The Hypergeometric Probability Distribution

Definition:

A set of N objects contains

- ***M objects classified as successes***
- ***N-M objects classified as failures***

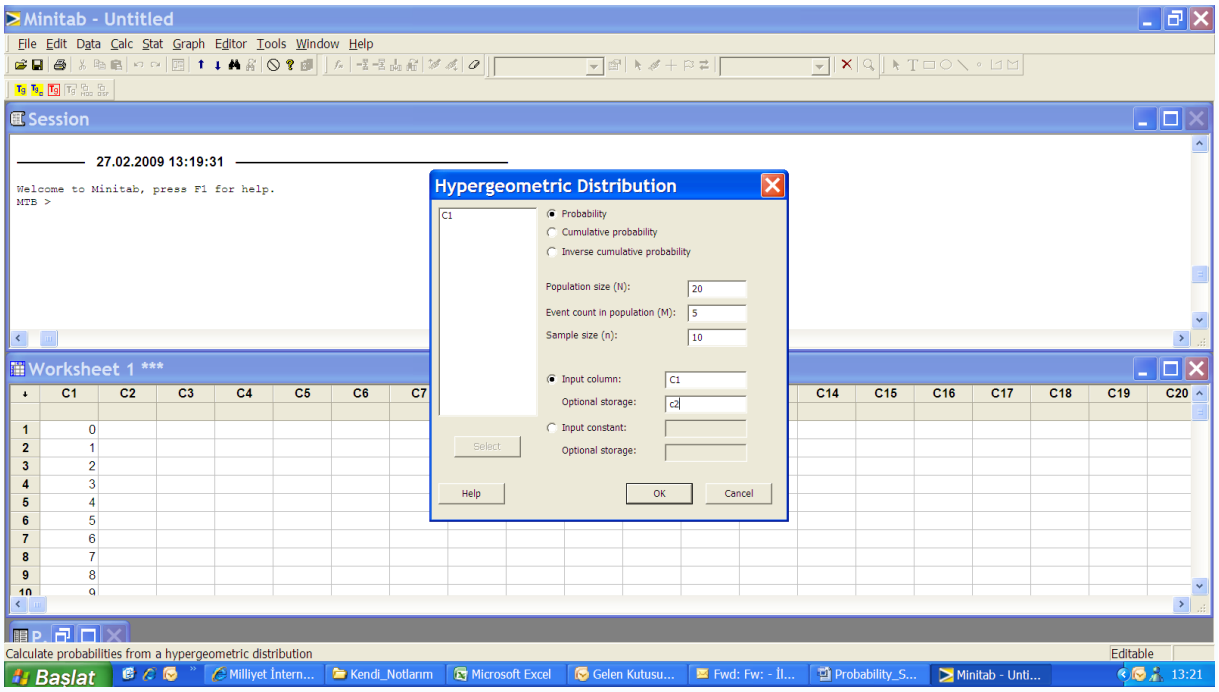
A sample of size n objects is selected randomly (without replacement) from the N objects, where M < N and n< N.

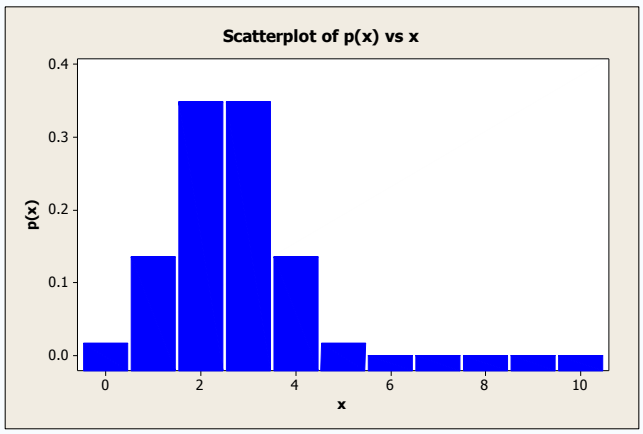
Let the random variable X denote the number successes in the sample. Then X is a Hypergeometric random variable and

