper geometric: N=50 M=25 n=5

Binomial: n=5 p=M/N=25/50=0.5

k	Hypergeometric	Binomial
0	0.025	0.031
1	0.149	0.156
2	0.326	0.312
3	0.326	0.312
4	0.149	0.156
5	0.025	0.031

Binomial: n=5 p=M/N=25/50=0.5

Poisson : λ = np=5*(0.5) = 2.5

k	Hypergeometric	Binomial	Poisson
0	0.025	0.031	0.082085
1	0.149	0.156	0.205212
2	0.326	0.312	0.256516
3	0.326	0.312	0.213763
4	0.149	0.156	0.133602
5	0.025	0.031	0.066801

Compare results

```
MTB > pdf c1 c2;
SUBC> hypergeometric 850 50 5.
MTB > pdf c1 c2;
SUBC> hypergeometric 50 25 5.
MTB > pdf c1 c3;
SUBC> Binomial 5 0.5.
MTB > pdf c1 c4;
SUBC> Poisson 2.5 .
MTB > print c1-c4
```

Data Display

```
Row   C1   Hypergeometric   Binomial    Poisson
  1    0         0.025076    0.03125   0.082085
  2    1         0.149262    0.15625   0.205212
  3    2         0.325662    0.31250   0.256516
  4    3         0.325662    0.31250   0.213763
  5    4         0.149262    0.15625   0.133602
  6    5         0.025076    0.03125   0.066801
```

More realistic approximation

Hyper geometric: N=500 M=10 n=10

Binomial: n=10 p=M/N=10/500=0.02

Poisson : λ = np=10*(0.02) = 0.2

k	Hypergeometric	Binomial	Poisson
0	0.815554	0.817073	0.818731
1	0.169554	0.166750	0.163746
2	0.014247	0.015314	0.016375
3	0.000629	0.000833	0.001092
4	0.000016	0.000030	0.000055
5	0.000000	0.000001	0.000002
6	0.000000	0.000000	0.000000
7	0.000000	0.000000	0.000000
8	0.000000	0.000000	0.000000
9	0.000000	0.000000	0.000000
10	0.000000	0.000000	0.000000

Compare results

```
MTB > pdf c1 c2;
SUBC> hypergeometric 500 10 10.
MTB > pdf c1 c3;
SUBC> binomial 10 0.02.
MTB > pdf c1 c4;
SUBC> poisson 0.2.
MTB > print c1-c4
```

Data Display

Row	C1	Hypergeometric	Binomial	Poisson
1	0	0.815554	0.817073	0.818731
2	1	0.169554	0.166750	0.163746
3	2	0.014247	0.015314	0.016375
4	3	0.000629	0.000833	0.001092
5	4	0.000016	0.000030	0.000055
6	5	0.000000	0.000001	0.000002
7	6	0.000000	0.000000	0.000000
8	7	0.000000	0.000000	0.000000
9	8	0.000000	0.000000	0.000000
10	9	0.000000	0.000000	0.000000
11	10	0.000000	0.000000	0.000000