Why a new method?

In school math, 'inference from sample to population' means using the sample to estimate the population proportion. For example, if we ask some randomly selected people whether they like chocolate, we would like to use the relative frequency of chocolate-liking people in the sample to estimate the proportion of people in the population who like chocolate. We will refer to this proportion in the population as the population proportion in the following. It is identical to the success probability or the probability of the event "A randomly selected person likes chocolate".
In technical terms, this is briefly referred to as the estimation of the parameter ***p***.

For this purpose, two methods are shown in the school:

1) The point estimation: the population proportion is set to the same value as the relative frequency.

2) Interval estimation with confidence intervals.

The problem

Method 1 gives us an optimal estimation in the following meaning: If a sample is given, we may have drawn it from different populations. We now identify that population which conveys the highest probability to get the present sample.
While this method is good and correct, it is hard for middle school students to understand *why* we obtain an optimal point estimator with this. Mistakenly, most students assume there is some type of natural law that forces the relative frequency of the sample to be similar to the population proportion.

Method 2 has the following problems:

- The method contradicts human intuition. In everyday life, we humans make inferences from the sample to the population by assigning probabilities to different possible populations. Thus, it might be desirable to have a method to infer mathematically correctly from the sample to the population in this way.

- To help students understand why statistics works, we must address the following question: When we conduct a survey, we only interview a small part of the population. But what do we know about the people we didn't ask? With the method of confidence intervals, the answer is possible but relatively difficult to understand.

- It supports the erroneous view that the confidence level indicates the probability with which the actual population proportion is located in the confidence interval.
As a result, there is often a misinterpretation of the p-value in hypothesis testing, which is mistaken for the probability that the hypothesis is true.
The American Statistical Association has apparently also recognized this problem so that in 2016 it felt compelled to publish a statement clarifying the existing misunderstandings.

(American Statistical Association Releases Statement on Statistical Significance and p-Values: http://amstat.tandfonline.com/doi/abs/10.1080/00031305.2016.1154108#.Vt2XIOaE2MN)*

---

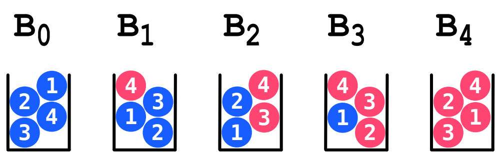
*See also: "A confidence interval is not a probability, and therefore it is not technically correct to say the probability is 95% that a given 95% confidence interval will contain the true value of the parameter being estimated."(https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2947664/)
and
"A 95% confidence level does not mean that for a given realized interval there is a 95% probability that the population parameter lies within the interval (i.e., a 95% probability that the interval covers the population parameter)."
(https://en.wikipedia.org/wiki/Confidence_interval#Common_misunderstandings)

The solution

There is a possibility to get rid of all these problems at once: With a method that allows us to infer directly from the sample to the population. This method is called directly inferring statistics. We can start teaching this method from day one in stochastics class - all we need is to count blue and red balls in boxes. With little effort, we will then expand this method into an easy-to-understand and very effective way to do statistics.

The beginning - boxes and balls

Let's start with 5 boxes, each containing 4 balls that are either blue or red. Let all possible proportions of red balls be represented. (We could also focus on the proportions of blue balls, but out of pure arbitrariness we chose the red balls).



Suppose we randomly draw with replacement 3 balls from one of the boxes without knowing from which one. We can then make a guess as to which box the sample was drawn from. If, for example, one blue and 2 red balls have been drawn, we could say: "It has probably been drawn from B_3. It could also have been drawn from B_1, but that would be improbable." If the sample consisted of 3 blue balls, we could think: "B_0 is most likely and B_3 is least likely."
We humans think like this. *We assign probabilities to populations*.

A brief comparison of probability theory and statistics will show how this works mathematically.

Mathematical note

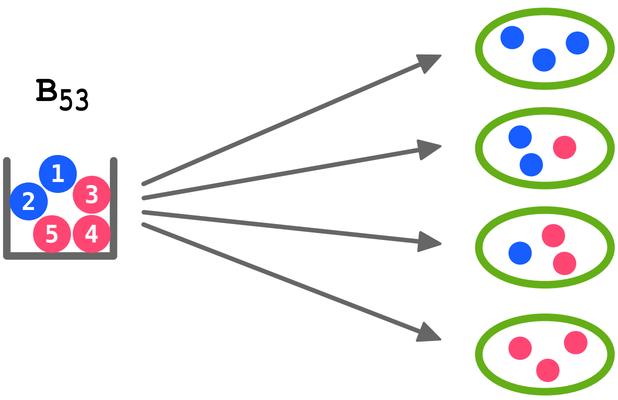
The method is based on the concept of conditional probability. In the above example, we can ask what the probability is that a specific sample is drawn from a specific box.
For understanding conditional probability may be difficult for students, this method simplifies the process by honing in on the number of elements in the population, making it just as accessible as the Laplace probability.
Fortunately, with the help of combinatorics and integrals, we can apply this concept to populations of any size, while the underlying idea remains the same: the probability of a population is determined by dividing the number of certain samples by the number of all samples.

