

## 12.5: The Matching Problem

### Definitions and Notation

#### The Matching Experiment

The *matching experiment* is a random experiment that can be formulated in a number of colorful ways:

- Suppose that  $n$  male-female couples are at a party and that the males and females are randomly paired for a dance. A match occurs if a couple happens to be paired together.
- An absent-minded secretary prepares  $n$  letters and envelopes to send to  $n$  different people, but then randomly stuffs the letters into the envelopes. A match occurs if a letter is inserted in the proper envelope.
- $n$  people with hats have had a bit too much to drink at a party. As they leave the party, each person randomly grabs a hat. A match occurs if a person gets his or her own hat.

These experiments are clearly equivalent from a mathematical point of view, and correspond to selecting a random permutation  $\mathbf{X} = (X_1, X_2, \dots, X_n)$  of the population  $D_n = \{1, 2, \dots, n\}$ . Here are the interpretations for the examples above:

- Number the couples from 1 to  $n$ . Then  $X_i$  is the number of the woman paired with the  $i$ th man.
- Number the letters and corresponding envelopes from 1 to  $n$ . Then  $X_i$  is the number of the envelope containing the  $i$ th letter.
- Number the people and their corresponding hats from 1 to  $n$ . Then  $X_i$  is the number of the hat chosen by the  $i$ th person.

Our modeling assumption, of course, is that  $\mathbf{X}$  is uniformly distributed on the sample space of permutations of  $D_n$ . The number of objects  $n$  is the basic parameter of the experiment. We will also consider the case of sampling *with* replacement from the population  $D_n$ , because the analysis is much easier but still provides insight. In this case,  $\mathbf{X}$  is a sequence of independent random variables, each uniformly distributed over  $D_n$ .

#### Matches

We will say that a *match* occurs at position  $j$  if  $X_j = j$ . Thus, number of matches is the random variable  $N$  defined mathematically by

$$N_n = \sum_{j=1}^n I_j \quad (12.5.1)$$

where  $I_j = \mathbf{1}(X_j = j)$  is the indicator variable for the event of match at position  $j$ . Our problem is to compute the probability distribution of the number of matches. This is an old and famous problem in probability that was first considered by Pierre-Remond Montmort; it is sometimes referred to as *Montmort's matching problem* in his honor.

#### Sampling With Replacement

First let's solve the matching problem in the easy case, when the sampling is with replacement. Of course, this is not the way that the matching game is usually played, but the analysis will give us some insight.

$(I_1, I_2, \dots, I_n)$  is a sequence of  $n$  Bernoulli trials, with success probability  $\frac{1}{n}$ .

Proof

The variables are independent since the sampling is with replacement. Since  $X_j$  is uniformly distributed,  $\mathbb{P}(I_j = 1) = \mathbb{P}(X_j = j) = \frac{1}{n}$ .

The number of matches  $N_n$  has the binomial distribution with trial parameter  $n$  and success parameter  $\frac{1}{n}$ .

$$\mathbb{P}(N_n = k) = \binom{n}{k} \left(\frac{1}{n}\right)^k \left(1 - \frac{1}{n}\right)^{n-k}, \quad k \in \{0, 1, \dots, n\} \quad (12.5.2)$$

Proof

This follows immediately from the previous result on [Bernoulli trials](#).

The mean and variance of the number of matches are

1.  $\mathbb{E}(N_n) = 1$
2.  $\text{var}(N_n) = \frac{n-1}{n}$

Proof

These results follow from the previous result on the [binomial distribution](#) of  $N_n$ . Recall that the binomial distribution with parameters  $n$  and  $p$  has mean  $np$  and variance  $np(1-p)$ .

The distribution of the number of matches converges to the Poisson distribution with parameter 1 as  $n \rightarrow \infty$ :

$$\mathbb{P}(N_n = k) \rightarrow \frac{e^{-1}}{k!} \text{ as } n \rightarrow \infty \text{ for } k \in \mathbb{N} \quad (12.5.3)$$

Proof

This is a special case of the convergence of the binomial distribution to the Poisson. For a direct proof, note that

$$\mathbb{P}(N_n = k) = \frac{1}{k!} \frac{n^{(k)}}{n^k} \left(1 - \frac{1}{n}\right)^{n-k} \quad (12.5.4)$$

But  $\frac{n^{(k)}}{n^k} \rightarrow 1$  as  $n \rightarrow \infty$  and  $\left(1 - \frac{1}{n}\right)^{n-k} \rightarrow e^{-1}$  as  $n \rightarrow \infty$  by a famous limit from calculus.