$$P(X=k)=\frac{C_k^M C_{n-k}^{N-M}}{C_n^N}$$

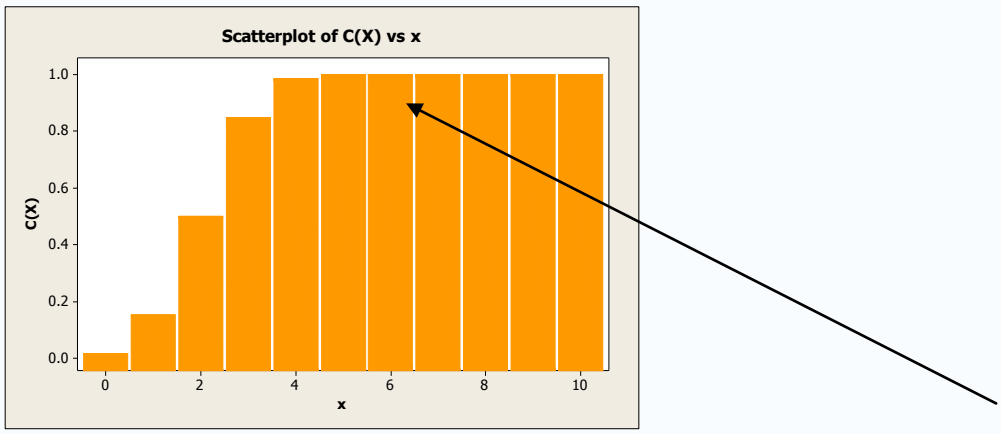
$$P(X=k)=\frac{\binom{M}{k}\binom{N-M}{n-k}}{\binom{N}{n}}$$

$$k=Max\{0,n+M-N\} \quad to \quad Min\{M,n\}$$





Hypergeometric Probability Distribution with N=20, M=5 and n=10



Cumulative Hypergeometric Distribution with N=20, M=5 and n=10

X=0…Min (M, n) Why?

N=20,M=5 and n=10

$$k = Max\{0, n+M-N\} \quad to \quad Min\{M, n\}$$

$$k = Max\{0, 10+5-20\} \quad to \quad Min\{5, 10\}$$

$$k = 0...5$$

x	Probability	Cumulative Probability
0	0.016254	0.01625
1	0.135449	0.15170
2	0.348297	0.50000
3	0.348297	0.84830
4	0.135449	0.98375
5	0.016254	1.00000
Sum	1.00	

```
MTB > PDF C1 C2;
SUBC>   Hypergeometric 20 5 10.
MTB > CDF C1 C3;
SUBC>   Hypergeometric 20 5 10.
MTB > PRINT C1-C3
```

Data Display

Row	C1	C2	C3
1	0	0.016254	0.01625
2	1	0.135449	0.15170
3	2	0.348297	0.50000
4	3	0.348297	0.84830
5	4	0.135449	0.98375
6	5	0.016254	1.00000

Example:

A day's production of **850** manufactured parts contains **50** parts that do not conform to customer requirements. **Two** parts are selected random, without replacement from the day's production. Let X the number of nonconforming parts in the sample. Find the following probabilities.

i. What is the probability that one of the parts in the sample do not conform?

ii. What is the probability that both parts do not conform?

$$P(X=k)=\frac{\binom{M}{k}\binom{N-M}{n-k}}{\binom{N}{n}}$$

i. P(X=1) =?

N=850 M= 50 n=2, and k=1

$$P(X=1)=\frac{\binom{50}{1}\binom{800}{1}}{\binom{850}{2}}=0.111$$

Alternative Solution :

P(X=1) =P (First part selected conforms and the second part selected does not, or the first part selected does not and the second part selected conforms)

= (800/850) (50/849) + (50/850) (800/849) =0.111

```
MTB > PDF 1;
SUBC> Hypergeometric 850 50 2.
```

Probability Density Function

```
Hypergeometric with N = 850, M = 50, and n = 2
```

```
x   P( X = x )
1     0.110857
```

ii. P(X=2)=?

