Hypergeometric distribution

The hypergeometric distribution is used to calculate probabilities when sampling without replacement. For example, suppose you first randomly sample one card from a deck of 52. Then, without putting the card back in the deck you sample a second and then (again without replacing cards) a third. Given this sampling procedure, what is the probability that exactly two of the sampled cards will be aces (4 of the 52 cards in the deck are aces). You can calculate this probability using the following formula based on the hypergeometric distribution:

:the following formula based on the hypergeometric distribution

where

$$p = \frac{({}_kC_x)({}_{(N-k)}C_{(n-x)})}{{}_NC_n}$$

k is the number of "successes" in the population
x is the number of "successes" in the sample
N is the size of the population
n is the number sampled
p is the probability of obtaining exactly x successes
${}_kC_x$ is the number of combinations of k things taken x at a time

The mean and standard deviation of the hypergeometric distribution are:

$$\text{mean} = \frac{(n)(k)}{N}$$

$$\text{sd} = \sqrt{\frac{(n)(k)(N-k)(N-n)}{N^2(N-1)}}$$

Theorem 5.2: The mean and variance of the hypergeometric distribution $h(x; N, n, k)$ are

$$\mu = \frac{nk}{N} \text{ and } \sigma^2 = \frac{N-n}{N-1} \cdot n \cdot \frac{k}{N}\left(1 - \frac{k}{N}\right).$$

In this example, k = 4 because there are four aces in the deck, x = 2 because the problem asks about the probability of getting two aces, N = 52 because there are 52 cards in a deck, and n = 3 because 3 cards were sampled. Therefore,

$$p = \frac{({}_4C_2)({}_{(52-4)}C_{(3-2)})}{{}_{52}C_3}$$

$$p = \frac{\frac{4!}{2!2!}\frac{48!}{47!1!}}{\frac{52!}{49!3!}} = 0.013$$

EX

A wallet contains 3 $100 bills and 5 $1 bills. You randomly choose 4 bills. What is the probability that you will choose exactly 2 $100 bills?

$$P(X=2)=(3C2)(5C2)/8C4$$

$$=0.42856$$

Example 5.8: A particular part that is used as an injection device is sold in lots of 10. The producer deems a lot acceptable if no more than one defective is in the lot. A sampling plan involves random sampling and testing 3 of the parts out of 10. If none of the 3 is defective, the lot is accepted. Comment on the utility of this plan.

Solution: Let us assume that the lot is truly unacceptable (i.e., that 2 out of 10 parts are defective). The probability that the sampling plan finds the lot acceptable is

$$P(X = 0) = \frac{\binom{2}{0}\binom{8}{3}}{\binom{10}{3}} = 0.467.$$

Thus, if the lot is truly unacceptable, with 2 defective parts, this sampling plan will allow acceptance roughly 47% of the time. As a result, this plan should be considered faulty.

Example 5.9: Lots of 40 components each are deemed unacceptable if they contain 3 or more defectives. The procedure for sampling a lot is to select 5 components at random and to reject the lot if a defective is found. What is the probability that exactly 1 defective is found in the sample if there are 3 defectives in the entire lot?

Solution: Using the hypergeometric distribution with $n = 5$, $N = 40$, $k = 3$, and $x = 1$, we find the probability of obtaining 1 defective to be

$$h(1; 40, 5, 3) = \frac{\binom{3}{1}\binom{37}{4}}{\binom{40}{5}} = 0.3011.$$

Once again, this plan is not desirable since it detects a bad lot (3 defectives) only about 30% of the time.

Example 5.11: Find the mean and variance of the random variable of Example 5.9

Solution: Since Example 5.9 was a hypergeometric experiment with $N = 40$, $n = 5$, and $k = 3$, by Theorem 5.2, we have

$$\mu = \frac{(5)(3)}{40} = \frac{3}{8} = 0.375,$$

and

$$\sigma^2 = \left(\frac{40-5}{39}\right)(5)\left(\frac{3}{40}\right)\left(1 - \frac{3}{40}\right) = 0.3113.$$

Poisson Distribution and the Poisson Process

Experiments yielding numerical values of a random variable X, the number of outcomes occurring during a given time interval or in a specified region, are called **Poisson experiments**. The given time interval may be of any length, such as a minute, a day, a week, a month, or even a year. For example, a Poisson experiment can generate observations for the random variable X representing the number of telephone calls received per hour by an office, the number of days school is closed due to snow during the winter, or the number of games postponed due to rain during a baseball season. The specified region could be a line segment, an area, a volume, or perhaps a piece of material. In such instances, X might represent the number of field mice per acre, the number of bacteria in a given culture, or the number of typing errors per page. A Poisson experiment is derived from the **Poisson process** and possesses the following properties.