## Sampling Without Replacement

Now let's consider the case of real interest, when the sampling is without replacement, so that  $\mathbf{X}$  is a random permutation of the elements of  $D_n = \{1, 2, \dots, n\}$ .

### Counting Permutations with Matches

To find the probability density function of  $N_n$ , we need to count the number of permutations of  $D_n$  with a specified number of matches. This will turn out to be easy once we have counted the number of permutations with no matches; these are called *derangements* of  $D_n$ . We will denote the number of permutations of  $D_n$  with exactly  $k$  matches by  $b_n(k) = \#\{N_n = k\}$  for  $k \in \{0, 1, \dots, n\}$ . In particular,  $b_n(0)$  is the number of derangements of  $D_n$ .

The number of derangements is

$$b_n(0) = n! \sum_{j=0}^n \frac{(-1)^j}{j!} \quad (12.5.5)$$

Proof

By the complement rule for counting measure  $b_n(0) = n! - \#(\bigcup_{i=1}^n \{X_i = i\})$ . From the inclusion-exclusion formula,

$$b_n(0) = n! - \sum_{j=1}^n (-1)^{j-1} \sum_{J \subseteq D_n, \#(J)=j} \#\{X_i = i \text{ for all } i \in J\} \quad (12.5.6)$$

But if  $J \subseteq D_n$  with  $\#(J) = j$  then  $\#\{X_i = i \text{ for all } i \in J\} = (n-j)!$ . Finally, the number of subsets  $J$  of  $D_n$  with  $\#(J) = j$  is  $\binom{n}{j}$ . Substituting into the displayed equation and simplifying gives the result.

The number of permutations with exactly  $k$  matches is

$$b_n(k) = \frac{n!}{k!} \sum_{j=0}^{n-k} \frac{(-1)^j}{j!}, \quad k \in \{0, 1, \dots, n\} \quad (12.5.7)$$

Proof

The following is two-step procedure that generates all permutations with exactly  $k$  matches: First select the  $k$  integers that will match. The number of ways of performing this step is  $\binom{n}{k}$ . Second, select a permutation of the remaining  $n-k$  integers with

no matches. The number of ways of performing this step is  $b_{n-k}(0)$ . By the multiplication principle of combinatorics it follows that  $b_n(k) = \binom{n}{k} b_{n-k}(0)$ . Using the result above for [derangements](#) and simplifying gives the results.

## The Probability Density Function

The probability density function of the number of matches is

$$\mathbb{P}(N_n = k) = \frac{1}{k!} \sum_{j=0}^{n-k} \frac{(-1)^j}{j!}, \quad k \in \{0, 1, \dots, n\} \quad (12.5.8)$$

Proof

This follows directly from the result above on [permutations with matches](#), since  $\mathbb{P}(N_n = k) = \#\{N_n = k\} / n!$ .

In the matching experiment, vary the parameter  $n$  and note the shape and location of the probability density function. For selected values of  $n$ , run the simulation 1000 times and compare the empirical density function to the true probability density function.

$$\mathbb{P}(N_n = n - 1) = 0.$$

Proof

A simple *probabilistic* proof is to note that the event is impossible—if there are  $n - 1$  matches, then there must be  $n$  matches. An algebraic proof can also be constructed from the [probability density function](#) of  $N_n$  above.

The distribution of the number of matches converges to the Poisson distribution with parameter 1 as  $n \rightarrow \infty$ :

$$\mathbb{P}(N_n = k) \rightarrow \frac{e^{-1}}{k!} \text{ as } n \rightarrow \infty, \quad k \in \mathbb{N} \quad (12.5.9)$$

Proof

From the power series for the exponential function,

$$\sum_{j=0}^{n-k} \frac{(-1)^j}{j!} \rightarrow \sum_{j=0}^{\infty} \frac{(-1)^j}{j!} = e^{-1} \text{ as } n \rightarrow \infty \quad (12.5.10)$$

So the result follows from the [probability density function](#) of  $N_n$  above.

The convergence is remarkably rapid.

In the matching experiment, increase  $n$  and note how the probability density function stabilizes rapidly. For selected values of  $n$ , run the simulation 1000 times and compare the relative frequency function to the probability density function.

## Moments

The mean and variance of the number of matches could be computed directly from the distribution. However, it is much better to use the representation in terms of indicator variables. The exchangeable property is an important tool in this section.

$$\mathbb{E}(I_j) = \frac{1}{n} \text{ for } j \in \{1, 2, \dots, n\}.$$

Proof

$X_j$  is uniformly distributed on  $D_n$  for each  $j$  so  $\mathbb{P}(I_j = 1) = \mathbb{P}(X_j = x) = \frac{1}{n}$ .

$$\mathbb{E}(N_n) = 1 \text{ for each } n$$

Proof

This follows from the [previous result](#) and basic properties of expected value.

