12 HYPERGEOMETRIC DISTRIBUTION

Examples:

1. Five cards are chosen from a well shuffled deck.

   $X =$ the number of diamonds selected.

2. An audio amplifier contains six transistors. It has been ascertained that three of the transistors are faulty but it is not known which three. Amy removes three transistors at random, and inspects them.

   $X =$ the number of defective transistors that Amy finds.

3. A batch of 20 integrated circuit chips contains 20% defective chips. A sample of 10 is drawn at random.

   $X =$ the number of defective chips in the sample.

4. A batch of 100 printed circuit cards is populated with semiconductor chips. 20 of these are selected without replacement for function testing. If the original batch contains 30 defective cards, how will these show up in the sample?

   $X =$ the number of defective cards in the sample.

Hypergeometric Distribution:

A finite population of size N consists of:

M elements called successes
L elements called failures

A sample of n elements are selected at random **without replacement**.
X = number of successes

$$P(X = x) = \frac{\begin{pmatrix} M \\ x \end{pmatrix}\begin{pmatrix} L \\ n-x \end{pmatrix}}{\begin{pmatrix} N \\ n \end{pmatrix}}$$

X is said to have a **hypergeometric distribution**

Example: Draw 6 cards from a deck without replacement.
What is the probability of getting two hearts?

Solution: Here

$$\begin{aligned} M &= 13 \text{ number of hearts} \\ L &= 39 \text{ number of non-hearts} \\ N &= 52 \text{ total} \end{aligned}$$

$$P(2\ hearts) = \frac{\begin{pmatrix} 13 \\ 2 \end{pmatrix}\begin{pmatrix} 39 \\ 4 \end{pmatrix}}{\begin{pmatrix} 52 \\ 6 \end{pmatrix}} = .31513$$

Check in *R*

```
> dhyper(2, 13, 39, 6)
[1] 0.3151299
> round(dhyper(2, 13, 39, 6), 5)
[1] 0.31513
```

In R :

- **Example 1:** Five cards from a deck

```
> x<-0:5
> hyperprob<-dhyper(x, 13, 39, 5)
> round(hyperprob, 4)
[1] 0.2215 0.4114 0.2743 0.0815 0.0107 0.0005
```

- **Example 2:** Three transisters from 6.

```
x<-0:3
dhyper(x, 3, 3, 3)
[1] 0.05 0.45 0.45 0.05
```

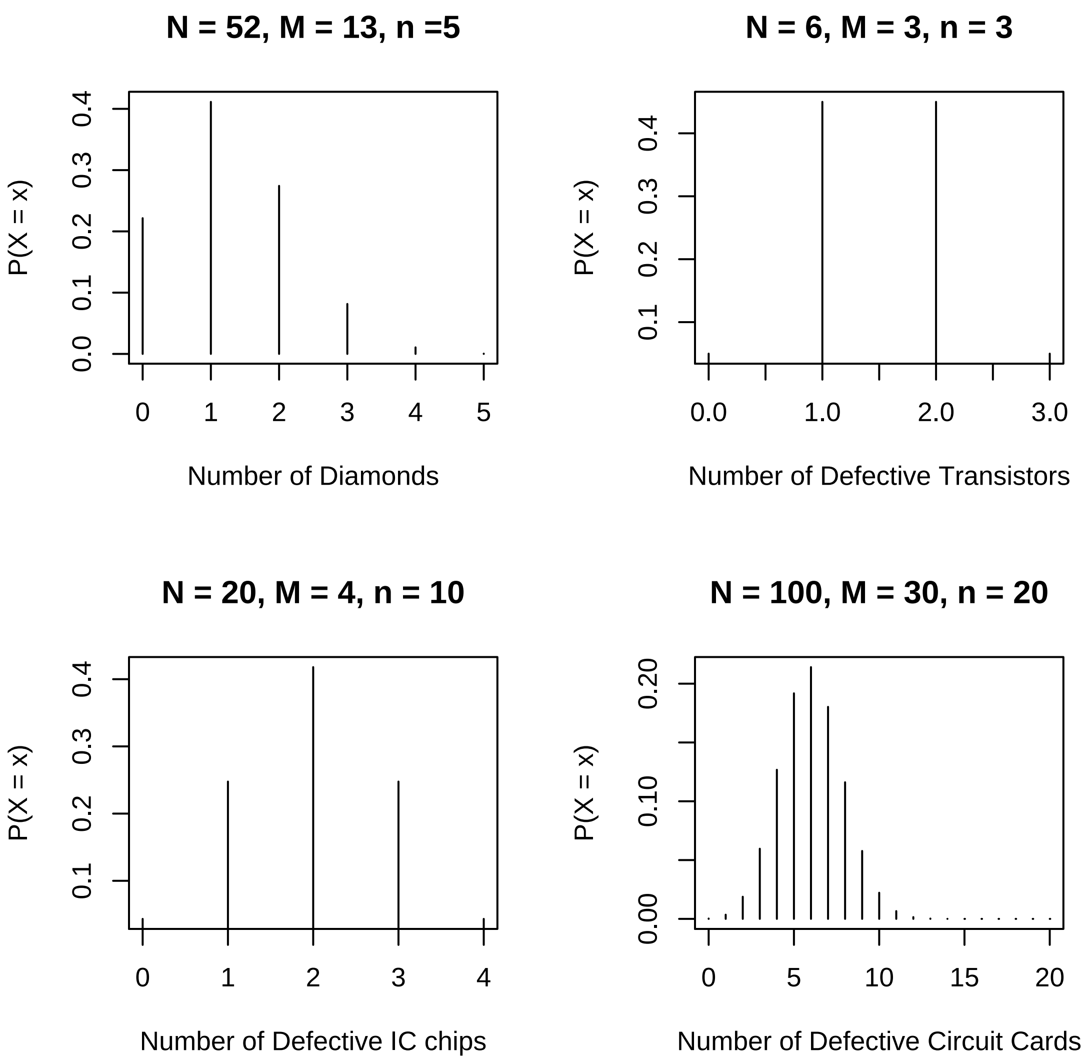
- **Example 3:** 10 IC chips from 20

```
x<-0:4
dhyper(x, 4, 16, 10)
[1] 0.04334365 0.24767802 0.41795666 0.24767802 0.04334365
```