By the way: How challenging it is to understand conditional probabilities can be seen from the fact that many people believe that the following problem is unsolvable: There are two black and two white balls in a box. The balls are randomly drawn twice without replacement. What is the probability that a black ball is drawn the first time under the condition that a white ball is drawn the second time?

Probability, Relative Frequency, Population Proportion

In probability, we have a population (in this case with a determined **population proportion** of red balls) and we assign **probabilities** to the possible samples (here with their different **relative frequencies** of red balls).

Shown is the threefold random drawing of ordered samples with replacement from population B_{53}. The samples with equal relative frequencies form the events that we assign probabilites to.



$\mathcal{P}\big((3,0)\big) = 0.064$

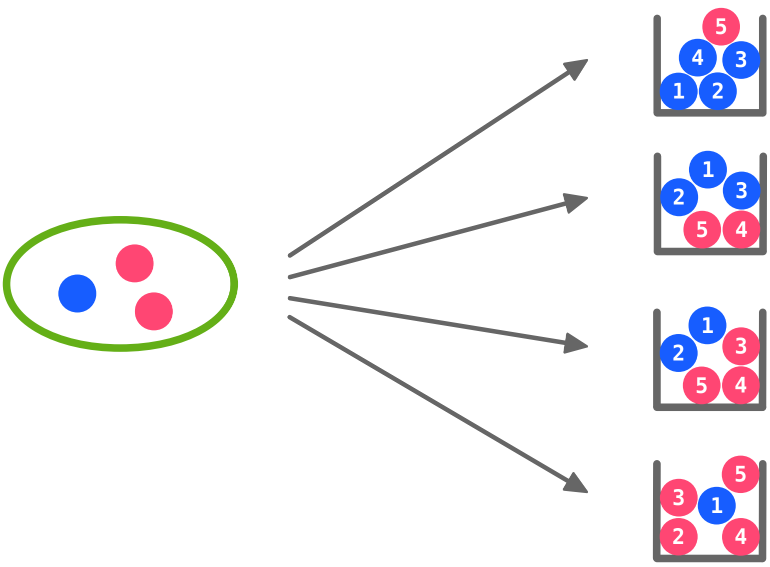
$\mathcal{P}\big((3,1)\big) = 0.288$

$\mathcal{P}\big((3,2)\big) = 0.432$

$\mathcal{P}\big((3,3)\big) = 0.216$

$\mathcal{P}\big((3,1)\big)$ denotes the probability that the relative frequency of red balls in a sample of size 3 is equal to $\frac{1}{3}$.

However, we can also assign **probabilities** (based on a sample with a certain **relative frequency** of red balls) to the populations (with their different **population proportions** of red balls) from which the sample can be drawn.



$\mathcal{P}\big((5,1)\big) = 0.08$

$\mathcal{P}\big((5,2)\big) = 0.24$

$\mathcal{P}\big((5,3)\big) = 0.36$

$\mathcal{P}\big((5,4)\big) = 0.32$

$\mathcal{P}\big((5,2)\big)$ denotes the probability that in a population of size 5, the population proportion of red balls is equal to $\frac{2}{5}$.

