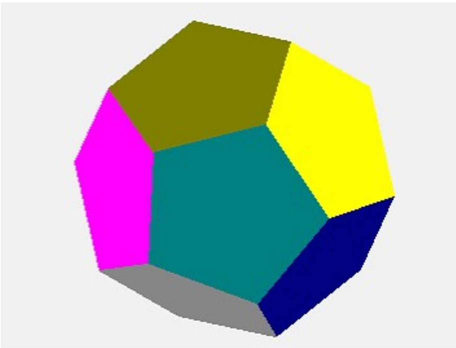
Probability Theory

An introduction and beyond

Chapter 4

Probability distributions

This is an article from my home page: www.olewitthansen.dk



Ole Witt-Hansen (2017) 1997

Contents

1. Repetition of an experiment

We shall now look at a situation where an experiment is repeated a certain number of times and where each performance is independent of the preceding experiments.
When doing the series of experiments, we are however only interested in whether a certain event occurs or not.
More specifically we take as an example a repeated throw with a dice and focus on the event that the dice shows six eyes or not. This choice of example, however, makes no constraint on the general results, since it can readily be generalized to a repetition of any experiment which complies with the demand of independence.
If the experiment is repeated n times, the probability distribution is the determination of the probabilities $P(0), P(1)...P(n)$, that event in question occurs $0…n$ times.

2. An example of the Binomial Distribution

The binomial distribution is a mathematical model of a repetition of an experiment n times, where the individual experiments are independent of each other. The aim is to establish the probabilities that a specific event occurs $0…n$ times.

Before we enter the general formulas, we shall consider the, where a dice is thrown 12 times, calculating the probability that the dice shows six eyes exactly two times.

Throwing a dice once, we have a finite probability field (U_1, P_1) with the outcome space $U_1 = \{1,2,3,4,5,6\}$.
If A is the event $\{6\}$ and $\overline{A} = \{1,2,3,4,5\}$ is the complementary event, then obviously $P_1(A) = 1/6$ and $P_1(\overline{A}) = 5/6$.
If we imagine that the experiment is carried out 12 times, then the outcome space U consists of the outcomes in each of the 12 repetitions of the outcome space U_1.

$U = (u_1, u_2, u_3, ... u_{11}, u_{12})$, where each of the outcomes u_i is one of the outcomes $\{1\},…,\{6\}$. The outcome space has 6^{12} elements, namely 6 for each repetition, and this is often written formally as: $U = U_1 x U_1 x ... x U_1 = U_1^{12}$.

We shall first look at the event A_4: The dice shows six eyes in the 4th throw.
As the 4th throw is independent of the other 11 throws, we have for the probability of that event $P(A_4) = P_1(6) = 1/6$. Correspondingly we have for the event A_7: The dice shows 6 eyes in the 7. throw that: $P(A_7) = P_1(6) = 1/6$.

Then we look at the event $A_4 \cap A_7$. The dice shows 6 eyes both in the 4th and the 7th throw.
We should notice, (and that is important), that we have assumed that the outcomes of the throws are *independent* of each other, and consequently we may find the probability of the "*both…and*" event multiplying the probabilities of the two events.

$$P(A_4 \cap A_7) = P(A_4) \cdot P(A_7) = \frac{1}{6} \cdot \frac{1}{6} = \frac{1}{36}$$

Using the same arguments, we may determine the probability of the event:
$A_4 \cap A_5 \cap A_7 \cap \overline{A}_8 \cap \overline{A}_{10}$, as $P(A_4 \cap A_5 \cap A_7 \cap \overline{A}_8 \cap \overline{A}_{10}) = 1/6 \cdot 1/6 \cdot 1/6 \cdot 5/6 \cdot 5/6 = (1/6)^3 \cdot (5/6)^2$

Namely the probability 1/6, for each time we get 6 eyes, and the probability 5/6 for each time we do not.
We are interested in the event: Exactly two times we get 6 eyes in 12 throws. One possibility could be 6 eyes in the first and the second throw, and not 6 eyes in the following 10 throws. The probability of this specific event is according to the considerations above:

$$P(\textit{6 eyes in the } 1. \textit{ and } 2. \textit{ throw and not 6 eyes in the } 3\text{ -}12 \textit{ throw}) = (1/6)^2 \cdot (5/6)^{10}$$

The two outcomes where we get 6 eyes could however also be in the 4th and the 7th throw or in any pair of numbers selected from the12 repetitions. For the event: That the dice shows 6 eyes in exactly two out of the 12 throws, it is immaterial in which pair of throws it occurs.

