13. The Weak Law and the Strong Law of Large Numbers

James Bernoulli proved the weak law of large numbers (WLLN) around 1700 which was published posthumously in 1713 in his treatise *Ars Conjectandi.* Poisson generalized Bernoulli's theorem around 1800, and in 1866 Tchebychev discovered the method bearing his name. Later on one of his students, Markov observed that Tchebychev's reasoning can be used to extend Bernoulli's theorem to dependent random variables as well.

In 1909 the French mathematician Emile Borel proved a deeper theorem known as the strong law of large numbers that further generalizes Bernoulli's theorem. In 1926 Kolmogorov derived conditions that were necessary and sufficient for a set of mutually independent random variables to obey the law of large numbers.[1]

Let X_i be independent, identically distributed Bernoulli random Variables such that

$$P(X_i) = p, \qquad P(X_i = 0) = 1 - p = q,$$

and let $k = X_1 + X_2 + \cdots + X_n$ represent the number of "successes" in n trials. Then the weak law due to Bernoulli states that [see Theorem 3-1, page 58, Text]

$$P\left\{ \left| \tfrac{k}{n} - p \right| > \varepsilon \right\} \le \frac{pq}{n\varepsilon^2}. \qquad (13\text{-}1)$$

i.e., the ratio "total number of successes to the total number of trials" tends to *p in probability* as *n* increases.

A stronger version of this result due to Borel and Cantelli states that the above ratio *k/n* tends to *p* not only *in probability*, but *with probability* 1. This is the strong law of large numbers (SLLN).

What is the difference between the weak law and the strong law?

The strong law of large numbers states that if $\{\varepsilon_n\}$ is a sequence of positive numbers converging to zero, then

$$\sum_{n=1}^{\infty} P\left\{\left|\tfrac{k}{n} - p\right| \geq \varepsilon_n\right\} < \infty. \tag{13-2}$$

From Borel-Cantelli lemma [see (2-69) Text], when (13-2) is satisfied the events $A_n \overset{\Delta}{=} \left\{\left|\tfrac{k}{n} - p\right| \geq \varepsilon_n\right\}$ can occur only for a finite number of indices n in an infinite sequence, or equivalently, the events $\left\{\left|\tfrac{k}{n} - p\right| < \varepsilon_n\right\}$ occur infinitely often, i.e., the event k/n converges to p *almost-surely.*

Proof: To prove (13-2), we proceed as follows. Since

$$\left|\tfrac{k}{n} - p\right| \geq \varepsilon \quad \Rightarrow \quad \left|k - np\right|^4 \geq \varepsilon^4 n^4$$

we have

$$\sum_{k=0}^{n} (k-np)^4 p_n(k) \geq \varepsilon^4 n^4 = \varepsilon^4 n^4 \left(P\left\{ \left| \tfrac{k}{n} - p \right| \geq \varepsilon \right\} + P\left\{ \left| \tfrac{k}{n} - p \right| < \varepsilon \right\} \right)$$

and hence

$$P\left\{ \left| \tfrac{k}{n} - p \right| \geq \varepsilon \right\} \leq \frac{\sum_{k=0}^{n} (k-np)^4 p_n(k)}{\varepsilon^4 n^4} \tag{13-3}$$

where

$$p_n(k) = P\left\{ \sum_{i=1}^{n} X_i = k \right\} = \binom{n}{k} p^k q^{n-k}$$

By direct computation

$$\sum_{k=0}^{n} (k-np)^4 p_n(k) = E\left\{ \left(\sum_{i=1}^{n} X_i - np \right)^4 \right\} = E\left\{ \left(\sum_{i=1}^{n} (X_i - p) \right)^4 \right\}$$

4

$$= E\{(\sum_{i=1}^{n} Y_i)^4\} = \sum_{i=1}^{n}\sum_{k=1}^{n}\sum_{j=1}^{n}\sum_{l=1}^{n} E(Y_i Y_k Y_j Y_l)$$