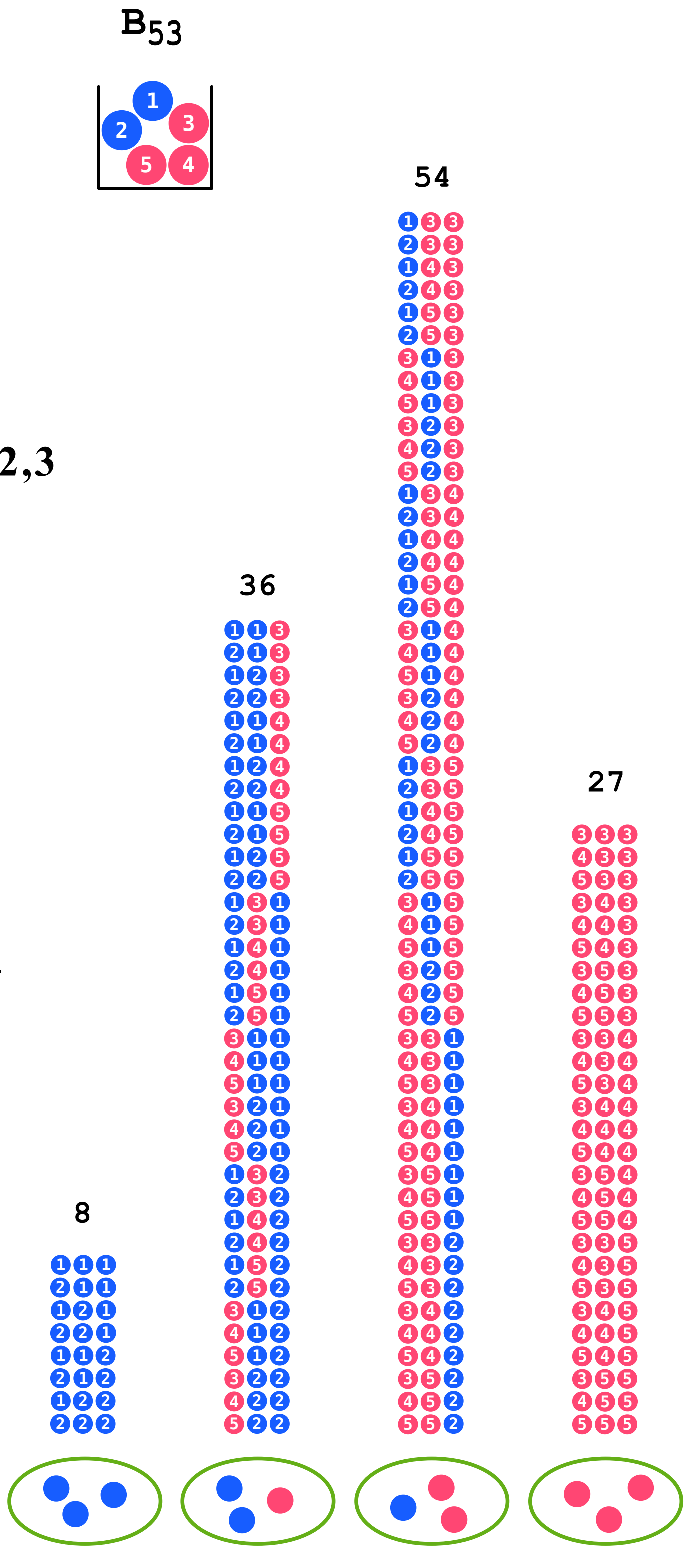
Probability - Counting Balls

In probability, we have a given population and calculate the probability of samples with a certain relative frequency by dividing the number of samples with this relative frequency by the number of all possible samples from this population.

Samples of size n with k red balls we further call (n, k)-samples.

Random experiment: Threefold drawing of an ordered sample with replacement

The red balls are counted.



number of balls in the sample: $n = 3$

numbers of red balls in the possible samples: $k = 0, 1, 2, 3$

number of balls in the population: $g = 5$

number of red balls in the population: $r = 3$

probability to draw a $(3,2)$-sample

$$\mathcal{P}\left((3,2)\right) = \frac{54}{8 + 36 + 54 + 27} = \frac{54}{125} = 0.432$$

