http://dx.doi.org/10.1090/stml/028/06

Chapter 5

The Weak Law of Large Numbers

The probability of success for a random experiment is usually obtained experimentally from the frequency of success in a sequence of (identical and independent) repetitions of that experiment. However, in our mathematical model the probability of success is given a priori. The law of large numbers reconciles these two notions of probability. In fact, it shows that for a large number of trials, it is very probable that the frequency of success is close to the theoretical probability. This is the first justification for our mathematical model of probability.

Consider the situation studied in the preceding section in which $\Omega_n = \{0,1\}^n$ is the space of outcomes for n trials of the experiment. Supposing that these trials are identical and independent, the space Ω_n is equipped with the product probability $P_n = (1-p,p)^{\otimes n}$, where p is a parameter between 0 and 1 that represents the theoretic probability of success (see Section 1.2). As before, we let S_n be the random variable that counts the number of successes

$$S_n(\omega) = \omega_1 + \omega_2 + \cdots + \omega_n, \qquad \text{where } \omega = (\omega_1, \omega_2, \ldots, \omega_n) \in \Omega_n.$$

The random variable S_n follows a binomial distribution with parameters n and p.

The *empirical* (or *experimental*) *probability* of success is $\frac{S_n}{n}$. For large n, we expect this frequency to be close to p. The following theorem says that this is very likely to be true. The result appeared for the first time in a posthumously published work by Jacob Bernoulli.[1]

Theorem 5.1 (weak law of large numbers). *For each $\epsilon > 0$,*

$$P_n\left(\left|\frac{S_n}{n} - p\right| > \epsilon\right) \longrightarrow 0$$

as n approaches infinity, and this convergence is uniform in p.

This was named the *weak law of large numbers* by Siméon Denis Poisson,[2] who generalized Bernoulli's result to cases where the probability of success varies from trial to trial. (We will discuss this result later; see Theorem 11.12.)

Proof. The variance of the random variable S_n is $\text{var}(S_n) = np(1-p)$. By the Bienaymé–Chebyshev inequality (Corollary 2.2),

$$P_n\left(|S_n - np| > n\epsilon\right) \le \frac{1}{(n\epsilon)^2}\text{var}(S_n) = \frac{p(1-p)}{n\epsilon^2} \le \frac{1}{4n\epsilon^2},$$

which proves the theorem. □

In the case of tossing a fair coin ($p = 1/2$), the weak law of large numbers says that the proportion of sequences of 0's and 1's of length n in which the frequency of 1's differs from $1/2$ by less than ϵ tends to 1 as n approaches infinity.

Bernoulli's weak law of large numbers has a nice application to the problem of uniformly approximating a continuous function on an interval of $\mathbb{R}$ by a polynomial. The Weierstrass approximation theorem states that, for any real function f defined and continuous on a closed and bounded interval $[a, b]$ of $\mathbb{R}$ and for every $\epsilon > 0$, there exists a real polynomial function g such that

$$\sup_{a \le x \le b} |f(x) - g(x)| < \epsilon.$$

---

[1] J. Bernoulli, *Ars conjectandi*, 1713.

[2] S. D. Poisson, *Recherches sur la probabilité des jugements en matière criminelle et en matiére civile*, Paris, 1837.

Serge Bernstein[3] gave a proof of this result using the weak law of large numbers. His method has the advantage that it gives an explicit formula for the approximating polynomials.

Proposition 5.2 (Bernstein polynomials). *Let f be a real function that is defined and continuous on the interval* $[0,1]$. *Then*

$$\sup_{0\le x\le 1}\left| f(x)-\sum_{k=0}^{n}\binom{n}{k} f\left(\frac{k}{n}\right) x^k(1-x)^{n-k}\right| \longrightarrow 0$$

as n approaches infinity.

This theorem implies that the polynomials approximating f on an arbitrary interval $[a,b]$ are given by

$$x\mapsto \sum_{k=0}^{n}\binom{n}{k} f\left(a+(b-a)\frac{k}{n}\right)\left(\frac{x-a}{b-a}\right)^k\left(\frac{b-x}{b-a}\right)^{n-k}.$$

Proof. Fix an $\epsilon>0$. Since f is uniformly continuous, there exists an $\eta>0$ such that

$$0\le x,y\le 1 \text{ and } |x-y|<\eta \Rightarrow |f(x)-f(y)|<\epsilon.$$

Now, consider the probability space (Ω_n,P_n) as previously defined and the random variable $f(\frac{S_n}{n})$. Bernstein's polynomials appear as the expected value of this random variable; indeed,

$$E_n\left[f\left(\frac{S_n}{n}\right)\right]=\sum_{k=0}^{n} f\left(\frac{k}{n}\right)P_n(S_n=k)$$
$$=\sum_{k=0}^{n}\binom{n}{k} f\left(\frac{k}{n}\right)p^k(1-p)^{n-k}.$$

By the law of large numbers, there exists an integer n_0, independent of the parameter p, such that

$$P_n\left(\left|\frac{S_n}{n}-p\right|>\eta\right)<\epsilon$$

for every $n\ge n_0$. We have

$$\left|E_n\left[f\left(\frac{S_n}{n}\right)\right]-f(p)\right|=\left|\sum_{k=0}^{n}\left(f\left(\frac{k}{n}\right)-f(p)\right)P_n(S_n=k)\right|,$$

[3] S. Bernstein, *Démonstration du théorème de Weierstrass fondée sur le calcul des probabilités*, Soobsch. Charkovskovo Mat. Obsch., vol. 13, pp. 1–2, 1912.

and the triangle inequality implies that an upper bound for this expression is

$$\sum_{|\frac{k}{n}-p|\le\eta}\left|f\left(\frac{k}{n}\right)-f(p)\right|P_n(S_n=k)$$
$$+\sum_{|\frac{k}{n}-p|>\eta}\left(\left|f\left(\frac{k}{n}\right)\right|+|f(p)|\right)P_n(S_n=k)$$
$$\le\sum_{|\frac{k}{n}-p|\le\eta}\epsilon P_n(S_n=k)+\sum_{|\frac{k}{n}-p|>\eta}2\sup_{0\le x\le 1}|f(x)|P_n(S_n=k)$$
$$=\epsilon P_n\left(\left|\frac{S_n}{n}-p\right|\le\eta\right)+2\sup_{0\le x\le 1}|f(x)|P_n\left(\left|\frac{S_n}{n}-p\right|>\eta\right).$$

Thus for every $n\ge n_0$,

$$\left|E_n\left[f\left(\frac{S_n}{n}\right)\right]-f(p)\right|\le\epsilon+2\epsilon\sup_{0\le x\le 1}|f(x)|.$$

This proves that $\left|E_n\left[f\left(\frac{S_n}{n}\right)\right]-f(p)\right|$ can be made arbitrarily small, uniformly with respect to p, by picking a large enough n. □