$i = 1 \to n$ can coincide with j, k or l, and the second variable takes (n-1) values

$$= \sum_{i=1}^{n} E(Y_i^4) + 4n(n-1)\sum_{i=1}^{n}\sum_{j=1}^{n} E(Y_i^3)\overset{0}{E(Y_j)} + 3n(n-1)\sum_{i=1}^{n}\sum_{j=1}^{n} E(Y_i^2)E(Y_j^2)$$

$$= n(p^3 + q^3)pq + 3n(n-1)(pq)^2 \leq [n + 3n(n-1)]pq$$

$$= 3n^2 pq, \tag{13-4}$$

since
$$p^3 + q^3 = (p+q)^3 - 3p^2q - 3pq^2 < 1, \quad pq \leq 1/2 < 1$$

Substituting (13-4) also (13-3) we obtain

$$P\left\{\left|\frac{k}{n} - p\right| \geq \varepsilon\right\} \leq \frac{3pq}{n^2\varepsilon^4}$$

Let $\varepsilon = \frac{1}{n^{1/8}}$ so that the above integral reads
and hence

$$\sum_{n=1}^{\infty} P\left\{\left|\frac{k}{n} - p\right| \geq \frac{1}{n^{1/8}}\right\} \leq 3pq\sum_{n=1}^{\infty}\frac{1}{n^{3/2}} \leq 3pq\left(1 + \int_1^{\infty} x^{-3/2}dx\right)$$

$$= 3pq(1+2) = 9pq < \infty, \tag{13-5}$$

thus proving the strong law by exhibiting a sequence of positive numbers $\varepsilon_n = 1/n^{1/8}$ that converges to zero and satisfies (13-2).

      We return back to the same question: "What is the difference between the weak law and the strong law?."

      The weak law states that for every n that is large enough, the ratio $(\sum_{i=1}^{n} X_i)/n = k/n$ is likely to be near p with certain probability that tends to 1 as n increases. However, it does not say that k/n is bound to stay near p if the number of trials is increased. Suppose (13-1) is satisfied for a given ε in a certain number of trials n_0. If additional trials are conducted beyond n_0, the weak law does not guarantee that the new k/n is bound to stay near p for such trials. In fact there can be events for which $k/n > p + \varepsilon$, for $n > n_0$ in some regular manner. The probability for such an event is the sum of a large number of very small probabilities, and the weak law is unable to say anything specific about the convergence of that sum.

      However, the strong law states (through (13-2)) that not only all such sums converge, but the total number of all such events

6

where $k/n > p + \varepsilon$ is in fact finite! This implies that the probability $\{|\frac{k}{n} - p| > \varepsilon\}$ of the events as n increases becomes and remains small, since with probability 1 only finitely many violations to the above inequality takes place as $n \to \infty$.

Interestingly, if it possible to arrive at the same conclusion using a powerful bound known as *Bernstein's inequality* that is based on the WLLN.

Bernstein's inequality : Note that

$$\left|\tfrac{k}{n} - p\right| > \varepsilon \qquad \Rightarrow \qquad k > n(p+\varepsilon)$$

and for any $\lambda > 0$, this gives $e^{\lambda(k-n(p+\varepsilon))} > 1$. Thus

$$P\{\tfrac{k}{n} - p > \varepsilon\} = \sum_{k=\lceil n(p+\varepsilon)\rceil}^{n} \binom{n}{k} p^k q^{n-k}$$

$$\leq \sum_{k=\lceil n(p+\varepsilon)\rceil}^{n} e^{\lambda(k-n(p+\varepsilon))} \binom{n}{k} p^k q^{n-k}$$

$$\leq \sum_{k=0}^{n} e^{\lambda(k-n(p+\varepsilon))} \binom{n}{k} p^k q^{n-k}$$

$$P\{\tfrac{k}{n} - p > \varepsilon\} = e^{-\lambda n \varepsilon} \sum_{k=0}^{n} \binom{n}{k} (pe^{\lambda q})^k (qe^{-\lambda p})^{n-k}$$

$$= e^{-\lambda n \varepsilon} (pe^{\lambda q} + qe^{-\lambda p})^n. \tag{13-6}$$