To find the probability of the event: The dice shows 6 eyes exactly two times in 12 throws, we need to add the probabilities of the outcomes, where this is the case. Since they have the same probability, we only need to find the number of such events.

We must therefore ask the question in how many different ways can one select 2 places out of twelve. This is however the same question that we have encountered before: The number of different ways to select a *q-subset* from a *n-set,* and the answer is well known as:

$$C(n,q) = \frac{n!}{(n-q)!q!}$$

We may then write the sought probability:

$$P(\textit{The dice shows 6 eyes exactly two times in } 12 \textit{ attempts}) = C(12,2) * (1/6)^2 * (5/6)^{10}$$

For this reason the numbers $C(n,q)$ are often called *binomial coefficients*, and in probability theory they are almost always written with the symbol:

$$\binom{n}{q} \qquad \text{Being the same as} \qquad C(n,q)$$

The new symbol is read (in Denmark) as: "*n over q*" .(Not to be confused with "n divided by *q*")

The probability that an event occurs exactly *q* times in *n* attempts, is called for a *binomial probability*, and the probabilities together are called the *binomial distribution*.

If $X = 2$ is the number of times the event "6 eyes" occur, and we use the new symbol for the binomial coefficients, we may write:

$$P(X=2) = \binom{12}{2}\left(\frac{1}{6}\right)^2\left(\frac{5}{6}\right)^{10} = \binom{12}{2}\left(\frac{1}{6}\right)^2\left(1-\frac{1}{6}\right)^{10}$$

The probability can be evaluated to: $P(X = 2) = 0.2961$.

In exactly the same manner we can establish the probabilities: $P(X = j);\ \ j = 0, 1, 2, ..., 12$.

$$P(X = j) = \binom{12}{j}\left(\frac{1}{6}\right)^{j}\left(\frac{5}{6}\right)^{12-j} = \binom{12}{j}\left(\frac{1}{6}\right)^{j}\left(1-\frac{1}{6}\right)^{12-j} \quad ; j = 0, 1, 2, ..., 12$$

In this manner we have established a probability field with an outcome space $U = \{0, 1, 2, ..., 12\}$, corresponding to that the dice shows 6 eyes exactly 0, 1, …,12 times. The probability field is not symmetric, however, since (most of) the probabilities are different

Although it would otherwise be surprising, we have not formally proved that the probabilities above form a probability field, that is $0 \le P(u) \le 1$ and $\sum_{u \in U} P(u) = 1$. However, this will be done below, when we have treated the general form of the binomial distribution.

3. The general binomial distribution

We now generalize to the case, where we have n independent versions of an experiment.
We are interested in whether an event A occurs in the j'th attempt or not. As usual in probability theory, we write the complementary event to an event A as $\overline{A}$. We put $P(A) = p$, which is also called *the primary probability*. Hence $P(\overline{A}) = 1 - p$

The probability that A occurs exactly "j" times out of n, and that A does not occur in the remaining "$n - j$" attempts, may be calculated in the same manner as we did in the more specific example of throwing a dice 12 times as: $p^j\,(1 - p)^{n-j}$.

To find the probability that A occurs exactly j times, independently of where in the sequence it happens is therefore $p^j\,(1-p)^{n-j}$ times the number of different ways we can select a *j-subset* from an *n-set.* The answer is as we have used it many times already:

$$\binom{n}{j} = C(n,q) = \frac{n!}{(n-q)!\,q!}$$

Thus we obtain the general expression for the *binomial distribution*

$$P(X = j) = \binom{n}{j} p^j (1-p)^{n-j} \quad ; j = 0, 1, 2, ..., n$$

4. The binomial distribution and Pascal's triangle

Pascal's triangle is named after Blaise Pascal, a French mathematician and physicist from the 17th century. He is often ascribed as the founder of probability theory, as well as constructor of the first mechanical calculator.
(Unfortunately he gave up his brilliant scientific career at an early age, and dedicated the rest of his life to the study of theology).