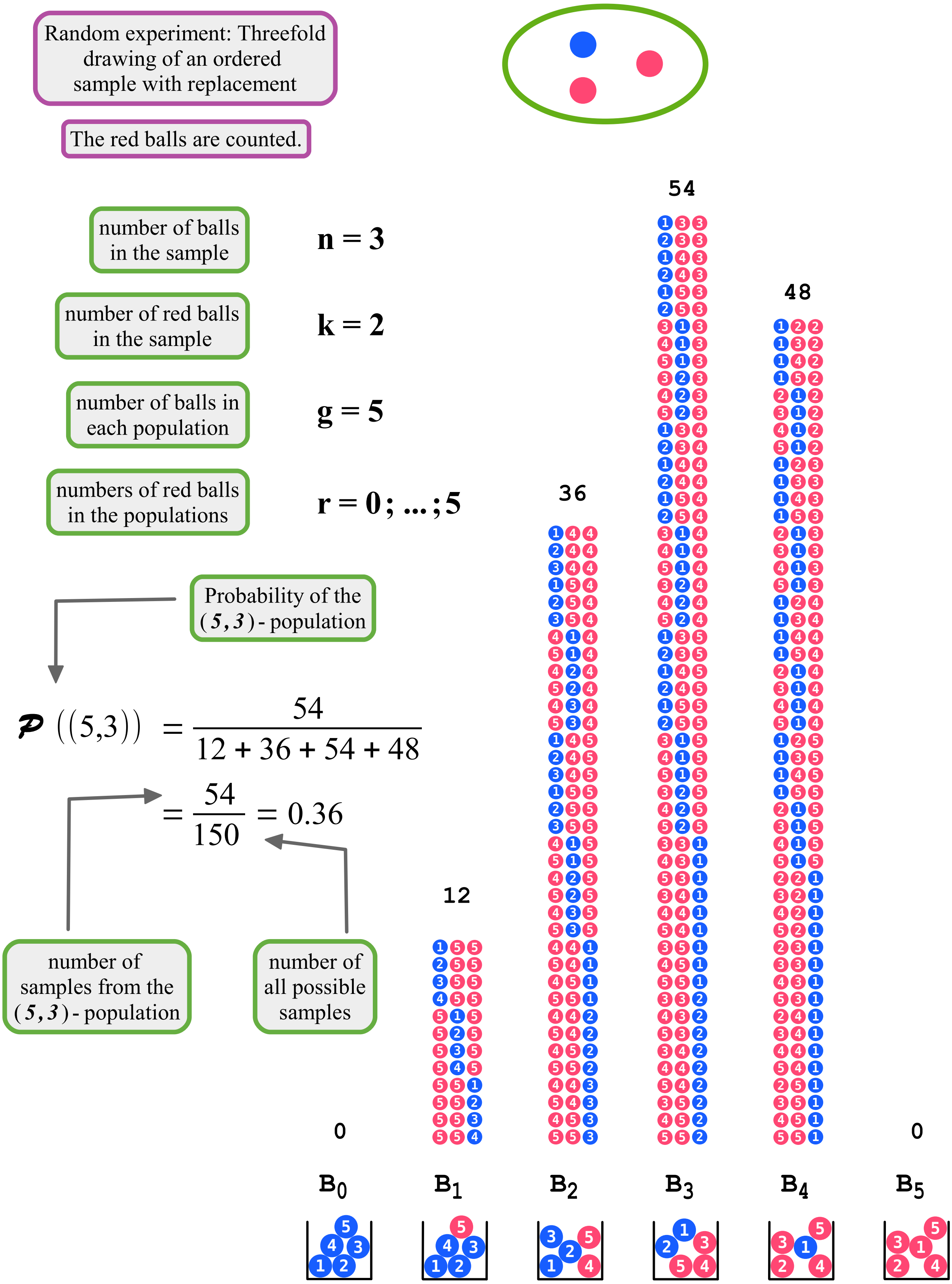
number of all $(3,2)$-samples

number of all samples of size 3

Directly Inferring Statistics - Counting Balls

Given a sample with a certain **relative frequency**, we calculate the probability of a population by dividing the number of samples from that population by the number of all samples from all populations.

Populations with g elements and r red balls we further call (g, r) - populations.



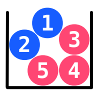
Probability - with combinatorics

If we have combinatorics at our disposal for probability, we can deal with larger populations and larger samples. However, the calculations remain the same in principle. Again, given a population, we will calculate the probability for the samples with a certain relative frequency by dividing the number of samples with this relative frequency by the number of all possible samples.

We now asume a population with $\boldsymbol{g}$ balls, of which $\boldsymbol{r}$ are red, and look for the probability of a sample of size $\boldsymbol{n}$ with $\boldsymbol{k}$ red balls. In other words, we are looking for the probability with which we can draw an $(\boldsymbol{n}, \boldsymbol{k})$-sample from a given $(\boldsymbol{g}, \boldsymbol{r})$-population.

Let's take a look at a specific example: We have a population with $\boldsymbol{g} = 5$ balls, of which $\boldsymbol{r} = 3$ are red. We want to draw a sample of size $\boldsymbol{n} = 20$ and ask for the probability that $\boldsymbol{k} = 11$ balls in the sample are red. In doing so, we draw ordered samples with replacement.

$\mathbf{B}_{53}$



n = 20 **g = 5**

k = 11 **r = 3**

We first calculate the number of all possible samples of size 20 with 11 red balls.

$$\mathcal{S}((20,\, 11)) = \binom{20}{11} \cdot 3^{11} \cdot (5-3)^{20-11}$$

Then we calculate the number of all possible samples of size 20.

$$\sum_{i=0}^{20} \binom{20}{i} \cdot 3^{i} \cdot (5-3)^{20-i}$$

The quotient of both numbers is the probability of drawing a sample of size $\boldsymbol{n} = 20$ with $\boldsymbol{k} = 11$ red balls from a given population with $\boldsymbol{g} = 5$ balls, of which $\boldsymbol{r} = 3$ are red.

$$\mathcal{P}((20,\, 11)) = \frac{\binom{20}{11} \cdot 3^{11} \cdot (5-3)^{20-11}}{\sum_{i=0}^{20} \binom{20}{i} \cdot 3^{i} \cdot (5-3)^{20-i}}$$

If we also take into account that

$$\sum_{i=0}^{20} \binom{20}{i} \cdot 3^{i} \cdot (5-3)^{20-i} = 5^{20}$$

we get the probability

$$\mathcal{P}((20,\, 11)) = \frac{\binom{20}{11} \cdot 3^{11} \cdot (5-3)^{20-11}}{5^{20}}$$

Hence:

$$\mathcal{P}((20,\, 11)) = \frac{15\,233\,848\,381\,440}{5^{20}}$$

$$\mathcal{P}((20,\, 11)) \approx 0{,}1597$$

Probability - with combinatorics

In general, wie have the number of $(\boldsymbol{n},\boldsymbol{k})$-samples from a $(\boldsymbol{g},\boldsymbol{r})$-population

$$\mathcal{S}((n,k)) = \binom{n}{k} \cdot r^k \cdot (g-r)^{n-k}$$

and the number of all samples of size $\boldsymbol{n}$ from a $(\boldsymbol{g},\boldsymbol{r})$-population,

$$\sum_{i=0}^{n} \binom{n}{i} \cdot r^i \cdot (g-r)^{n-i}$$

from which we can calculate the probability of drawing an $(\boldsymbol{n},\boldsymbol{k})$-sample from a given $(\boldsymbol{g},\boldsymbol{r})$-population.

$$\mathcal{P}((n,k)) = \frac{\binom{n}{k} \cdot r^k \cdot (g-r)^{n-k}}{\sum_{i=0}^{n} \binom{n}{i} \cdot r^i \cdot (g-r)^{n-i}}$$