Thus, the expected number of matches is 1, regardless of  $n$ , just as when the sampling is [with replacement](#).

$$\text{var}(I_j) = \frac{n-1}{n^2} \text{ for } j \in \{1, 2, \dots, n\}.$$

Proof

This follows from  $\mathbb{P}(I_j = 1) = \frac{1}{n}$ .

A match in one position would seem to make it more likely that there would be a match in another position. Thus, we might guess that the indicator variables are positively correlated.

For distinct  $j, k \in \{1, 2, \dots, n\}$ ,

$$1. \text{cov}(I_j, I_k) = \frac{1}{n^2(n-1)}$$

$$2. \text{cor}(I_j, I_k) = \frac{1}{(n-1)^2}$$

Proof

Note that  $I_j I_k$  is the indicator variable of the event of a match in position  $j$  and a match in position  $k$ . Hence by the exchangeable property  $\mathbb{P}(I_j I_k = 1) = \mathbb{P}(I_j = 1) \mathbb{P}(I_k = 1 | I_j = 1) = \frac{1}{n} \frac{1}{n-1}$ . As before,  $\mathbb{P}(I_j = 1) = \mathbb{P}(I_k = 1) = \frac{1}{n}$ . The results now follow from standard computational formulas for covariance and correlation.

Note that when  $n = 2$ , the event that there is a match in position 1 is perfectly correlated with the event that there is a match in position 2. This makes sense, since there will either be 0 matches or 2 matches.

$$\text{var}(N_n) = 1 \text{ for every } n \in \{2, 3, \dots\}.$$

Proof

This follows from the previous two results on the [variance](#) and the [covariance](#) of the indicator variables, and basic properties of covariance. Recall that  $\text{var}(N_n) = \sum_{j=1}^n \sum_{k=1}^n \text{cov}(I_j, I_k)$ .

In the matching experiment, vary the parameter  $n$  and note the shape and location of the mean  $\pm$  standard deviation bar. For selected values of the parameter, run the simulation 1000 times and compare the sample mean and standard deviation to the distribution mean and standard deviation.

For distinct  $j, k \in \{1, 2, \dots, n\}$ ,  $\text{cov}(I_j, I_k) \rightarrow 0$  as  $n \rightarrow \infty$ .

Thus, the event that a match occurs in position  $j$  is nearly independent of the event that a match occurs in position  $k$  if  $n$  is large. For large  $n$ , the indicator variables behave nearly like  $n$  Bernoulli trials with success probability  $\frac{1}{n}$ , which of course, is what happens when the sampling is [with replacement](#).

## A Recursion Relation

In this subsection, we will give an alternate derivation of the distribution of the number of matches, in a sense by *embedding* the experiment with parameter  $n$  into the experiment with parameter  $n + 1$ .

The probability density function of the number of matches satisfies the following recursion relation and initial condition:

$$1. \mathbb{P}(N_n = k) = (k+1) \mathbb{P}(N_{n+1} = k+1) \text{ for } k \in \{0, 1, \dots, n\}.$$

$$2. \mathbb{P}(N_1 = 1) = 1.$$

Proof

First, consider the random permutation  $(X_1, X_2, \dots, X_n, X_{n+1})$  of  $D_{n+1}$ . Note that  $(X_1, X_2, \dots, X_n)$  is a random permutation of  $D_n$  if and only if  $X_{n+1} = n+1$  if and only if  $I_{n+1} = 1$ . It follows that

$$\mathbb{P}(N_n = k) = \mathbb{P}(N_{n+1} = k+1 | I_{n+1} = 1), \quad k \in \{0, 1, \dots, n\} \quad (12.5.11)$$

From the definition of conditional probability argument we have

$$\mathbb{P}(N_n = k) = \mathbb{P}(N_{n+1} = k+1) \frac{\mathbb{P}(I_{n+1} = 1 \mid N_{n+1} = k+1)}{\mathbb{P}(I_{n+1} = 1)}, \quad k \in \{0, 1, \dots, n\} \quad (12.5.12)$$

But  $\mathbb{P}(I_{n+1} = 1) = \frac{1}{n+1}$  and  $\mathbb{P}(I_{n+1} = 1 \mid N_{n+1} = k+1) = \frac{k+1}{n+1}$ . Substituting into the last displayed equation gives the recurrence relation. The initial condition is obvious, since if  $n = 1$  we must have one match.

This result can be used to obtain the probability density function of  $N_n$  recursively for any  $n$ .

