

A GENERALIZATION OF THE MATCHING DISTRIBUTION

Stefan Niermann
Institut für Quantitative Wirtschaftsforschung
Universität Hannover
30167 Hannover

The Matching Distribution converges to a Poisson Distribution with $\lambda = 1$ as the parameter n converges to infinity. A generalization of the Matching Distribution is proposed. The properties of this Generalized Matching Distribution (GMD) turn out to be analogical to the case with $\lambda = 1$.

1 INTRODUCTION

Let us consider the *Matching Problem*, where we are interested in the number of permutations of the integers $1, 2, \dots, n$ having the property that exactly x elements remain fix. An integer that remains in its natural position is called a match. The random variable X_n denotes the number of matches. It is assumed that all $n!$ possible permutations occur with the same probability.

Problems of this kind arise in a variety of applications. The Matching Problem occurs when n cards (numbered $1, 2, \dots, n$) are laid out onto n fixed positions, also marked $1, 2, \dots, n$, with each position receiving one card. A match occurs if the card i is placed in position i . Montmort (1708) was the first to give a solution to this problem for $n=13$.

This problem might also be framed in the following scenario. A number of children write letters to their fathers who serve in a war. The letters are randomly delivered and we are interested in the number of correctly delivered letters. Notably, on the average one letter will be delivered correctly – no matter how many letters are sent. Krämer (1995, p. 24) refers to such an event during the Gulf War. Therefore, the Matching Distribution serves as an illustration for the fact, that even if the probability of an event to happen to a certain person may be very low, the probability of this event happening to any person can be considerably large. This result is important in the statistical analysis of coincidences.

The Matching Distribution also serves as a reference distribution if one wishes to decide, whether systematic matching performs significantly better than random matching.

Is it really possible to identify individuals just by looking at their horoscopes? Is it possible to identify certain aspects of personality by looking at the handwriting?

These questions might be affirmed if the number of observed matches in an experiment is significantly higher than the number of expected random matches. More examples and applications in the Social Sciences are given by Vernon (1936).

2 THE DISTRIBUTION OF THE NUMBER OF MATCHES

The distribution of X_n (the Matching Distribution) is given by

<p>Distribution of the Number of Matches</p> $P(X_n = x) = \frac{1}{x!} \cdot \sum_{i=0}^{n-x} \frac{(-1)^i}{i!}$ <p>for $x = 0, \dots, n$.</p>
--

For arbitrary n we have $E[X] = V(X) = 1$. The derivation of the pdf and the proof of these properties can be found in Dwass (1970, pp. 52.).

Looking at the series representation of the exponential function for $x = -1$:

$e^{-1} = \sum_{i=0}^{\infty} (-1)^i \cdot \frac{1}{i!} = \sum_{i=2}^{\infty} (-1)^i \cdot \frac{1}{i!}$, it can be seen, that the distribution of the number of matches converges to a Poisson Distribution with $\lambda = 1$. Irwin (1955) showed numerically that for $\lambda = 10$ there is at most a discrepancy in the fifth decimal place.

3 The Generalized Matching Distribution

We are interested in the question, whether there is a family of distributions – depending on two parameters, λ and n , which converges to a Poisson Distribution with parameter λ for large n .

Analogous to the above case we claim the expectation and the variance both equal λ .

Considering the structure of the Matching random variable and the relation to the Poisson Distribution, it seems plausible to look at expressions of the following form:

$P(X = x) = f(x; n, \lambda) = \frac{\lambda^x}{x!} \sum_{i=0}^{n-x} \frac{(-\lambda)^i}{i!} = \frac{\lambda^x}{x!} \cdot S_{n-x}$ <p>for $0 < \lambda \leq 1, x = 0, 1, \dots, n$.</p>

For $0 < \lambda \leq 1$, a valid probability function emerges. This random variable also has the desired properties (convergence to a Poisson random variable, $E[X_{n,\lambda}] = V(X_{n,\lambda}) = \lambda$).

First, it will be shown that the proposed probability function is valid.