Properties of the Poisson Process

1. The number of outcomes occurring in one time interval or specified region of space is independent of the number that occur in any other disjoint time interval or region. In this sense we say that the Poisson process has no memory.
2. The probability that a single outcome will occur during a very short time interval or in a small region is proportional to the length of the time interval or the size of the region and does not depend on the number of outcomes occurring outside this time interval or region.
3. The probability that more than one outcome will occur in such a short time interval or fall in such a small region is negligible.

The number X of outcomes occurring during a Poisson experiment is called a **Poisson random variable**, and its probability distribution is called the **Poisson**

distribution. The mean number of outcomes is computed from $\mu = \lambda t$, where t is the specific "time," "distance," "area," or "volume" of interest. Since the probabilities depend on λ, the rate of occurrence of outcomes, we shall denote them by $p(x; \lambda t)$. The derivation of the formula for $p(x; \lambda t)$, based on the three properties of a Poisson process listed above, is beyond the scope of this book. The following formula is used for computing Poisson probabilities.

The probability distribution of the Poisson random variable X, representing the number of outcomes occurring in a given time interval or specified region denoted by t, is

$$p(x; \lambda t) = \frac{e^{-\lambda t}(\lambda t)^x}{x!}, \quad x = 0, 1, 2, \ldots,$$

where λ is the average number of outcomes per unit time, distance, area, or volume and $e = 2.71828\ldots$.

The average (mean) number of outcomes (mean of X) in the time interval or region t is:

$$\mu = \lambda t$$

The random variable X is called a Poisson random variable with parameter μ ($\mu=\lambda t$), and we write X~Poisson(μ), if its probability distribution is given by:

$$f(x) = P(X = x) = p(x;\mu) = \begin{cases} \dfrac{e^{-\mu}\mu^{x}}{x!} & ; \quad x = 0, 1, 2, 3, \ldots \\ 0 & ; \quad otherwise \end{cases}$$

The mean and the variance of the Poisson distribution Poisson(x;μ) are:

$$\mu = \lambda t$$

$$\sigma^2 = \mu = \lambda t$$

Note:

- λ is the average (mean) of the distribution in the unit time (t=1).
- If X=The number of calls received in a month (unit time t=1 month) and X~Poisson(λ), then:

(i) Y = number of calls received in a year.

Y ~ Poisson (μ); μ=12λ (t=12)

(ii) W = number of calls received in a day.

W ~ Poisson (μ); μ=λ/30 (t=1/30)

Example:

Suppose that the number of typing errors per page has a Poisson distribution with average 6 typing errors.

(1) What is the probability that in a given page:

(i) The number of typing errors will be 7?

(ii) The number of typing errors will at least 2?

(2) What is the probability that in 2 pages there will be 10 typing errors?

(3) What is the probability that in a half page there will be no typing errors?

Solution: **X = number of typing errors per page.**

X ~ Poisson (6) **(t=1, λ=6, μ=λt=6)**

$$f(x) = P(X=x) = p(x;6) = \frac{e^{-6}6^x}{x!}\ ; \quad x = 0, 1, 2, \dots$$

(i) $f(7) = P(X=7) = p(7;6) = \frac{e^{-6}6^7}{7!} = 0.13768$

(ii) $P(X \geq 2) = P(X=2) + P(X=3) + \dots = \sum_{x=2}^{\infty} P(X=x)$

$P(X \geq 2) = 1 - P(X<2) = 1 - [P(X=0) + P(X=1)]$

$= 1 - [f(0) + f(1)] = 1 - \left[\frac{e^{-6}6^0}{0!} + \frac{e^{-6}6^1}{1!}\right]$

$= 1 - [0.00248 + 0.01487]$

$= 1 - 0.01735 = 0.982650$

(2) **X = number of typing errors in 2 pages**

X ~ Poisson(12) **(t=2, λ=6, μ=λt=12)**

$$f(x) = P(X=x) = p(x;12) = \frac{e^{-12}12^x}{x!} : \quad x = 0, 1, 2 \dots$$

$$f(10) = P(X=10) = \frac{e^{-12}12^{10}}{10} = 0.1048$$

(3) X = number of typing errors in a half page.