If we also consider this equation

$$\sum_{i=0}^{n} \binom{n}{i} \cdot r^i \cdot (g-r)^{n-i} = g^n$$

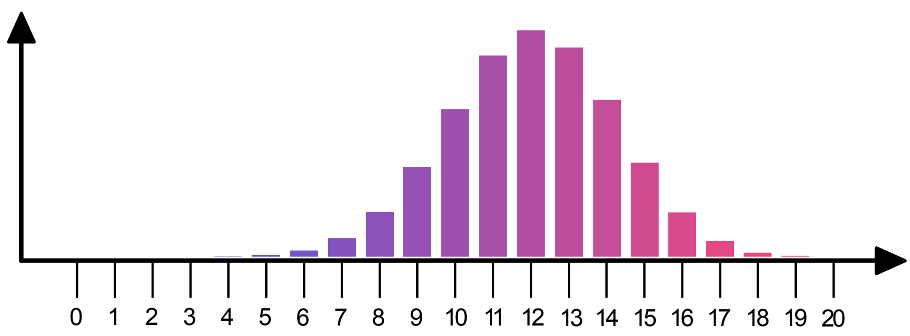
we can write

$$\mathcal{P}((n,k)) = \frac{\binom{n}{k} \cdot r^k \cdot (g-r)^{n-k}}{g^n}$$

After a brief reshaping, this calculation looks like the binomial distribution we are used to.

$$\mathcal{P}((n,k)) = \binom{n}{k} \cdot \left(\frac{r}{g}\right)^k \cdot \left(1-\frac{r}{g}\right)^{n-k}$$

The probabilities of the possible samples for $\boldsymbol{g} = 5$, $\boldsymbol{r} = 3$ and $\boldsymbol{n} = 20$ are shown below as a bar chart.



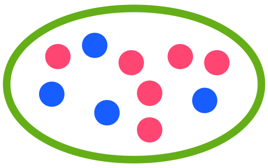
k=0	1 048 576	k=7	1 388 840 878 080	k=14	11 864 824 220 160
k=1	31 457 280	k=8	3 385 299 640 320	k=15	7 118 894 532 096
k=2	448 266 240	k=9	6 770 599 280 640	k=16	3 336 981 811 920
k=3	4 034 396 160	k=10	11 171 488 813 056	k=17	1 177 758 286 560
k=4	25 719 275 520	k=11	15 233 848 381 440	k=18	294 439 571 640
k=5	123 452 522 496	k=12	17 138 079 429 120	k=19	46 490 458 680
k=6	462 946 959 360	k=13	15 819 765 626 880	k=20	3 486 784 401

Statistics - with combinatorics

If we have combinatorics available for directly inferring statistics, we can also deal with larger samples and larger populations. Here too, the calculations remain the same in principle. Again, given a sample with a certain **relative frequency**, we will determine the **probability** of a **population** by dividing the number of samples from this population by the number of all possible samples.

We now assume a sample of size $\boldsymbol{n}$ with $\boldsymbol{k}$ red balls and look for the probability of drawing this sample from a population with $\boldsymbol{g}$ balls, of which $\boldsymbol{r}$ are red. In other words, we are looking for the probability with which we can draw a given $(\boldsymbol{n},\boldsymbol{k})$-sample from a $(\boldsymbol{g},\boldsymbol{r})$-population.

Let's take a look at a specific example: We have given a sample of size $\boldsymbol{n} = 10$ with $\boldsymbol{k} = 6$ red balls and populations with $\boldsymbol{g} = 20$ balls each. We ask for the probability with which a sample can be drawn from a population with $\boldsymbol{r} = 13$ red balls. We draw ordered samples with replacement.



n = 10 **g = 20**

k = 6 **r = 13**

We first calculate the number of all possible samples that can be drawn from a population of 20 balls, 13 of which are red.

$$\mathcal{S}\big((20,13)\big) = \binom{10}{6}\cdot 13^6 \cdot (20-13)^{10-6}$$

Then we calculate the number of all possible samples that can be drawn from populations with 20 balls.

$$\sum_{i=0}^{20}\binom{10}{6}\cdot i^6\cdot(20-i)^{10-6}$$

The quotient of both numbers is the probability of drawing a given sample of size $\boldsymbol{n}=10$ with $\boldsymbol{k}=6$ red balls from a population with $\boldsymbol{g}=20$ balls, of which $\boldsymbol{r}=13$ are red.

$$\mathcal{P}\big((20,13)\big) = \frac{\binom{10}{6}\cdot 13^6\cdot(20-13)^{10-6}}{\sum_{i=0}^{20}\binom{10}{6}\cdot i^6\cdot(20-i)^{10-6}}$$

Hence we have:

$$\mathcal{P}\big((20,13)\big) = \frac{2433725365890}{18618197650500}$$

We have thus proceeded in a very similar way to the probability calculation. There we calculated the probability of a possible sample from a given population; here we calculate the probability of a possible population for a given sample.