Pascal's triangle is in principle just an exposure of the binomial coefficients in a triangular scheme.

$$
\begin{array}{lc}
n=0 & \binom{0}{0} \\
n=1 & \binom{1}{0}\ \binom{1}{1} \\
n=2 & \binom{2}{0}\ \binom{2}{1}\ \binom{2}{2} \\
n=3 & \binom{3}{0}\ \binom{3}{1}\ \binom{3}{2}\ \binom{3}{3} \\
n=4 & \binom{4}{0}\ \binom{4}{1}\ \binom{4}{2}\ \binom{4}{3}\ \binom{4}{4} \\
\ldots\ldots & \\
n-1 & \binom{n-1}{0}\cdots\cdots\binom{n-1}{q-1}\ \binom{n-1}{q}\cdots\cdots\binom{n-1}{n-1} \\
n & \binom{n}{0}\cdots\cdots\binom{n}{q-1}\ \binom{n}{q}\ \binom{n}{q+1}\cdots\cdots\binom{n}{n}
\end{array}
\qquad
\begin{array}{c}
1 \\
1\quad 1 \\
1\quad 2\quad 1 \\
1\quad 3\quad 3\quad 1 \\
1\quad 4\quad 6\quad 4\quad 1
\end{array}
$$

In the scheme above to the left, we have established the binomial coefficients in symbolic form $\binom{n}{q} = C(n,q)$ for $n = 0 \ldots 4$. Then we have added two rows, corresponding to $n - 1$, and n.
To the right we have evaluated the binomial coefficients for $n = 1 \ldots 4$
You should notice that along the sides of the triangle we have ones only.
This because $\binom{n}{0} = 1$, and $\binom{n}{n} = 1$, since there is only one way to select zero elements from a *n-set.*
You don't have to watch Pascal's triangle for a very long time to discover a very simple system, when creating a new row from the one above, since each number is the sum of the two neighbour numbers (the numbers to the right and to the left) in the row above. This should indicate that:

$$\binom{n}{q} = \binom{n-1}{q-1} + \binom{n-1}{q} \qquad \text{or} \qquad C(n,q) = C(n-1,q-1) + C(n-1,q)$$

The formula can be proved algebraically, when we use the formula for $C(n,q)$.

$$\binom{n-1}{q-1} + \binom{n-1}{q} = \frac{(n-1)!}{(q-1)!(n-q)!} + \frac{(n-1)!}{q!(n-q-1)!} = \frac{(n-1)!}{(q-1)!(n-q-1)!}\left(\frac{1}{(n-q)} + \frac{1}{q}\right)$$

$$= \frac{(n-1)!}{(q-1)!(n-q-1)!}\left(\frac{n}{(n-q)q}\right) = \frac{n!}{q!(n-q)!} = \binom{n}{q}$$

But it may also be proved using logic:
Let us assume that we have a set with n elements: $U = \{a_1, a_2, a_3, \ldots, a_n\}$. We shall then divide the *q-subsets* into two groups:

1. The *q-subsets* that a_1 belongs to. There are then $n - 1$ elements left, and q -1 to be selected which can be don in $C(n - 1, q - 1)$ different ways.
2. The subsets, that a_1 does not belong to. Then we must select q elements from n -1 elements. This may be done in $C(n - 1, q)$ different ways. Together they include all combinations, so:

$C(n, q) = C(n - 1, q - 1) + C(n - 1, q)$ or $\binom{n}{q} = \binom{n-1}{q-1} + \binom{n-1}{q}$ As asserted.