- **Example 4:** 20 printed circuit cards from 100

```
x<- 0:10
dhyper(x, 30,70, 20)
```

Hypergeometric probability density functions (pdfs)



Example: Lotto

42 balls are numbered 1 - 42.

You select six numbers between 1 and 42. (The ones you write on your lotto card)

Number of possible ways to draw six mumbers in the range [1, 42] $= \binom{42}{6}$

What is the probability that they contain

(i) match 6?

(ii) match 5?

(ii) match 4?

(iii) match 3?

Solution:

Total = 42; Favourable = 6; Non-Favourable = 36.

Sample size n = 6.

$$P(\text{match } 4) = \frac{\binom{6}{4}\binom{36}{2}}{\binom{42}{6}} = .0018$$

ODDS OF ABOUT 1 in 500

Winning the Jackpot:

6 correct numbers, just one way of winning:

With 36 numbers

```
choose(36, 6)
```

```
1,947,792
```

- a chance of about 1 in 2 million.

With 39 numbers

```
choose(39, 6)
```

```
3,262,623
```

- a chance of less than 1 in 3 million.

With 42 numbers

```
choose(42, 6)
```

```
5,245,786
```

less than 1 in 5 million chance.

The current situation of 45 numbers:

```
choose(45, 6)|
```

```
8,145,060
```

- a chance of less than 1 in 8 million.

Cumulative Distribution Function

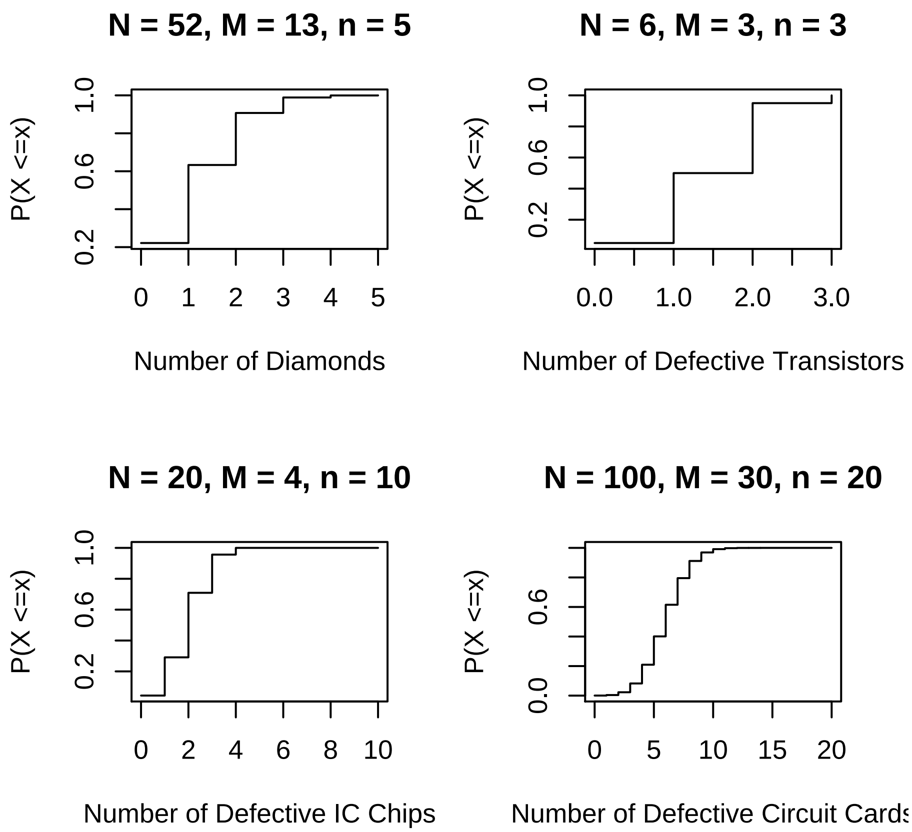
Want is the probability of getting at most 2 diamonds in the 5 selected without replacement from a well shuffled deck?