- $0 \leq f(x; n, \lambda) \leq 1$

The proposed probabilities are a product of the factor $\frac{\lambda^x}{x!}$ and the partial sum $S_{n-x} = \sum_{i=0}^{n-x} \frac{(-\lambda)^i}{i!}$. The first factor is bounded by 0 and 1. For the partial sum we have $S_0 = 1$ and $S_1 = 1 - \lambda$, or $0 < S_0 \leq 1$ and $0 < S_1 \leq 1$. The partial sums can be calculated recursively.

$$\begin{aligned} \text{k even} &\Rightarrow S_k = S_{k-1} + \frac{\lambda^k}{k!} > S_{k-1} \geq 0 \\ &\Rightarrow S_k = S_{k-2} - \frac{\lambda^{k-1}}{(k-1)!} + \frac{\lambda^k}{k!} < S_{k-2} \leq 1 \\ \text{k odd} &\Rightarrow S_k = S_{k-1} - \frac{\lambda^k}{k!} < S_{k-1} \leq 1 \\ &\Rightarrow S_k = S_{k-2} + \frac{\lambda^{k-1}}{(k-1)!} - \frac{\lambda^k}{k!} > S_{k-2} \geq 0 \end{aligned}$$

Hence, we get $0 < S_k \leq 1$ for all natural numbers k.

- The fact that the sum of all probabilities equals one will be proved by induction. For $n = 1$ we get $f(0; 1, \lambda) = 1 - \lambda$ and $f(1; 1, \lambda) = \lambda$. For a GMD with $0 < \lambda \leq 1$ and $n + 1$ follows

$$\begin{aligned} \sum_{x=0}^{n+1} f(x; n+1, \lambda) &= \sum_{x=0}^n f(x; n+1, \lambda) + f(n+1; n+1, \lambda) \\ &= \sum_{x=0}^n \frac{\lambda^x}{x!} \cdot \left(\sum_{i=0}^{n+1-x} \frac{(-\lambda)^i}{i!} \right) + \frac{\lambda^{n+1}}{(n+1)!} \\ &= \sum_{x=0}^n \frac{\lambda^x}{x!} \cdot \sum_{i=0}^{n-x} \frac{(-\lambda)^i}{i!} + \sum_{x=0}^n \frac{\lambda^x}{x!} \cdot \frac{(-\lambda)^{n+1-x}}{(n+1-x)!} + \frac{\lambda^{n+1}}{(n+1)!} \\ &= \sum_{x=0}^n f(x; n, \lambda) + \sum_{x=0}^n \frac{\lambda^x \cdot \lambda^{n+1-x} \cdot (-1)^{n+1-x}}{x! \cdot (n+1-x)!} + \frac{\lambda^{n+1}}{(n+1)!} \\ &= 1 + \frac{\lambda^{n+1}}{(n+1)!} \left[\sum_{x=0}^n \binom{n+1}{x} \cdot (-1)^{n+1-x} + 1 \right] \\ &= 1 + \frac{\lambda^{n+1}}{(n+1)!} \sum_{x=0}^{n+1} \binom{n+1}{x} \cdot (-1)^{n+1-x} \cdot 1^x \\ &= 1 + (1-1)^{n+1} = 1. \end{aligned}$$

$E[X_{n,\lambda}] = V(X_{n,\lambda}) = \lambda$ will also be proved by induction.

- The expected value of a GMD with $n = 1$ equals $0 \cdot (1 - \lambda) + 1 \cdot \lambda = \lambda$.

$$\begin{aligned} E[X_{n+1}] &= \sum_{x=0}^{n+1} x \cdot f(x; n+1, \lambda) = \sum_{x=1}^{n+1} \frac{\lambda^x}{(x-1)!} \cdot \sum_{i=0}^{n+1-x} \frac{(-\lambda)^i}{i!} \\ &= \sum_{y=1}^n \frac{\lambda^{y+1}}{y!} \cdot \sum_{i=0}^{n-y} \frac{(-\lambda)^i}{i!} = \lambda \cdot \sum_{y=0}^n f(y; n, \lambda) = \lambda \end{aligned}$$