Since $e^x \leq x + e^{x^2}$ for any real x,

$$pe^{\lambda q} + qe^{-\lambda p} \leq p(\lambda q + e^{\lambda^2 q^2}) + q(-\lambda p + e^{\lambda^2 p^2})$$

$$= pe^{\lambda^2 q^2} + qe^{\lambda^2 p^2} \leq e^{\lambda^2}. \tag{13-7}$$

Substituting (13-7) into (13-6), we get

$$P\{\tfrac{k}{n} - p > \varepsilon\} \leq e^{\lambda^2 n - \lambda n \varepsilon}.$$

But $\lambda^2 n - \lambda n \varepsilon$ is minimum for $\lambda = \varepsilon / 2$ and hence

$$P\{\tfrac{k}{n} - p > \varepsilon\} \leq e^{-n\varepsilon^2/4}, \qquad \varepsilon > 0. \tag{13-8}$$

Similarly

$$P\{\tfrac{k}{n} - p < -\varepsilon\} \leq e^{-n\varepsilon^2/4}$$

and hence we obtain *Bernstein's inequality*

$$P\{\left|\tfrac{k}{n} - p\right| > \varepsilon\} \le 2\, e^{-n\varepsilon^2/4}. \qquad (13\text{-}9)$$

Bernstein's inequality is more powerful than Tchebyshev's inequality as it states that the chances for the relative frequency k/n exceeding its probability p tends to zero exponentially fast as $n \to \infty$.

Chebyshev's inequality gives the probability of k/n to lie between $p - \varepsilon$ and $p + \varepsilon$ for a specific n. We can use Bernstein's inequality to estimate the probability for k/n to lie between $p - \varepsilon$ and $p + \varepsilon$ for *all* large n.

Towards this, let

$$y_n = \{p - \varepsilon \le \tfrac{k}{n} < p + \varepsilon\}$$

so that

$$P(y_n^c) = P\{\left|\tfrac{n}{k} - p\right| > \varepsilon\} \le 2\, e^{-n\varepsilon^2/4}$$

To compute the probability of the event $\bigcap_{n=m}^{\infty} y_n$, note that its complement is given by $(\bigcap_{n=m}^{\infty} y_n)^c = \bigcup_{n=m}^{\infty} y_n^c$

9

PILLAI

and using Eq. (2-68) Text,

$$P(\bigcup_{n=m}^{\infty} y_n^c) \le \sum_{n=m}^{\infty} P(y_n^c) \le \sum_{n=m}^{\infty} 2\, e^{-n\varepsilon^2/4} = \frac{2e^{-m\varepsilon^2/4}}{1-e^{-\varepsilon^2/4}}.$$

This gives

$$P(\bigcap_{n=m}^{\infty} y_n) = \{1 - P(\bigcup_{n=m}^{\infty} \bar{y}_n)\} \ge 1 - \frac{2\,e^{-m\varepsilon^2/4}}{1-e^{-\varepsilon^2/4}} \to 1 \quad \text{as} \quad m \to \infty$$

or,

$$P\{p-\varepsilon \le \tfrac{k}{n} \le p+\varepsilon, \;\; \textit{for all} \;\; n \ge m\} \to 1 \quad \text{as} \quad m \to \infty.$$

Thus k/n is bound to stay near p for all large enough n, in probability, a conclusion already reached by the SLLN.