$$
\begin{aligned}
P(X \leq 2) &= P(X=0) + P(X=1) + P(X=2) \\
&= = \frac{\binom{39}{5}}{\binom{52}{5}} + \frac{\binom{13}{1}\binom{39}{4}}{\binom{52}{5}} + \frac{\binom{13}{2}\binom{39}{3}}{\binom{52}{5}} \\
&= 0.222 + 0.411 + 0.274 \\
&= .907
\end{aligned}
$$

In *R*

```
phyper(2, 13, 39, 5)
[1] 0.9072329
```

Hypergeometric cdfs



R Code:

```
 par(mfrow = c(2,2))
x<- 0:5 #Example 1
plot(x, phyper(x, 13, 39, 5),
 xlab = "Number of Diamonds",
 type = "s", ylab = "P(X <=x)",
 main = "N = 52, M = 13, n = 5")

x<- 0:3 #Example 2
plot(x, phyper(x, 3, 3, 3),
 xlab = "Number of Defective Transistors",
 type = "s", ylab = "P(X <=x)",
 main = "N = 6, M = 3, n = 3")

x<- 0:10 #Example 12.3
plot(x, phyper(x, 4, 16, 10),
  xlab = "Number of Defective IC Chips",
   type = "s", ylab = "P(X <=x)",
   main = "N = 20, M = 4, n = 10")

x<- 0:20 #Example 12.4
plot(x, phyper(x, 30, 70, 20),
   xlab = "Number of Defective Circuit Cards",
   type = "s", ylab = "P(X <=x)",
   main = "N = 100, M = 30, n = 20")
```

Binomial or Hypergeometric?

Boxes contain 20 items of which 10% are defective.
Find the probability that no more than 2 defectives will be obtained in a sample of size 10.
Let X = no of defectives.

$$P(X \leq 2) = P(X = 0) + P(X = 1) + P(X = 2)$$

With Replacement Sampling

$$\begin{aligned}
P(X = 0) &= .9^{10} \qquad = .3487 \\
P(X = 1) &= \binom{10}{1} .1^1 .9^9 = .3874 \\
P(X = 2) &= \binom{10}{2} .1^2 .9^8 = .1937
\end{aligned}$$

Without Replacement Sampling

$$\begin{aligned}
P(X = 0) &= \frac{\binom{18}{10}}{\binom{20}{10}} = .2368 \\
P(X = 1) &= \frac{\binom{2}{1}\binom{18}{9}}{\binom{20}{10}} = .5263 \\
P(X = 2) &= \frac{\binom{2}{2}\binom{18}{8}}{\binom{20}{10}} = .2368
\end{aligned}$$

Binomial or Hypergeometric?

Boxes contain 200 items of which 10% are defective.
Find the probability that no more than 2 defectives will be obtained in a sample of size 10.
Let X = no of defectives.

$$P(X \le 2) = P(X = 0) + P(X = 1) + P(X = 2)$$

With Replacement Sampling

$$\begin{aligned} P(X = 0) &= .9^{10} \quad = .3487 \\ P(X = 1) &= \binom{10}{1} .1^1 .9^9 = .3874 \\ P(X = 2) &= \binom{10}{2} .1^2 .9^8 = .1937 \end{aligned}$$

Without Replacement Sampling

$$\begin{aligned} P(X = 0) &= \frac{\binom{180}{10}}{\binom{200}{10}} = .3398 \\ P(X = 1) &= \frac{\binom{20}{1}\binom{180}{9}}{\binom{200}{10}} = .3974 \\ P(X = 2) &= \frac{\binom{20}{2}\binom{180}{8}}{\binom{200}{10}} = .1975 \end{aligned}$$

Binomial or Hypergeometric?

