The Weak Law of Large Numbers

In fact, the probability P that the relative frequency of an event E lies close to the probability P_E of that event becomes larger and larger as the number of trials increases. This is, so to speak, the everyday-language version of the weak law of large numbers.

The probability that the relative frequency of an event lies within a certain interval around the probability of that event increases as the number of trials grows, and it even converges to 1.

We need a clear understanding of the connections between probability and relative frequency, or between sample and population, and we cannot wait until the mathematical tools for binomial distributions and statistical inference are introduced. We use probability from the very beginning to make statements about the future. For example, we want to know how many winning tickets we can expect if we buy 10 tickets and the winning probability is $\frac{1}{20}$. In everyday life, we often infer probability from (often very small) samples: if someone rolls a six three times in a row, something seems suspicious, doesn't it?

We will now look at two methods by which we can classify the relationships between these basic concepts. 1) From the population to the sample, or from probability to relative frequency: If we understand the empirical law of large numbers as a topic of combinatorics, we can reason mathematically exactly without relying on some vague "stabilizing" or "settling down." In other words, the weak law of large numbers has its roots in combinatorics. 2) From the sample to the population, or from relative frequency to probability: With direct inferential statistics we can work with very small populations and equally small samples and establish the relationships without guessing or estimating.

1) Let's start with the simplest random experiment imaginable: We have a box with one blue and one red ball. We draw with replacement and order matters. Now we can list all possible outcomes of drawing two, three, four, etc. times, arranged by relative frequencies.

Since the set of outcomes of, say, drawing five times is completely known, we can regard this fivefold draw as a single draw of an 8-tuple. The probability for a relative frequency of, for example, $\frac{3}{8}$ is then the ratio of the number of 8-tuples with two red balls to the total number of 8-tuples, that is $\frac{56}{256} = 0.21875$.

What do we know about the future?

What have we achieved so far? With probability theory, we want to predict what will happen in the future. If we know how many blue and red balls are in the box, then in the random experiment "drawing once from the box," we cannot predict what will happen on the next draw, but we can specify the probabilities for "blue" and "red." If we want to draw five times, we still cannot predict the outcome, but we can assign probabilities. For example: How likely is it that the relative frequency of "red" is similar to the proportion of "red" in the box? If "similar" means that the relative frequency is equal to $\frac{3}{8}$, $\frac{4}{8}$ or $\frac{5}{8}$, then we divide the number of 8-tuples with 3, 4, or 5 red balls by the total number of 8-tuples and get the probability $\frac{182}{256} = 0.7109375$.
Even though the methods used so far are still quite down-to-earth, we can already see the whole principle: Once we decide how many times we want to perform the random experiment (that is, we decide which "future" we are talking about), we can list the corresponding set of outcomes and assign probabilities to all possible future scenarios. We do not need to discuss whether the relative frequency "approaches" or "stabilizes," nor do we need to speculate about what would happen in infinitely many trials.
We see that there are more possibilities in the middle than at the edges, and that this difference increases the more often we draw. It may be a bit tedious to list and count all possibilities, but it is possible even without any combinatorial knowledge. So it becomes more and more likely that the relative frequency of red balls is close to 1/2. Thus, we move directly toward the weak law of large numbers — and that only by counting blue and red balls.
We can also see that, no matter how many times we draw, there are always results whose relative frequencies are as far as possible from the base probability. Such a result is just as likely as any other and therefore can occur as well.
And we can also see this: exact predictions are less likely than approximate ones. If we want to know how likely it is that the relative frequency is roughly 0.5, we get a greater probability than if we ask whether the relative frequency is exactly 0.5. Interestingly, the probability that the relative frequency is exactly 0.5 becomes smaller and smaller and even goes to 0, although that might not be apparent from the graphs.
The empirical law of large numbers is not correct, not only because it was not formulated by Bernoulli, but also because it is usually expressed vaguely. It cannot be correct in a mathematical sense.
We know what actually converges: Not the relative frequency of an event toward the probability of the event, but the probability that the relative frequency of an event differs from the probability of the event by more than a given amount converges to 0.
In fact, the weak law of large numbers (among other things) can be expressed in a simplified way as follows: The probability that the relative frequency of an event differs from the probability of the event by more than a given amount converges to 0 as the number of trials approaches infinity.
Also because it is often claimed that the weak law of large numbers ensures the

empirical law of large numbers or that it means the relative frequency converges to the probability, we will show here what this law actually says.
The weak law of large numbers states:

$$\lim_{n\to\infty} P\left(|h_n(E) - p| > \epsilon\right) = 0$$

where $h_n(E)$ is the relative frequency of an event E after n trials, P is the probability of the event in parentheses, n is the number of trials, p is the probability of event E, and ϵ is a real number greater than 0.
Let's apply this law to coin tossing:
Let the event E consist only of the outcome H. Then the probability p of this event is 0.5. To make things concrete, let's choose a specific number for n, say 20. So we consider the 20-fold coin toss. Every sequence of H's and T's of length 20 is a possible outcome e of the 20-fold toss. To each of these sequences, we assign the relative frequency of H. The sequence
$(H; H; H; H; H; H; H; H; H; H; H; H; H; H; H; H; H; H; H; H)$
has the relative frequency
$h_{20}(H) = \frac{20}{20} = 1$,
the sequence
$(T; T; H; H; H; T; H; T; T; T; H; H; T; H; H; H; H; T; T; H)$
has the relative frequency
$h_{20}(H) = \frac{11}{20} = 0.55$,
and the sequence
$(T; T; T; T; T; T; T; T; T; T; T; T; T; T; T; T; T; T; T; T)$
has the relative frequency
$h_{20}(H) = \frac{0}{20} = 0$.
Now we choose a value for ϵ. According to the weak law of large numbers, this ϵ should be any number greater than 0. Let's choose 0.1.
With $\epsilon = 0.1$ and the relative frequency $h_{20}(H)$ for each outcome, we can now define events — for example, the event that contains all outcomes whose relative frequency of H differs by more than 0.1 from the probability $p(H) = 0.5$. Then, for example, the outcomes
$(H; H; H; H; H; H; H; H; H; H; H; H; H; H; H; H; H; H; H; H)$ and
$(H; H; H; H; H; H; H; H; T; H; H; H; H; H; H; H; H; T; H; H)$
are elements of this event, while
$(T; T; H; H; H; T; H; T; T; T; H; H; T; H; H; H; H; T; T; H)$
is not an element of this event, because the relative frequency $h_{20}(H) = \frac{11}{20} = 0.55$ does not differ by more than 0.1 from 0.5.
We will denote this event as follows:

$$\left|h_{20}(H) - 0.5\right| > 0.1$$

Each outcome e_{20} of the 20-fold coin toss — that is, each 20-sequence of H's and T's — has the probability

$$p(e_{20}) = \frac{1}{2^{20}}.$$

Since all outcomes have the same probability, we can assign a probability to any event by dividing the number of outcomes in that event by 2^{20}.
Thus, our event $\left|h_{20}(H) - 0.5\right| > 0.1$ has a probability that we denote as
$P\left(\left|h_{20}(H) - 0.5\right| > 0.1\right)$.
We can now form a sequence of probabilities by substituting $n = 21, 22, 23, \ldots$. This sequence then has the terms
$P\left(\left|h_{21}(H) - 0.5\right| > 0.1\right)$,
$P\left(\left|h_{22}(H) - 0.5\right| > 0.1\right)$,
$P\left(\left|h_{23}(H) - 0.5\right| > 0.1\right)$, etc.
If we were to calculate these probabilities, we would find that they become smaller and smaller. In fact, we would find that these probabilities can get arbitrarily close to 0 as long as we choose n large enough. This simply means that the limit of this sequence of probabilities is 0. We express this as

$$\lim_{n\to\infty} P\left(\left|h_n(H) - 0.5\right| > 0.1\right) = 0.$$

Thus, we have what the weak law of large numbers states when E is the event that heads occur in a coin toss, $p = 0.5$, and $\epsilon = 0.1$.
In everyday language, this means that the probability that the relative frequency of H differs from the probability of H — namely 0.5 — by more than 0.1 becomes smaller and smaller and even approaches 0 if we repeat the coin toss often enough. Since we can choose ϵ freely, we can also use 0.01 or 0.001 instead of 0.1. This means: Even if we make the deviation of event H from its probability very small, the probability of such a deviation becomes smaller and smaller if we repeat the random experiment often enough. Conversely, this means: As the number of coin tosses increases, the probability that the relative frequency of heads is arbitrarily close to 0.5 also becomes larger and even approaches 1 if we repeat the coin toss infinitely many times.
By the way, the reasoning for convergence in probability of the relative frequency when there are more than two possible outcomes of the underlying random experiment lies in the multinomial distributions: the more evenly the numbers in the denominator are distributed, the smaller the product of the factorials becomes. Then the number of permutations becomes the largest. For example, if we have the balls 1, 2, ..., 6 in the box (that is, like rolling an ideal die), we compute the probabilities of the different relative frequencies with respect to the counting measure (counting density) using the multinomial distribution

$$p(n_1; \ldots; n_6) = \binom{n}{n_1; \ldots; n_6} \left(\frac{1}{6}\right)^{n_1} \times \ldots \times \left(\frac{1}{6}\right)^{n_6}$$

where n is the number of draws. Since $n_1 + \ldots + n_6 = n$, the different probabilities arise only from the multinomial coefficients

$$\binom{n}{n_1; \ldots; n_6} = \frac{n!}{n_1! \times \ldots \times n_6!}$$

For example, if we draw 60 times, the denominator of the multinomial coefficient is smallest when $n_1 = ... = n_6 = 10$.
But then the relative frequency, for example of "2," is also equal to $\frac{1}{6}$.