N=850 M= 50 n=2, and k=2

$$P(X=2)=\frac{\binom{50}{2}\binom{800}{0}}{\binom{850}{2}}=0.003$$

Alternative Solution :

P(X=2) =P (Both parts do not conform)

= (50/850) (49/849) =0.003

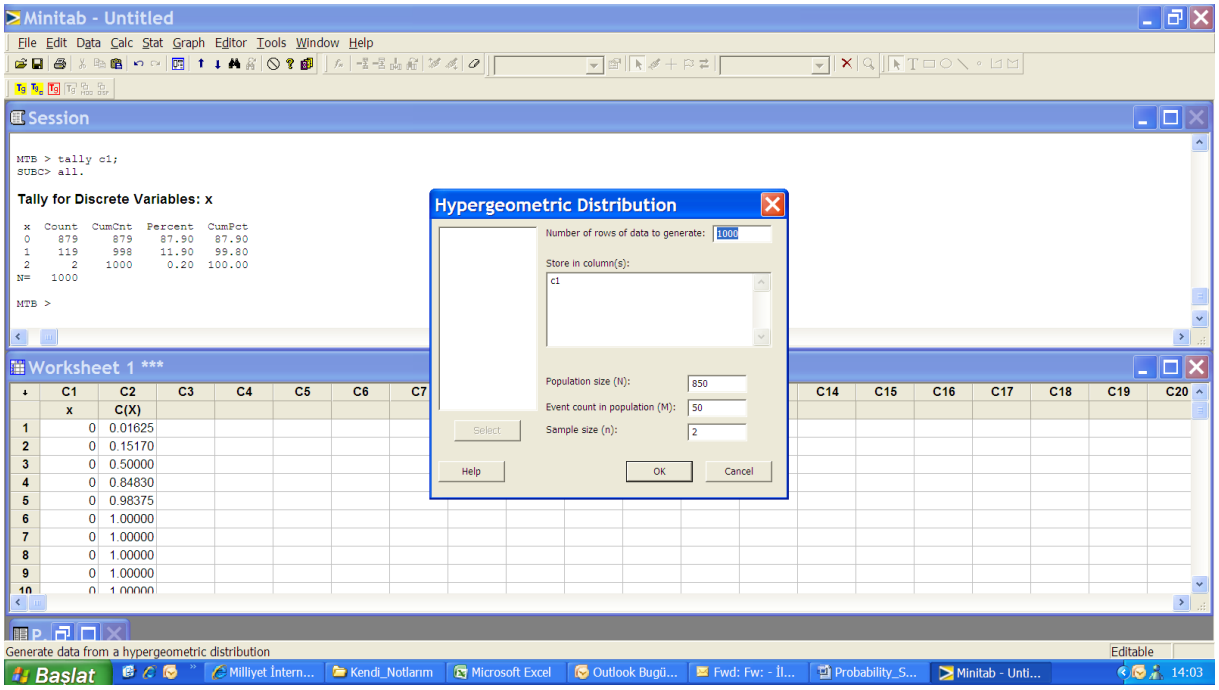
```
MTB > PDF 2;
SUBC> Hypergeometric 850 50 2.
```

Probability Density Function

```
Hypergeometric with N = 850, M = 50, and n = 2
```

```
x   P( X = x )
2    0.0033950
```

Obtaining probabilities with simulated data




```
MTB > Random 10000 c1;
SUBC> Hypergeometric 850 50 2.
MTB > print c1
```