Next we shall look at a so called binomial: $(a + b)^n$, where a and b are numbers, and n is an integer. Most students from public school, know how to write it down for $n = 2$: $(a + b)^2 = a^2 + b^2 + 2ab$. But what about n =4 or n =6? Before analyzing, we present the result below from n = 0..5.

$$
\begin{aligned}
(a + b)^0 &= 1\\
(a + b)^1 &= 1\cdot a + 1\cdot b\\
(a + b)^2 &= 1\cdot a^2 + 2ab + 1\cdot b^2\\
(a + b)^3 &= 1\cdot a^3 + 3a^2b + 3ab^2 + 1\cdot b^3\\
(a + b)^4 &= 1\cdot a^4 + 4a^3b + 6a^2b^2 + 4ab^3 + 1\cdot b^4\\
(a + b)^5 &= 1\cdot a^5 + 5a^4b + 10a^3b^2 + 10a^2b^3 + 5ab^4 + 1\cdot b^5
\end{aligned}
$$

What you discover is that the coefficients (arranged in this manner) when evaluating the binomial are equal to the numbers in Pascal's triangle.
What might come as a surprise is actually rather obvious. For example if we look at $(a + b)^5$.

$$(a + b)^5 = (a + b)\,(a + b)\,(a + b)\,(a + b)\,(a + b)$$

Each term in the evaluation of the products comes from taking (b) from j parenthesis and (a) from $5 - j$ parenthesis and multiply them to get $b^j a^{5-j}$. The number of different ways, one may select j parenthesis out of 5 is of course $C(5,j) = \binom{5}{j}$. The same applies for all terms, which explains that the coefficients of the different terms are the binomial coefficients.

$$(a+b)^5 = \binom{5}{0}a^5b^0 + \binom{5}{1}a^4b^1 + \binom{5}{2}a^3b^2 + \binom{5}{3}a^2b^3 + \binom{5}{4}a^1b^4 + \binom{5}{5}a^0b^5$$

The formula is trivially generalized to evaluating $(a + b)^n$, but it is then more suitable written by the help of a summation sign.

$$(a+b)^n = \sum_{j=0}^{n} \binom{n}{j} a^{n-j} b^j$$

This formula is called the *binomial formula*.

And again we stress that the coefficients to the terms $a^j\, b^{n-j}$, when evaluating $(a + b)^n$ are equal to the nth row in Pascal's triangle.

We only need (formally) to establish that the binomial distribution is a probability distribution, that is, the sum of the probabilities $P(X=j),\ j = 0 \ldots n$ are equal to 1.
This, however, is straightforward, once we have established the binomial distribution.

Suppose namely that we have two numbers s and t for which $s + t = 1$, and therefore $t = 1 - s$.
It then follows that $(s + t)^n = 1$. We then apply the binomial formula to evaluate $(s + t)^n$.

$$1 = 1^n = (s+t)^n = \binom{n}{0} s^n t^0 + \binom{n}{1} s^{n-1} t^1 + \ldots + \binom{n}{j} s^{n-j} t^j + \ldots + \binom{n}{n} s^0 t^n$$

If we in this formula put $t = p$ and $s = 1 - p$ we find again the binomial distribution.

$$1 = \binom{n}{0} p^0 (1-p)^n + \binom{n}{1} p^1 (1-p)^{n-1} + \ldots + \binom{n}{j} p^j (1-p)^{n-j} + \ldots + \binom{n}{n} p^n (1-p)^0$$

From which it appears that the sum of the binomial probabilities is equal to 1

In a manner of speaking it is often referred to as a ”success”, when the primary event occurs (e.g. the dice shows 6 eyes), and a “failure” if it does not occur. (You should not associate any semantics to these expressions, since the primary event might be the probability that you get killed in a traffic accident, or the probability of getting divorced).

The probability $P(X = j)$ then means the probability of getting j “successes” in n attempts, and there exist an commonly used abbreviation $b(j;\, n, p)$ or just b_j , when n and p are understood.