Discussion: Let $\varepsilon = 0.1$. Thus if we toss a fair coin 1,000 times, from the weak law

$$P\{|\tfrac{k}{n} - \tfrac{1}{2}| \ge 0.01\} \le \frac{1}{40}.$$

Thus on the average 39 out of 40 such events each with 1000 or more trials will satisfy the inequality $\left\{\left|\frac{k}{n} - \frac{1}{2}\right| \leq 0.1\right\}$ or, it is quite possible that one out of 40 such events may not satisfy it. As a result if we continue the coin tossing experiment for an additional 1000 more trials, with k representing the total number of successes up to the current trial n, for $n = 1000 \rightarrow 2000$, it is quite possible that for few such n the above inequality may be violated. This is still consistent with the weak law, but "not so often" says the strong law. According to the strong law such violations can occur only a finite number of times each with a finite probability in an infinite sequence of trials, and hence almost always the above inequality will be satisfied, i.e., the sample space of k/n coincides with that of p as $n \rightarrow \infty$.

Next we look at an experiment to confirm the strong law:

Example: $2n$ red cards and $2n$ black cards (all distinct) are shuffled together to form a single deck, and then split into half. What is the probability that each half will contain n red and n black cards?[11]

Solution: From a deck of $4n$ cards, $2n$ cards can be chosen in $\binom{4n}{2n}$ different ways. To determine the number of favorable draws of n red and n black cards in each half, consider the unique draw consisting of $2n$ red cards and $2n$ black cards in each half. Among those $2n$ red cards, n of them can be chosen in $\binom{2n}{n}$ different ways; similarly for each such draw there are $\binom{2n}{n}$ ways of choosing n black cards. Thus the total number of favorable draws containing n red and n black cards in each half are $\binom{2n}{n}\binom{2n}{n}$ among a total of $\binom{4n}{2n}$ draws. This gives the desired probability p_n to be

$$p_n \approx \frac{\binom{2n}{n}\binom{2n}{n}}{\binom{4n}{2n}} = \frac{(2n!)^4}{(4n)!(n!)^4}.$$

For large n, using Stingling's formula we get

$$p_n \approx \frac{[\sqrt{2\pi(2n)}\,(2n)^{2n}e^{-2n}]^4}{\sqrt{2\pi(4n)}\,(4n)^{4n}e^{-4n}[\sqrt{2\pi n}\,n^n e^{-n}]^4} = \sqrt{\frac{2}{\pi n}}$$

For a full deck of 52 cards, we have $n = 13,$ which gives

$$p_n \approx 0.221$$

and for a partial deck of 20 cards (that contains 10 red and 10 black cards), we have $n = 5$ and $p_n \approx 0.3568.$
One summer afternoon, 20 cards (containing 10 red and 10 black cards) were given to a 5 year old child. The child split that partial deck into two equal halves and the outcome was declared a success if each half contained exactly 5 red and 5 black cards. With adult supervision (in terms of shuffling) the experiment was repeated 100 times that very same afternoon. The results are tabulated below in Table 13.1, and the relative frequency vs the number of trials plot in Fig 13.1 shows the convergence of k/n to p.

Table 13.1

Expt	Number of successes	Expt	Number of successes	Expt	Number of successes	Expt	Number of successes	Expt	Number of successes
1	0	21	8	41	14	61	23	81	29
2	0	22	8	42	14	62	23	82	29
3	1	23	8	43	14	63	23	83	30
4	1	24	8	44	14	64	24	84	30
5	2	25	8	45	15	65	25	85	30
6	2	26	8	46	16	66	25	86	31
7	3	27	9	47	17	67	25	87	31
8	4	28	10	48	17	68	25	88	32
9	5	29	10	49	17	69	26	89	32
10	5	30	10	50	18	70	26	90	32
11	5	31	10	51	19	71	26	91	33
12	5	32	10	52	20	72	26	92	33
13	5	33	10	53	20	73	26	93	33
14	5	34	10	54	21	74	26	94	34
15	6	35	11	55	21	75	27	95	34
16	6	36	12	56	22	76	27	96	34
17	6	37	12	57	22	77	28	97	34
18	7	38	13	58	22	78	29	98	34
19	7	39	14	59	22	79	29	99	34
20	8	40	14	60	22	80	29	100	35

The figure below shows results of an experiment of 100 trials.

Fig 13.1