Data Display

C1

1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0
0	0	0	1	0	0	0	0	0	0	1	0	0	0	0	0	1	0	0
0	0	0	0	0	1	0	0	0	0	0	0	1	0	0	0	0	0	0
0	0	0	0	0	0	0	0	1	1	1	0	0	0	0	0	0	0	0
0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0
0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0
0	0	0	1	0	1	0	0	0	0	0	0	0	1	0	0	0	0	0
0	0	0	0	0	1	0	1	0	0	1	0	0	0	0	0	0	1	0
0	1	0	0	0	0	0	0	1	0	0	0	1	0	0	0	1	0	0
1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
0	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0
0	0	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	1	0	0	0	1	0	0	1	1
0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0
0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	1	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	2	0	0	0	0	1	1	0	0	1

```
0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 1 0 0 0
0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 1 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0
1 0 0 0 0 0 0 1 1 0 0 0 0 0 0 0 0 0 0
0 0 0 0 1 0 1 1 0 0 0 0 0 0 0 0 0 1 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 1 0 0 0 1 0 0 1 0 0 1 0 0 0 0 0 0
1 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 1 0 0
0 0 0 1 0 0 0 0 1 0 0 0 0 0 0 1 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 1 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0
0 0 1 0 0 1 0 0 0 0 0 0 0 0 0 0 1 0 0
0 0 0 1 0 0 0 0 1 0 0 0 0 0 0 0 1 0 0
0 1 0 0 0 1 0 0 0 0 0 0 1 0 0 0 0 0 0
0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 1
0 0 0 0 0 0 0 0 1 1 1 0 0 0 0 0 0 0 0
0 0 1 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 1 0 0 0 0 0 1 0 0 1 0
0 0 1 0 0 0 1 0 0 0 0 0 0 0 0 0 1 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1
0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0
1 0 0 0 0 1 1 0 1 0 0 1 0 0 0 0 0 0 0
0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 1
0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0
0 0 1 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1
0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 1 0 0 0 0 1 0 0 0 0 0 1 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 1
1 0 0 0 1 0 0 0 0 0 1 0 0 0 0 0 0 0 0
1 0 0 0 1 0 0 0 0 0 0 1 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0
0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 1 1 0 0
0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0
1 0 0 0 0 1 0 0 0 0 1 0 0 0 0 0 0 0 0
0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1
1 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0
0 1 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 1 0
0 0 0 1 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 1 0 0 0 0 0 1 0 0 0 0
0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0
0 1 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0
1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 1 0 1 0 0 0 0 1 0 0 0 0
0 0 0 0 1 0 1 0 0 0 0 0 0 0 0 0 0 1 0
0 0 0 0 0 0 0 0 0 1 0 0 1 0 0 0 0 0 0
0 0 0 0 0 0 0 0 1 0 0 0 1 0 0 1 0 0 0
0 0 0 0 1 0 0 0 0 0 0 0 1 0 0 1 0 0 0
```

. . . . . .

```
MTB > tally c1;
SUBC> all.
```

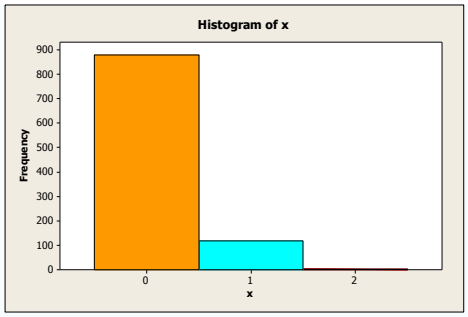
Tally for Discrete Variables: C1

C1	Count	CumCnt	Percent	CumPct
0	8817	8817	88.17	88.17
1	1150	9967	11.50	99.67
2	33	10000	0.33	100.00
N=	10000			

```
MTB > pdf c1 c2;
SUBC> Hypergeometric 850 50 2.
MTB > print c1 c2
```

Data Display

Row	x	P(X)
1	0	0.885748
2	1	0.110857
3	2	0.003395



Mean and Variance for the Hypergeometric Random Variable

If X is a Hypergeometric random variable with parameters N, M and n.

$$\mu = n\left(\frac{M}{N}\right)$$

and

$$\sigma^2 = n\left(\frac{M}{N}\right)\left(\frac{N-M}{N}\right)\left(\frac{N-n}{N-1}\right)$$

The term in the variance of a Hypergeometric random variable

$$\left(\frac{N-n}{N-1}\right)$$

is called the finite population correction factor.

Example:

Find mean and variance above example.

N=850 M= 50 n=2

$$\mu = n\left(\frac{M}{N}\right) = 2(50/850) = 0.1176$$

$$\sigma^2 = n\left(\frac{M}{N}\right)\left(\frac{N-M}{N}\right)\left(\frac{N-n}{N-1}\right)$$
$$= 2(50/850)(800/850)(848/849) = 0.110596$$

$$\sigma = \sqrt{0.110596} = 0.33256$$

```
desc c1
```

Descriptive Statistics: C1

```
Variable       N      Mean   SE Mean     StDev   Minimum        Q1
Median
C1         10000   0.12130   0.00336   0.33556   0.00000   0.00000
0.00000

Variable        Q3   Maximum
C1         0.00000   2.00000
```

Example:

A batch of parts contains 100 parts from a local supplier and 200 parts from a supplier in the next state. If four parts are selected randomly and without replacement, **what is the probability they are all from the local supplier?**

Let X equals the number of parts from the local supplier. Then, X has a Hypergeometric distribution and he required probability is

$$P(X=4)=\frac{\binom{100}{4}\binom{200}{0}}{\binom{300}{4}}=0.0119$$

Exercise:

What is the probability that two or more parts are from the local supplier?