Often we are interested in the number j giving the largest probability, that is, for which j, $b_j = b(j;\, n, p)$ is largest. To do so, we look at the ratio b_j / b_{j-1}.

$$\frac{b_j}{b_{j-1}} = \frac{\binom{n}{j} p^j (1-p)^{n-j}}{\binom{n}{j-1} p^{j-1} (1-p)^{n-j+1}} = \frac{p(n-j+1)}{(1-p)j}$$

The last expression comes from writing out the binomial coefficients and reducing.
First we write the condition that the sequence b_0 , b_1 , b_2 , ... , b_n is increasing.

$$b_j > b_{j-1} \iff \frac{b_j}{b_{j-1}} > 1 \iff \frac{p(n-j+1)}{(1-p)j} > 1 \iff$$
$$p(n-j+1) > (1-p)j \iff j < np + p$$

From which we conclude that if: $j \le np$ then $b_j > b_{j-1}$ meaning that the sequence is increasing.
In a quite similar way, if we reverse the inequality sign, we find that the sequence is decreasing if and only if : $j > np+p$, from which we conclude that:

$j \ge np+1 \;\Rightarrow\; b_j < b_{j-1}$, or $j \ge np$ then $b_{j+1} < b_j$

Implying that the sequence is decreasing.

So the sequence $b_0, b_1, b_2, \ldots, b_n$. is increasing for $j \le np$ and decreasing for $j \ge np$.

- The largest probability must therefore be at $j = np$ if np is an integer, otherwise the largest value for the probability is found among the two neighbouring integers to np.

5. Stick diagrams for the binomial distribution

To get a overview of distribution of the binomial distribution, it is often illustrated by a so called "stick diagram". Below the distribution is depicted for $n = 12$ and $p = 1/3$. Notice that the largest probability is for $j = 3$, where $np = 3$, as we derived above.

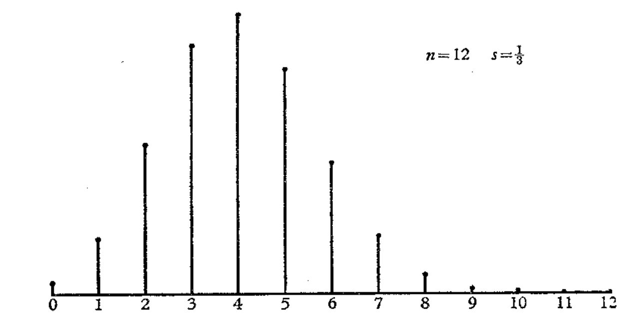


Fig. 14

44

6. Cumulated probabilities

It may – even after the appearance of mathematical calculators - be rather circumstantial to calculate binomial probabilities for high values of n.
Even if one could do numerical calculation with a pocket calculator since 1970, it was not until the mid 90'ties with the appearance of mathematical pocket calculators that one could directly evaluate formulas from combinatorics and probability. Before that one was forced to use entries in tables and often do manual interpolation.
Tables are still used, but they have always, (when it concerns probabilities) been designed to have *cumulated* probabilities, that is, sums of probabilities, from the low end or the high end.

As an example, we shall again consider 12 throws with a dice where the primary event is that the dice shows 6 eyes. The probability that we have success only two times, we have up till now written $P(X=2)$, but similarly, we may write e.g. $P(X \le 4)$ for the probability that we get at most 4 eyes.

$P(X \le 4)$ is an example of a cumulated probability, having the precise meaning:

$$P(X \le 4) = P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3) + P(X = 4)$$

In general for arbitrary n and p, we have:

$$P(X \le q) = \sum_{j=0}^{q} P(X = j) \quad \text{and} \quad P(X \ge q) = \sum_{j=q}^{n} P(X = j)$$

As mentioned before, at the time, where we used tables in high school, the tables only had cumulated probabilities.

Eksemples

1. (This is an outdated example, from the period, where we used tables). We want to find the probability of getting 6 eyes on a dice two times, throwing a dice 12 times. In the table we find for $n = 12$ and $p = 1/6$. $P(X \leq 2) = 0.6774$ and $P(X \leq 1) = 0.3813$.

From which we find: $P(X = 2) = P(X \leq 2) - P(X \leq 1) = 0.2961$.