$$\mathcal{P}\big((20,13)\big) \approx 0{,}1307$$

Statistics - with combinatorics

In general, we have the number of $(\boldsymbol{n},\boldsymbol{k})$-samples from a $(\boldsymbol{g},\boldsymbol{r})$-population

$$\mathcal{S}((n,k)) = \binom{n}{k} \cdot r^k \cdot (g-r)^{n-k}$$

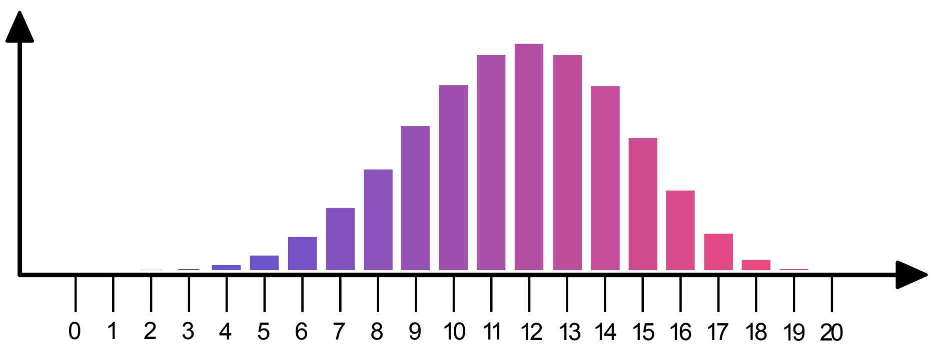
and the number of $(\boldsymbol{n},\boldsymbol{k})$-samples from all populations with $\boldsymbol{g}$ elements,

$$\sum_{i=0}^{n} \binom{n}{i} \cdot r^i \cdot (g-r)^{n-i}$$

from which we can calculate the probability of drawing a given $(\boldsymbol{n},\boldsymbol{k})$-sample from a $(\boldsymbol{g},\boldsymbol{r})$-population.

$$\mathcal{P}((g,r)) = \frac{\binom{n}{k} \cdot r^k \cdot (g-r)^{n-k}}{\sum_{i=0}^{g} \binom{n}{k} \cdot i^k \cdot (g-i)^{n-k}}$$

If we visualise the probabilities for each possible population with $\boldsymbol{g}$ = 20 balls (for a given (10 , 6) - sample) in a bar chart, a similar picture emerges as in the case of the binomial distribution. This fits with what we intuitively expect anyway: we have probably drawn a sample with a certain relative frequency of red balls from a population with a similar proportion of red balls. We may also have drawn this sample from a very dissimilar population. But that would be unlikely.

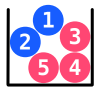


r=0	0	r=7	705 636 348 690	r=14	2 049 238 517 760
r=1	27 367 410	r=8	1 141 521 776 640	r=15	1 495 019 531 250
r=2	1 410 877 440	r=9	1 633 973 813 010	r=16	901 943 132 160
r=3	12 786 229 890	r=10	2 100 000 000 000	r=17	410 580 048 690
r=4	56 371 445 760	r=11	2 440 874 461 410	r=18	114 281 072 640
r=5	166 113 281 250	r=12	2 568 423 997 440	r=19	9 879 635 010
r=6	376 390 748 160	r=13	2 433 725 365 890	r=20	0

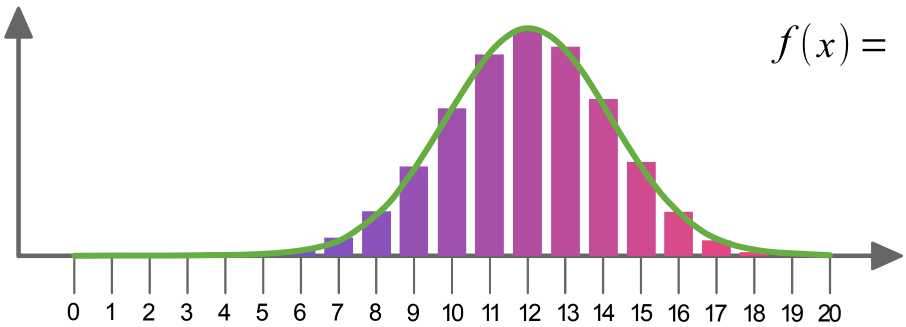
Probability and statistics with continuous functions

Probability:
If the numbers become too large with larger sample sizes, we determine the probabilities of samples by using the density function of the normal distribution.

B_{53}

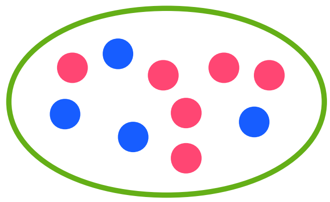


$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

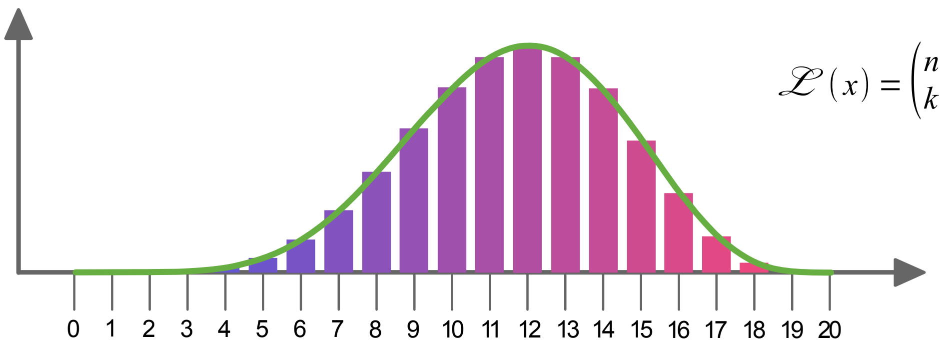


Directly Inferring Statistics:
If we do not want to limit ourselves in the number of populations from which we can draw a sample, we can use the likelihood function to determine the probabilities of arbitrary populations.
In this case, the function term of the likelihood function looks almost like the Bernoulli formula - only in the Bernoulli formula ***k*** is the variable, while the function variable ***x*** of the likelihood function stands for the probability of success (or the population proportion).