What is the probability that at least one part in the sample is from the local supplier?

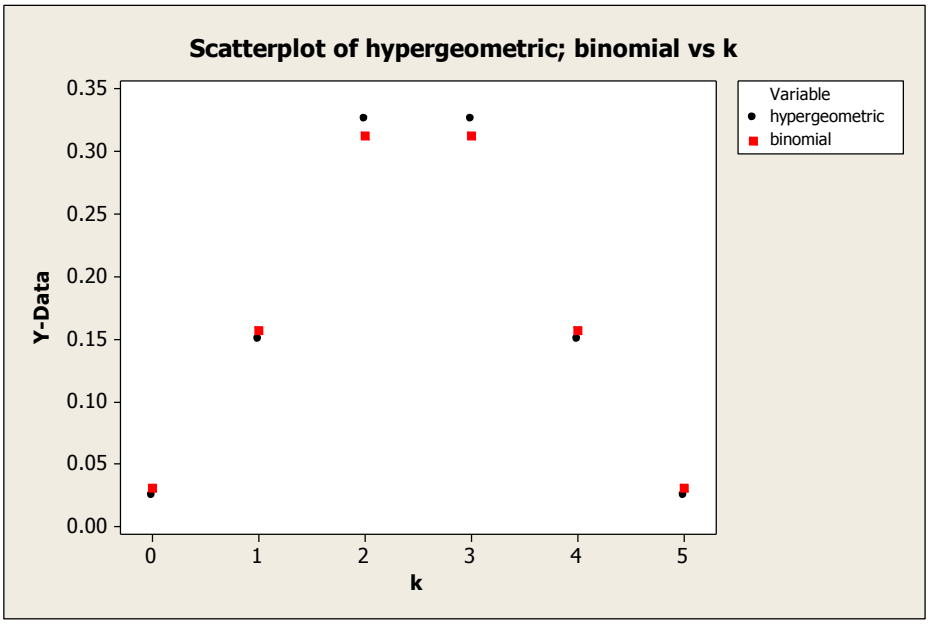
Binomial Approximation to Hypergeometric Distribution

If $n \leq 0.005N$ then the binomial distribution can be used two solve two-outcome sampling without replacement problems. This approximation is called binomial approximation to the Hypergeometric distribution.

Hypergeometric: 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



Generalization of the Hypergeometric Probability Distribution

- *A random sample of n objects is taken from a finite population of N_T objects by sampling without replacement.*
- *Of the N_T objects in the population, N_1 are type of one, N_2 are type of two,…, N_k are type of k, and*

$$N_T = N_1 + N_2 + ... + N_k$$

- *Discrete random variable $X_1, X_2, \ldots, X_k$ are used to count the number of times type one, two,…,k appear in the sample with the actual count values*

$$X_1 = x_1, X_2 = x_2, ..., X_k = x_k$$

The probability x_1 of type one, x_2 of type two, ..., x_k of type k

$$P(X_1 = x_1, X_2 = x_2, ..., X_k = x_k) = \frac{\binom{N_1}{x_1}\binom{N_2}{x_2}\cdots\binom{N_k}{x_k}}{\binom{N_T}{n}} \quad where\ n = x_1 + x_2 + ... + x_k$$

Example:

$$X_1 = 2, X_2 = 2, X_3 = 2, X_4 = 4$$

N_T=52, N_1= N_2= N_3= N_4=13 and n=10

$$P(X_1 = 2, X_2 = 2, X_3 = 2, X_4 = 4) = \frac{\binom{13}{2}\binom{13}{2}\binom{13}{2}\binom{13}{4}}{\binom{52}{10}} = 0.0214$$

[Wikipedia]

Multivariate Hypergeometric Distribution

Probability mass function

Probability mass function (pmf)

$$\frac{\prod_{i=1}^{c}\binom{m_i}{k_i}}{\binom{N}{n}}$$

Mean

$$E(X_i) = \frac{nm_i}{N}$$

Variance

$$var(X_i) = \frac{m_i}{N}\left(1 - \frac{m_i}{N}\right)n\frac{N-n}{N-1}$$

$$cov(X_i, X_j) = -\frac{nm_i m_j}{N^2}\frac{N-n}{N-1}$$