X ~ Poisson (3) **(t=1/2, λ=6, $\mu=\lambda t$=6/2=3)**

$$f(x) = P(X = x) = p(x;3) = \frac{e^{-3}\,3^x}{x!}; \quad x = 0, 1, 2\ldots$$

$$f(0) = P(X = 0) = \frac{e^{-3}(3)^0}{0!} = 0.0497871$$

Example 5.20: read: During a laboratory experiment the average number of radioactive particles (الجزيئات المشعة) passing through a counter in **1 millisecond is** 4. What is the probability that 6 particles enter the counter in a given millisecond?

$$p(6;4) = \frac{e^{-4}(4)^6}{6!} = 0.1042$$

Example 5.18: Ten is the average number of oil tankers arriving each day at a certain port. The facilities at the port can handle at most 15 tankers per day. What is the probability that on a given day tankers have to be turned away?

$$P(X > 15) = 1 - P(X \le 15) = 1 - \sum_{x=0}^{15} p(x;10) = 1 - 0.9513 = 0.0487$$

Table A.2 (continued) Poisson Probability Sums $\sum_{x=0}^{r} p(x;\mu)$

	μ								
r	**10.0**	**11.0**	**12.0**	**13.0**	**14.0**	**15.0**	**16.0**	**17.0**	**18.0**
0	0.0000	0.0000	0.0000						
1	0.0005	0.0002	0.0001	0.0000	0.0000				
2	0.0028	0.0012	0.0005	0.0002	0.0001	0.0000	0.0000		
3	0.0103	0.0049	0.0023	0.0011	0.0005	0.0002	0.0001	0.0000	0.0000
4	0.0293	0.0151	0.0076	0.0037	0.0018	0.0009	0.0004	0.0002	0.0001
5	0.0671	0.0375	0.0203	0.0107	0.0055	0.0028	0.0014	0.0007	0.0003
6	0.1301	0.0786	0.0458	0.0259	0.0142	0.0076	0.0040	0.0021	0.0010
7	0.2202	0.1432	0.0895	0.0540	0.0316	0.0180	0.0100	0.0054	0.0029
8	0.3328	0.2320	0.1550	0.0998	0.0621	0.0374	0.0220	0.0126	0.0071
9	0.4579	0.3405	0.2424	0.1658	0.1094	0.0699	0.0433	0.0261	0.0154
10	0.5830	0.4599	0.3472	0.2517	0.1757	0.1185	0.0774	0.0491	0.0304
11	0.6968	0.5793	0.4616	0.3532	0.2600	0.1848	0.1270	0.0847	0.0549
12	0.7916	0.6887	0.5760	0.4631	0.3585	0.2676	0.1931	0.1350	0.0917
13	0.8645	0.7813	0.6815	0.5730	0.4644	0.3632	0.2745	0.2009	0.1426
14	0.9165	0.8540	0.7720	0.6751	0.5704	0.4657	0.3675	0.2808	0.2081
15	0.9513	0.9074	0.8444	0.7636	0.6694	0.5681	0.4667	0.3715	0.2867
16	0.9730	0.9441	0.8987	0.8355	0.7559	0.6641	0.5660	0.4677	0.3751
17	0.9857	0.9678	0.9370	0.8905	0.8272	0.7489	0.6593	0.5640	0.4686
18	0.9928	0.9823	0.9626	0.9302	0.8826	0.8195	0.7423	0.6550	0.5622
19	0.9965	0.9907	0.9787	0.9573	0.9235	0.8752	0.8122	0.7363	0.6509
20	0.9984	0.9953	0.9884	0.9750	0.9521	0.9170	0.8682	0.8055	0.7307
21	0.9993	0.9977	0.9939	0.9859	0.9712	0.9469	0.9108	0.8615	0.7991
22	0.9997	0.9990	0.9970	0.9924	0.9833	0.9673	0.9418	0.9047	0.8551
23	0.9999	0.9995	0.9985	0.9960	0.9907	0.9805	0.9633	0.9367	0.8989
24	1.0000	0.9998	0.9993	0.9980	0.9950	0.9888	0.9777	0.9594	0.9317
25		0.9999	0.9997	0.9990	0.9974	0.9938	0.9869	0.9748	0.9554
26		1.0000	0.9999	0.9995	0.9987	0.9967	0.9925	0.9848	0.9718
27			0.9999	0.9998	0.9994	0.9983	0.9959	0.9912	0.9827
28			1.0000	0.9999	0.9997	0.9991	0.9978	0.9950	0.9897
29				1.0000	0.9999	0.9996	0.9989	0.9973	0.9941
30					0.9999	0.9998	0.9994	0.9986	0.9967
31					1.0000	0.9999	0.9997	0.9993	0.9982
32						1.0000	0.9999	0.9996	0.9990
33							0.9999	0.9998	0.9995
34							1.0000	0.9999	0.9998
35								1.0000	0.9999
36									0.9999
37									1.0000