2. Same experiment, but now we wish to find the probability that we get at least two six in twelve attempts

$P(\text{At least two six}) = P(X \geq 2) = 1 - P(X \leq 1) = 1 - 0.3813 = 0.6187$

What is the most probable outcome of the experiment described above? Since $np = 2$, it is not surprising that it is 2

3. A family is planning to have 5 children. Let the primary event (success) be that they have a girl with probability ½.
A: What is the probability that they have at least 2 girls, and
B: what is the most likely outcome.
We find from a table $(X \leq 2) = 1/2$.
$np = 5/2$, so we look at: $P(X=2) = P(X \leq 2) - P(X \leq 1) = 0.5000 – 0.1875 = 0.3025$, and $P(X=3) = P(X \leq 3) - P(X \leq 2) = 0.8125 – 0.5000 = 0.3125$. So the most likely outcome is (marginal) 3 girls.

Exercises

1. A dice is rolled 40 times. Find the probability that:
   a) That the outcome will be 6 eyes exactly 6 times.
   b) That the outcomes of 6 eyes will be less than 4 times.
   c) That the outcomes of 6 eyes will be between 4 and 7 times.

2. What has the greatest probability when rolling dices. To get 6 eyes once in 6 throws, or to get exactly 6 eyes twice in 12 throws with two dices?

3. According to statistics 73.3% of 43 year will be alive at the age of 68. (In Denmark you pass the 9-12 graduation exams at the age of 18). 20 persons are gathered at their 25 year student anniversary.
   a) What is the probability that the same group will present at their 50 year student anniversary?
   b) What is the most probable number of persons at their 50 year student anniversary?
   c) What is the probability that exactly this number of persons will be present.

4. During an epidemic of a certain cattle disease, one counted on that the chance of infection is 25%. 10 animals were vaccinated with a new developed vaccine, and the outcome was that only one of these animals got infected. Is this outcome significant that the vaccine is effective, if you claim that among the non vaccinated animals should be less than 5% probability of such or more favourable result.

7. Samples with replacement

Having a large sample of objects, which has to be examined for errors or for making a quality control it is seldom possible, (or just too time consuming), to examine each object. The sample should therefore be representative for the whole set of objects. Probability deals with the question whether the sample is representative for the set as a whole, within some limits.
For example on can decide that within 95% it is representative, meaning that the sample should be within the interval, where 95% of the samples are statistically.

This version of taking a sample is based on the binomial distribution. We have (usually a large) set of elements, from which we select a sample (a minor subset) to be examined.

The notion "with replacement" actually imply, that you replace the element taken and examined, making it possible to examine the same element twice, but in this connection the sets to examine are assumed to be so large, that it makes no difference whether you replace it or not, if the sample is much smaller than the set itself.
Especially in the context of opinion pools, it is common to use the notion of "a population" instead of "set", and this notion has been broaden to include other sets as well.

Example

1. Let us assume that a person buys a large number of seed. The dealer insures that at least 80% of the seed will sprouts, when treated properly. The person sows 50 seeds and 34 sprouts (68%). Is that statistically acceptable, if the dealer's assertion is correct?
   The number of seeds that sprouts are binomially distributed with n=50 and P=0.80. We wish to find the probability $P(X \leq 34)$ that at least 34 of the seeds sprouts: If you consult a table of binomial probabilities or use a mathematical calculator we find $P(X \leq 34) = 0.0308$. In which case we should normally reject the hypothesis that at least 80% of the seed sprouts, since the are less than 5% probability of this outcome, if the fluctuations are only statistical.

8. Hypothesis testing

The example above is a characteristic example of hypothesis testing by selecting a sample.
When you do the test, you should first decide on a *level of significance*. This could for example be 95%, meaning that if the probability of the result of the sample is less than 5% you will reject the hypothesis (the assertion). With this level of significance, you would have rejected the assertion concerning the seed. However with a level of significance of 99% it would have been accepted.

Levels of significance are usually chosen as 90%, 95% or 99%. There are however drawbacks in choosing a high or a low levels of significance.
With a low level of significance, there is a risk of rejecting a hypothesis for statistical reasons, even if the hypothesis is actually correct, and with a higher level of significance, you are at risk to accept a hypothesis, even if it is wrong.
The question which level of significance (depending of n and p) one should choose, to make the largest probability of reliance of the sample test, is certainly not an easy question to answer.