(***10*** , ***6***) - sample



$$\mathscr{L}(x) = \binom{n}{k} \cdot x^k \cdot (1-x)^{n-k}$$



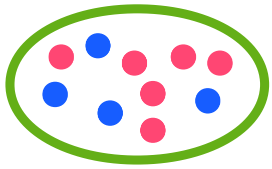
Likelihood function and probabilities of intervals

For every sample there is exactly one likelihood function. We denote the likelihood function to the

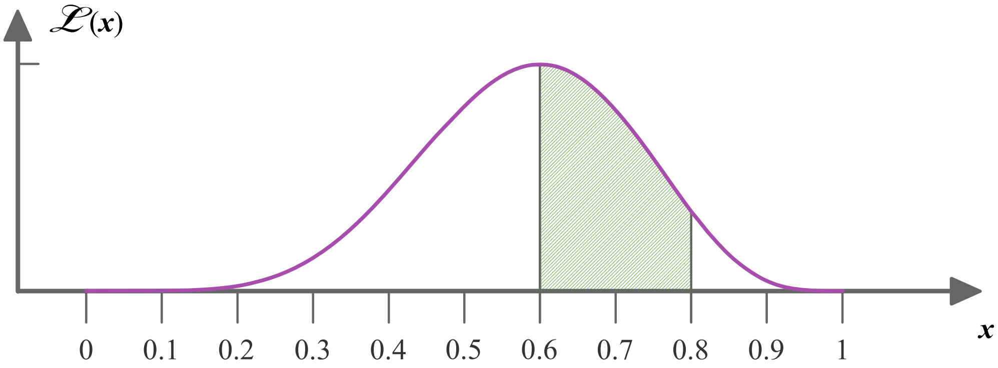
$(\boldsymbol{10}, \boldsymbol{6})$-sample by $\mathscr{L}_{10,6}$.

n = 10

k = 6



The probability of drawing an $(\boldsymbol{10}, \boldsymbol{6})$-sample from a population with a population proportion of, for example, 0.6 to 0.8 is equal to the quotient of the *marked* area and the *total* area under the curve between 0 and 1.



To calculate the probability, we first determine the area between the curve and the x-axis in the limits from 0 to 1.

$$\int_0^1 \mathscr{L}_{10,6}(x)\,dx = \int_0^1 \binom{10}{6} x^6 (1-x)^{10-6}\,dx = \frac{1}{11}$$

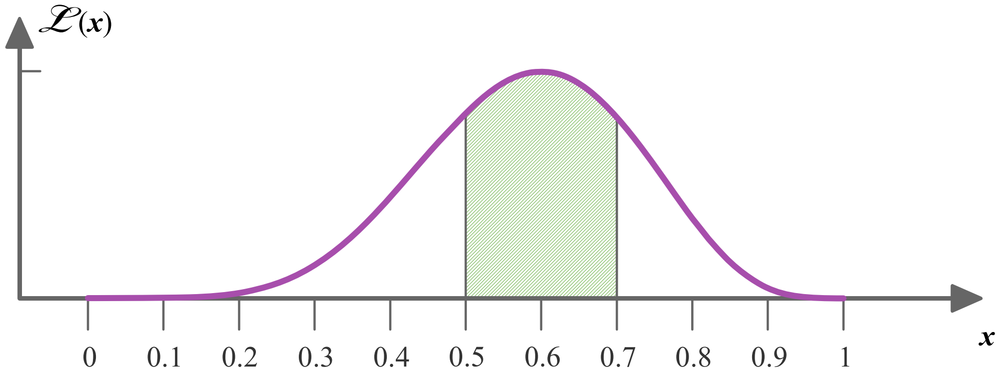
The definite integral multiplied by 11 in the limits from 0.6 to 0.8 then gives the sought probability.

$$\mathcal{P}((0.6,\,0.8)) = 11 \cdot \int_{0.6}^{0.8} \mathscr{L}_{10,6}(x)\,dx \;=\; 11 \cdot \int_{0.6}^{0.8} \binom{10}{6} x^6 (1-x)^{10-6}\,dx \approx 0.417$$