Boxes contain 2000 items of which 10% are defective.
Find the probability that no more than 2 defectives will be obtained in a sample of size 10.
Let X = no of defectives.

$$P(X \leq 2) = P(X = 0) + P(X = 1) + P(X = 2)$$

With Replacement Sampling

$$\begin{aligned} P(X=0) &= .9^{10} \quad = .3487 \\ P(X=1) &= \binom{10}{1} .1^1 .9^9 = .3874 \\ P(X=2) &= \binom{10}{2} .1^2 .9^8 = .1937 \end{aligned}$$

Without Replacement Sampling

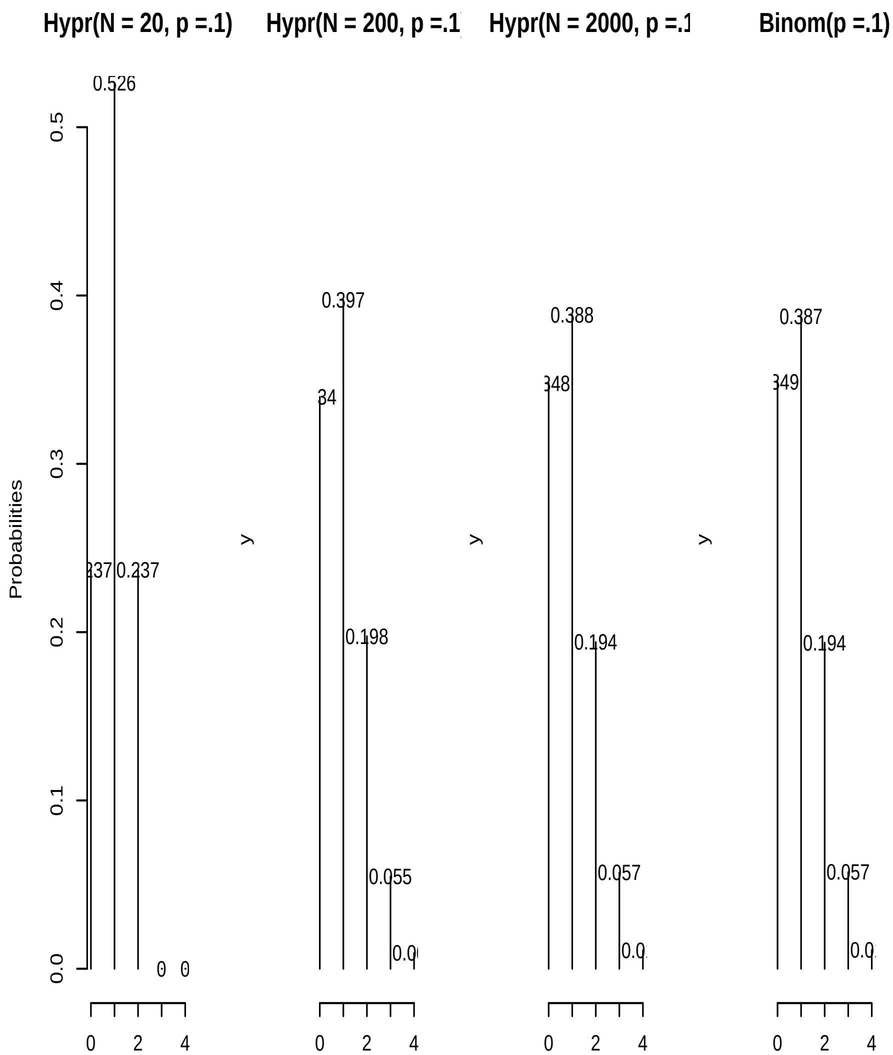
$$\begin{aligned} P(X=0) &= \frac{\binom{1800}{10}}{\binom{2000}{10}} = .3476 \\ P(X=1) &= \frac{\binom{200}{1}\binom{1800}{9}}{\binom{2000}{10}} = .3881 \\ P(X=2) &= \frac{\binom{200}{2}\binom{1800}{8}}{\binom{2000}{10}} = .1939 \end{aligned}$$

Binomial or Hypergeometric?