We will elaborate a little on the example above with the sprouting of the seeds. First we shall find the smallest number of plants which sprouts, so that we can accept the hypothesis: At least 80% sprout, at a significance level 95%. So we must determine the smallest number q, such that $P(X \leq \text{q}) > 0.05$. We have already seen, that $P(X \leq 34) = 0.0308$, and from a table or otherwise we find $P(X \leq 35) = 0.0607$. So we should have accepted the hypothesis (at a significance level 95%), if 35 of the 80 seeds had sprouted.
If we have a level of significance of 90% or 99% the number of seeds that would reject the hypothesis be: 35 and 32, since: $P(X \leq 35) = 0.0607$ and $P(X \leq 36) = 0.1106$. $P(X \leq 32) = 0.0063$ and $P(X \leq 33) = 0.0144$

2. A political party is claimed to have a support of 20% of the voters. To examine this, 40 voters are asked if they would vote for the party. If the hypothesis is correct, one would expect that 8 confirmed. But the result is that only 5 answer yes. On these grounds, is it possible to reject the assertion at a level of significance of 90%?
   We have $n = 40$ and and $p = 0.20$, and we must find the largest j for which $P(X \leq j) \leq 0.10$. In the table or otherwise we find: $P(X \leq 4) = 0.0759$ and $P(X \leq 5) = 0.1613$. Since the last probability is greater than 0.1000, we may not reject the hypothesis on the ground of the sample examined, since to reject it, the support (statistically) should have been less than 4.

9. Samples without replacement. The hyper geometric distribution

Samples, without replacement is most often illustrated by taking at random balls of various colours from a jar.
But it can actually be applied each time you take a subset of a set having elements of two, three or more different kinds, and you wish to calculate the probability of a certain composition of kinds in the sample. In the section of combinatorics we calculated the probability of the distribution of gender in a committee of 4, when selected from a class of 24. To convert the number of possibilities to probabilities, you need only to divide by the total number of possibilities, in this case $C(24,4)$.

For a specific situation the distribution of probabilities of a specific event, e.g. the number of boys in a committee of 4, the number of red balls in a sample of 6, is called the *hyper geometric distribution*. It has widespread appliance in all part of human life and the society.

Example

1. In a jar there are 8 red and 12 blue balls. We select a sample of 5, and we want to find the probability distribution for the number of red balls. That is, the probabilities $P(X=j)$, where j = 0, 1 ,2 ,3 ,4 , 5. There are C(20,5) possibilities to pick the sample. We shall as an example calculate the $P(X=3)$. The 3 red balls can be selected in C(8,3) different ways, and the blue balls can be selected in C(12,2) different ways. From this we find.

$$P(3) = \frac{C(8,3)C(12,2)}{C(20,5)} = \frac{56 \cdot 66}{15504} = 0{,}2384$$

If you have a set of n elements, from which q of them is of a certain kind (say red) different form the $n - q$ others, and if we pick a sample of r elements, then if $r < q$, we may the write the probabilities to get $j = 0, 1, 2, \ldots r$ "red" elements in the sample.
In the context of the hyper geometric distribution, these probabilities are usually written using with the binomial coefficients $\binom{n}{q} = C(n,q)$.

Below is shown the general expression for the hyper geometric distribution.

$$P(j) = \frac{\binom{q}{j}\binom{n-q}{r-j}}{\binom{n}{r}} \quad j = 0,1,2\ldots r$$

4. The Poisson distribution

The Poisson distribution is traditionally introduced as the probability distribution, representing the number of calls per unit of time to a manually served telephone central (more than 60 years ago). But it also applies to the number of nuclei decaying in a certain time from a radioactive specimen.

We shall assume that the calls to a central comes at random, and that the probability of a call is λ per unit of time. The probability of a call in the small time Δt is then $\lambda\Delta t$ and therefore the probability of no calls in Δt is $(1 - \lambda\Delta t)$.
We will then establish a recursion relation for the probability $P_n(t)$ that there are n calls to the central during the time t. It goes as follows:

The probability $P_{n+1}(t+\Delta t)$ that the central receives n +1 calls in the time $t + \Delta t$,is equal to the probability $P_{n+1}(t)(1- \lambda\Delta t)$ that it receives n+1 calls in the time t and 0 calls in the time Δt, plus the probability $P_n(t)\ \lambda\Delta t$ that the central receives n calls in time t and 1 call in time Δt, plus the probability $P_{n-1}(t)(\lambda\Delta t)^2$ that the central receives n -1 calls in time t and 2 calls in time Δt plus….plus.
We have multiplied the probabilities, since they are assumed to be independent.