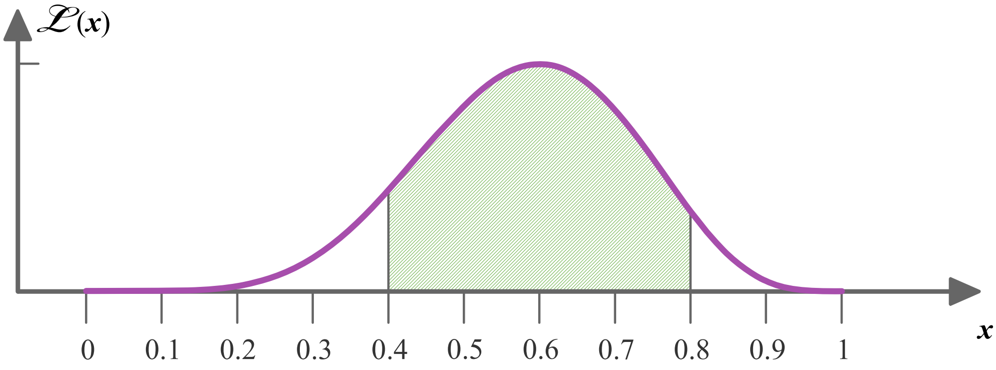
Likelihood Function - Precision

If we estimate the population proportion from a sample with the likelihood function, it happens exactly what we humans intuitively expect from such estimates. E.g.: The larger the estimate, the more probable it is.

$$\mathcal{P}((0.5,\ 0.7)) = 11 \cdot \int_{0.5}^{0.7} \mathcal{L}_{10,6}(x)\,dx = 11 \cdot \int_{0.5}^{0.7} \binom{10}{6}\ x^6 (1-x)^{10-6}\,dx \approx 0.515$$



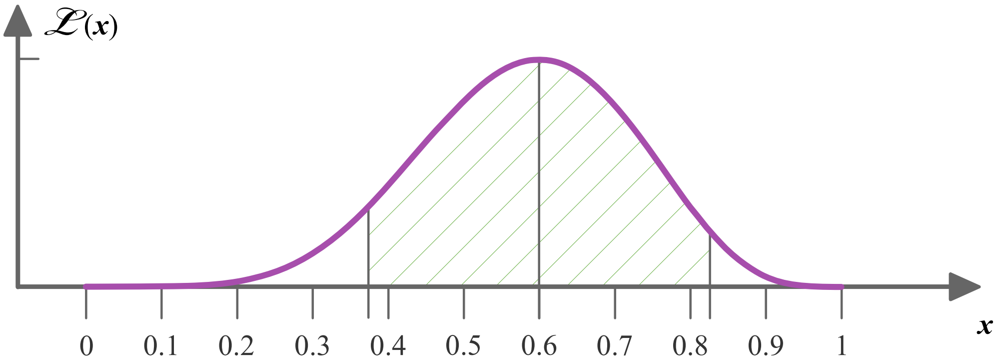
$$\mathcal{P}((0.4,\ 0.8)) = 11 \cdot \int_{0.4}^{0.8} \mathcal{L}_{10,6}(x)\,dx = 11 \cdot \int_{0.4}^{0.8} \binom{10}{6}\ x^6 (1-x)^{10-6}\,dx \approx 0.850$$



Likelihood function and sample size

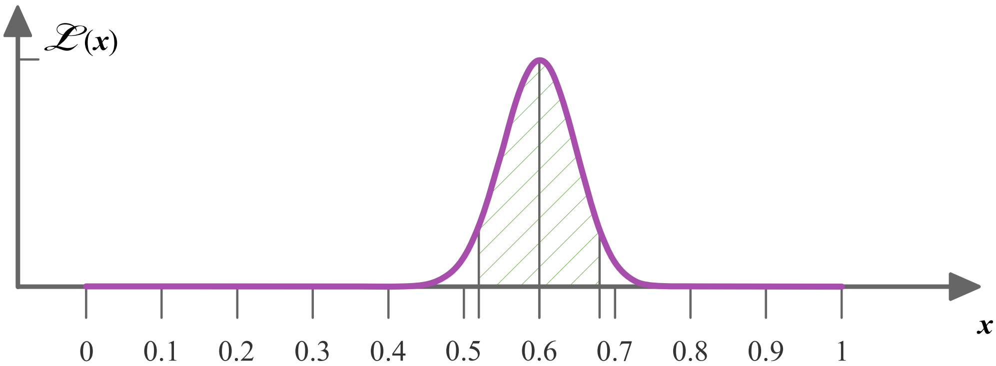
The larger the sample size, the more precise the estimate.
For sample sizes 10, 100 and 400, the interval is shown in which the population proportion lies with a probability of approx. 90 %.

$$\mathcal{P}((0.374, 0.826)) = 11 \cdot \int_{0.374}^{0.826} \mathcal{L}_{10,6}(x)\,dx = 11 \cdot \int_{0.374}^{0.826} \binom{10}{6} x^6 (1-x)^{10-6}\,dx \approx 0.900003$$



$$\mathcal{P}((0.520, 0.680)) =$$

$$101 \cdot \int_{0.520}^{0.680} \mathcal{L}_{100,60}(x)\,dx = 101 \cdot \int_{0.520}^{0.680} \binom{100}{60} x^{60} (1-x)^{100-60}\,dx \approx 0.902025$$



$$\mathcal{P}((0.5598, 0.6402)) =$$

$$401 \cdot \int_{0.5598}^{0.6402} \mathcal{L}_{400,240}(x)\,dx = 401 \cdot \int_{0.5598}^{0.6402} \binom{400}{240} x^{240} (1-x)^{400-240}\,dx \approx 0.900364$$