### The Probability Generating Function

Next recall that the probability generating function of  $N_n$  is given by

$$G_n(t) = \mathbb{E}(t^{N_n}) = \sum_{j=0}^n \mathbb{P}(N_n = j)t^j, \quad t \in \mathbb{R} \quad (12.5.13)$$

The family of probability generating functions satisfies the following differential equations and ancillary conditions:

1.  $G'_{n+1}(t) = G_n(t)$  for  $t \in \mathbb{R}$  and  $n \in \mathbb{N}_+$
2.  $G_n(1) = 1$  for  $n \in \mathbb{N}_+$

Note also that  $G_1(t) = t$  for  $t \in \mathbb{R}$ . Thus, the system of differential equations can be used to compute  $G_n$  for any  $n \in \mathbb{N}_+$ .

In particular, for  $t \in \mathbb{R}$ ,

1.  $G_2(t) = \frac{1}{2} + \frac{1}{2}t^2$
2.  $G_3(t) = \frac{1}{3} + \frac{1}{2}t + \frac{1}{6}t^3$
3.  $G_4(t) = \frac{3}{8} + \frac{1}{3}t + \frac{1}{4}t^2 + \frac{1}{24}t^4$

For  $k, n \in \mathbb{N}_+$  with  $k < n$ ,

$$G_n^{(k)}(t) = G_{n-k}(t), \quad t \in \mathbb{R} \quad (12.5.14)$$

Proof

This follows from [differential equation](#) for the PGF given above.

For  $n \in \mathbb{N}_+$ ,

$$\mathbb{P}(N_n = k) = \frac{1}{k!} \mathbb{P}(N_{n-k} = 0), \quad k \in \{0, 1, \dots, n-1\} \quad (12.5.15)$$

Proof

This follows from the [previous result](#) and basic properties of generating functions.

### Examples and Applications

A secretary randomly stuffs 5 letters into 5 envelopes. Find each of the following:

1. The number of outcomes with exactly  $k$  matches, for each  $k \in \{0, 1, 2, 3, 4, 5\}$
2. The probability density function of the number of matches.
3. The covariance and correlation of a match in one envelope and a match in another envelope.

Answer

1.	$k$	0	1	2	3	4	5
	$b_5(k)$	44	45	20	10	0	1

2.	$k$	0	1	2	3	4	5

$\mathbb{P}(N_5 = k)$	0.3667	0.3750	0.1667	0.0833	0	0.0083
-----------------------	--------	--------	--------	--------	---	--------

3. Covariance:  $\frac{1}{100}$ , correlation  $\frac{1}{16}$

Ten married couples are randomly paired for a dance. Find each of the following:

1. The probability density function of the number of matches.
2. The mean and variance of the number of matches.
3. The probability of at least 3 matches.

Answer

1. $k$	$\mathbb{P}(N_{10} = k)$
0	$\frac{16\,481}{44\,800} \approx 0.3678795$
1	$\frac{16\,687}{45\,360} \approx 0.3678792$
2	$\frac{2119}{11\,520} \approx 0.1839410$
3	$\frac{103}{1680} \approx 0.06130952$
4	$\frac{53}{3456} \approx 0.01533565$
5	$\frac{11}{3600} \approx 0.003055556$
6	$\frac{1}{1920} \approx 0.0005208333$
7	$\frac{1}{15\,120} \approx 0.00006613757$
8	$\frac{1}{80\,640} \approx 0.00001240079$
9	0
10	$\frac{1}{3\,628\,800} \approx 2.755732 \times 10^{-7}$

2.  $\mathbb{E}(N_{10}) = 1$ ,  $\text{var}(N_{10}) = 1$

3.  $\mathbb{P}(N_{10} \geq 3) = \frac{145\,697}{1\,814\,400} \approx 0.08030037$

In the matching experiment, set  $n = 10$ . Run the experiment 1000 times and compare the following for the number of matches:

1. The true probabilities
2. The relative frequencies from the simulation
3. The limiting Poisson probabilities

Answer

1. See part (a) of the [previous problem](#).

3. $k$	$e^{-1} \frac{1}{k!}$
0	0.3678794
1	0.3678794
2	0.1839397
3	0.06131324
4	0.01532831
5	0.003065662
6	0.0005109437
7	0.00007299195

8	$9.123994 \times 10^{-6}$
9	$1.013777 \times 10^{-6}$
10	$1.013777 \times 10^{-7}$

This page titled [12.5: The Matching Problem](#) is shared under a [CC BY 2.0](#) license and was authored, remixed, and/or curated by [Kyle Siegrist](#) ([Random Services](#)) via [source content](#) that was edited to the style and standards of the LibreTexts platform.