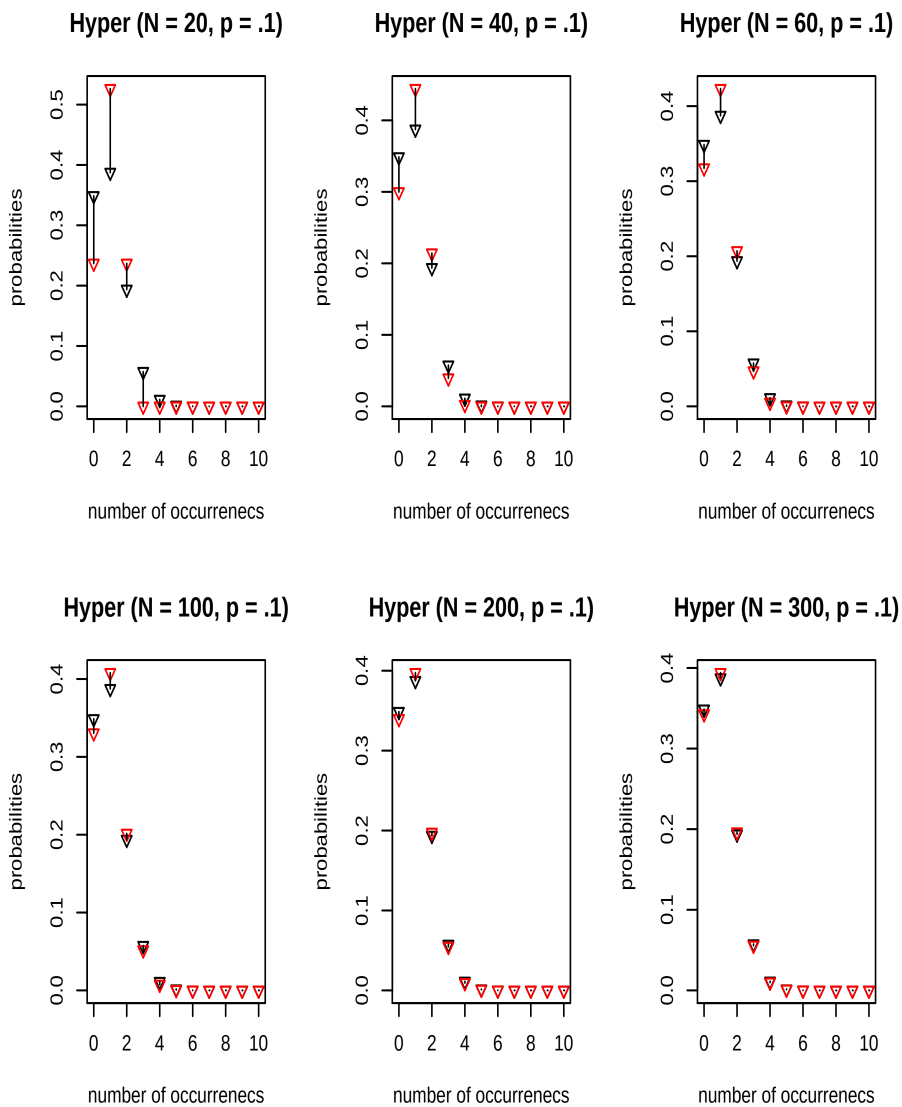
What is the probability of getting no more than 2 defectives in a random sample drawn without replacement from a batch which has 10% defectives?

Batch Size	20	200	2000	Bin. Approx
P(X=0)	.2368	.3398	.3476	.3487
P(X=1)	.5263	.3974	.3881	.3874
P(X=2)	.2368	.1975	.1939	.1937
P(X$\leq$ 2)	.999	.9347	.9296	.9298

Hypergeometric and binomial pdfs with $n = 10$, $p = .1$



Hypergeometric and binomial pdfs with $n = 10$, $p = .1$



As $N \to \infty$, the hypergeometric distribution converges to the binomial.

$$\begin{aligned} \text{Population Size} &= N \\ \text{Proportion of successes} &= p \\ \text{Number of successes in N} &= Np \\ \text{Number of failures} &= N(1-p) \end{aligned}$$

Let X = number of successes in s sample of size n drawn without replacement from N

Np Successes	N(1-p) Failures

Then

$$\begin{aligned} P(X=x) &= \frac{\binom{Np}{x}\binom{N(1-p)}{n-x}}{\binom{N}{x}} \\ &\to \binom{n}{x} p^x q^{n-x} \text{ as } N \to \infty \end{aligned}$$