- Similar, it can be shown that the variance of a GMD equals λ for $n > 1$. For $n = 2$

$$E[X_{2,\lambda}^2] = 1 \cdot f(1, 2, \lambda) + 4 \cdot f(2, 2, \lambda) = \lambda^2 + \lambda.$$

And for a GMD with λ and $(n+1)$

$$\begin{aligned} E[X_{n+1}^2] &= \sum_{x=0}^{n+1} x^2 \cdot f(x; n+1, \lambda) = \sum_{x=1}^{n+1} x \cdot \frac{\lambda^x}{(x-1)!} \cdot \sum_{i=0}^{n+1-x} \frac{(-\lambda)^i}{i!} \\ &= \sum_{x=0}^n (x+1) \cdot \frac{\lambda^{x+1}}{x!} \cdot \sum_{i=0}^{n-x} \frac{(-\lambda)^i}{i!} \\ &= \lambda \cdot \sum_{x=0}^n (x+1) f(x; n, \lambda) \\ &= \lambda \cdot E[X_n + 1] = \lambda \cdot (\lambda + 1). \end{aligned}$$

Thus, $V(X_{n,\lambda}) = \lambda$ for $n \geq 2$ and $0 < \lambda \leq 1$.

When the above cases are combined using the convolution method, it is possible to derive GMDs for any $\lambda > 0$.

4 EXAMPLES AND SPEED OF CONVERGENCE

The GMD with $0 < \lambda < 1$ is an appropriate model for situations, in which each match of a standard matching distribution ($\lambda = n = 1$) is counted with probability λ . In the case of the n letters from children to soldiers mentioned above, we additionally assume that some letters get lost and are not delivered at all. If each letter is delivered with probability λ , then the Generalized Matching Distribution with n and λ is an appropriate model for the number of counted matches.

The generalized Matching random variable converges to a Poisson random variable very fast. This is obviously due to the fact that the power series expansion of the exponential function converges very fast. The speed of convergence is even faster for small values of λ . To demonstrate this fact we will consider four Matching Distributions and compare them with the corresponding Poisson Distributions.

x	$f(x, n, 0.5)$			$f(x, n, 1)$		
	$n = 10$	$n = 5$	$\frac{e^{-0.5 \cdot 0.5^x}}{x!}$	$n = 10$	$n = 5$	$\frac{e^{-1}}{x!}$
0	0.6065	0.6065	0.6065	0.3679	0.3667	0.3679
1	0.3033	0.3034	0.3033	0.3679	0.3750	0.3679
2	0.0758	0.0755	0.0758	0.1839	0.1667	0.1839
3	0.0126	0.0130	0.0126	0.0613	0.0833	0.0613
4	0.0016	0.0013	0.0016	0.0153	0.0000	0.0153
5	0.0002	0.0003	0.0002	0.0031	0.0083	0.0031
6	0.0000		0.0000	0.0005		0.0005
7	0.0000		0.0000	0.0001		0.0001
8	0.0000		0.0000	0.0000		0.0000

It can be seen that the GMD converges to a Poisson Distribution very fast, too.

5 REFERENCES

Alt, F. B. (1985), The Matching Problem, Encyclopedia of Statistical Sciences, Vol. 5, S. Kotz, N. L. Johnson and C. B. Read (editors), New York, Wiley, 293-296.

Dwass, M. (1970), Probability and Statistics: An Undergraduate Course, W. A. Benjamin, INC., Menlo Park, California.

Irwin, J. O. (1955), A unified Derivation of some well-known Frequency Distributions of Interest in Biometry and Statistics, Journal of the Royal Statistical Society, Series A, Vol. 100, 266-268.

Krämer, W. (1995), Denkste! Trugschlüsse aus der Welt des Zufalls und der Zahlen, Campus Verlag, Frankfurt and New York.

Montmort, P. R. de (1713), Essay d'Analyse sur les Jeux de Hazard, 2nd ed. Paris. (Reprint: Chelsea House, New York, 1981)

Vernon, P. E. (1936), The Matching Method applied to Investigations of Personality, Psychological Bulletin, Vol. 33, 149-177.