We shall omit the last and the following terms, since we shall eventually divide by Δt and take the limit $\Delta t \rightarrow 0$, so all terms with Δt having a higher power than 1, will go to zero.
We may the establish our recursion equation.

$$P_{n+1}(t+\Delta t) = P_{n+1}(t)(1-\lambda\Delta t) + P_n(t)\lambda\Delta t \quad \Leftrightarrow$$
$$\frac{P_{n+1}(t+\Delta t) - P_{n+1}(t)}{\Delta t} = -\lambda P_{n+1}(t) + \lambda P_n(t)$$

And taking the limit for $\Delta t \rightarrow 0$

$$\frac{dP_{n+1}(t)}{dt} = -\lambda P_{n+1}(t) + \lambda P_n(t)$$

The equation is valid for n >0, but if we formally put $P_{-1}(t) = 0$ (since any other choice would be meaningless) we are left with a simple differential equation for $P_0(t)$.

$$\frac{dP_0(t)}{dt} = -\lambda P_0(t)$$

This equation may also be established directly, since the probability $P_0(t+\Delta t)$ that there are no calls in the time $t + \Delta t$ must be equal to the probability $P_0(t)(1- \lambda\Delta t)$ that are no calls in the time t times the probability of no calls in Δt , which is $(1- \lambda\Delta t)$. This leads to the equation

$$P_0(t+\Delta t) = P_0(t)(1-\lambda\Delta t) \quad \Leftrightarrow \quad \frac{P_0(t+\Delta t) - P_0(t)}{\Delta t} = -\lambda P_0(t) \quad \Rightarrow \quad \frac{dP_0(t)}{dt} = -\lambda P_0(t)$$

The last equation has the well known solution:

$$P_0(t) = e^{-\lambda t}$$

Where we have used that $P_0(0) = 1$. If we insert the expression for P_0 in the recursion equation we get:

$$\frac{dP_1(t)}{dt} = -\lambda P_1(t) + \lambda e^{-\lambda t} \quad \Leftrightarrow \quad \frac{d}{dt}(P_1(t)e^{\lambda t}) = \lambda \quad \Leftrightarrow$$
$$P_1(t) = \frac{(\lambda t)}{1!}e^{-\lambda t}$$

We could go on and solve for $P_2(t)$, and so on, until we have discovered a systematic development. This we shall not do however, but satisfy ourselves by showing that the general solution is:

$$P_n(t) = \frac{(\lambda t)^n}{n!}\, e^{-\lambda t}$$

We then write the recursion equation substituting *n*+1 with *n*.

$$\frac{dP_n(t)}{dt} = -\lambda P_n(t) + \lambda P_{n-1}(t)$$

$$\frac{d}{dt}\left(\frac{(\lambda t)^n}{n!}\, e^{-\lambda t}\right) = \frac{n\lambda(\lambda t)^{n-1}}{n!}\, e^{-\lambda t} - \lambda \frac{(\lambda t)^n}{n!}\, e^{-\lambda t} = \frac{\lambda(\lambda t)^{n-1}}{(n-1)!}\, e^{-\lambda t} - \lambda \frac{(\lambda t)^n}{n!}\, e^{-\lambda t} = -\lambda P_n(t) + \lambda P_{n-1}(t)$$

Which is seen to agree with the recursion equation above.

For the decay of nuclei, we have exactly the same equations. Assuming that the probability that a nucleus decays in The time dt is λdt, and let $P(t)$ be the probability that a nucleus that has not decayed at $t = 0$, has survived at time t, This leads to the equation: $P(t+dt) = P(t)(1 - \lambda dt)$, which is easily transformed into: $dP/dt = -\lambda\, P(t)$.
The same equation as above.