Chebyshev's inequality,

$$\mu_n\left\{k\in[n]\,:\,\left|\frac{\psi(k)-\mathbf{E}_n[\psi]}{\log\log n}\right|>\delta\right\}\le\frac{\mathrm{Var}_n(\psi)}{\delta^2(\log\log n)^2}=O\left(\frac{1}{\log\log n}\right).$$

From the asymptotics $\mathbf{E}_n[\psi]=\log\log n+O(n^{-\frac{3}{4}})$ we also get (for n large enough)

$$\mu_n\left\{k\in[n]\,:\,\left|\frac{\psi(k)}{\log\log n}-1\right|>\delta\right\}\le\frac{\mathrm{Var}_n(\psi)}{\delta^2(\log\log n)^2}=O\left(\frac{1}{\log\log n}\right).\qquad\blacksquare$$

Exercise 2.18. $\sum\limits_{p\le\sqrt[4]{n}}\frac{1}{p}\sim\log\log n$

2.6. Weak law of large numbers

If a fair coin is tossed 100 times, we expect that the number of times it turns up heads is close to 50. What do we mean by that, for after all the number of heads could be any number between 0 and 100? What we mean of course, is that the number of heads is unlikely to be far from 50. The weak law of large numbers expresses precisely this.

Theorem 2.19 (Kolmogorov). *Let $X_1,X_2\dots$ be i.i.d random variables. If $\mathbf{E}[|X_1|]<\infty$, then for any $\delta>0$, as $n\to\infty$, we have*

$$\mathbf{P}\left(\left|\frac{X_1+\ldots+X_n}{n}-\mathbf{E}[X_1]\right|>\delta\right)\to0.$$

In language to be introduced later, we shall say that S_n/n converges to zero in probability *and write $\frac{S_n}{n}\xrightarrow{P}\mathbf{E}[X_1]$*

PROOF. **Step 1:** First assume that X_i have finite variance σ^2. Without loss of generality take $\mathbf{E}[X_1]=0$ (or else replace X_i by $X_i-\mathbf{E}[X_1]$. Then, $\mu=\mathbf{E}[X_1]$. Then, by the first moment method (Chebyshev's inequality), $\mathbf{P}(|n^{-1}S_n|>\delta)\le n^{-2}\delta^{-2}\mathrm{Var}(S_n)$. By the independence of X_is, we see that $\mathrm{Var}(S_n)=n\sigma^2$. Thus, $\mathbf{P}(|\frac{S_n}{n}|>\delta)\le\frac{\sigma^2}{n\delta^2}$ which goes to zero as $n\to\infty$, for any fixed $\delta>0$.
Step 2: Now let X_i have finite expectation (which we assume is 0), but not necessarily any higher moments. Fix n and write $X_k=Y_k+Z_k$, where $Y_k:=X_k\mathbf{1}_{|X_k|\le A_n}$ and $Z_k:=X_k\mathbf{1}_{|X_k|>A_n}$ for some A_n to be chosen later. Then, Y_i are i.i.d, with some mean $\mu_n:=\mathbf{E}[Y_1]=-\mathbf{E}[Z_1]$ that depends on A_n and goes to zero as $A_n\to0$. We shall choose A_n going to infinity, so that for large enough n, we do have $|\mu_n|<\delta$ (for an arbitrary fixed $\delta>0$).

$|Y_1|\le A_n$, hence $\mathrm{Var}(Y_1)\le\mathbf{E}[Y_1^2]\le A_n\mathbf{E}[|X_1|]$. By the Chebyshev bound that we used in step 1,

$$\mathbf{P}\left(\left|\frac{S_n^Y}{n}-\mu_n\right|>\delta\right)\le\frac{\mathrm{Var}(Y_1)}{n\delta^2}\le\frac{A_n\mathbf{E}[|X_1|]}{n\delta^2}.\tag{2.7}$$

Further, if n is large, then $|\mu_n|<\delta$ and then

$$\mathbf{P}\left(\left|\frac{S_n^Z}{n}+\mu_n\right|>\delta\right)\le\mathbf{P}\left(S_n^Z\ne0\right)\le n\mathbf{P}(Z_1\ne0)=n\mathbf{P}(|X_1|>A_n).\tag{2.8}$$

Thus, writing $X_k = (Y_k - \mu_n) + (Z_k + \mu_n)$, we see that

$$\begin{aligned}
\mathbf{P}\left(\left|\frac{S_n}{n}\right| > 2\delta\right) &\le \mathbf{P}\left(\left|\frac{S_n^Y}{n} - \mu_n\right| > \delta\right) + \mathbf{P}\left(\left|\frac{S_n^Z}{n} + \mu_n\right| > \delta\right) \\
&\le \frac{A_n \mathbf{E}[|X_1|]}{n\delta^2} + n\mathbf{P}(|X_1| > A_n) \\
&\le \frac{A_n \mathbf{E}[|X_1|]}{n\delta^2} + \frac{n}{A_n}\mathbf{E}[|X_1|\ \mathbf{1}_{|X_1|>A_n}].
\end{aligned}$$

Now, we take $A_n = \alpha n$ with $\alpha := \delta^3 \mathbf{E}[|X_1|]^{-1}$. The first term clearly becomes less than δ. The second term is bounded by $\alpha^{-1}\mathbf{E}[|X_1|\ \mathbf{1}_{|X_1|>\alpha n}]$, which goes to zero as $n \to \infty$ (for any fixed choise of $\alpha > 0$). Thus, we see that

$$\limsup_{n\to\infty} \mathbf{P}\left(\left|\frac{S_n}{n}\right| > 2\delta\right) \le \delta$$

which gives the desired conclusion. ■

2.7. Applications of weak law of large numbers

We give three applications, two "practical" and one theoretical.

Application 1: Bernstein's proof of Wierstrass' approximation theorem.

Theorem 2.20. *The set of polynomials is dense in the space of continuous functions (with the sup-norm metric) on an interval of the line.*

PROOF. (**Bernstein**) Let $f \in C[0,1]$. For any $n \ge 1$, we define the *Bernstein polynomials* $Q_{f,n}(p) := \sum_{k=0}^{n} f\left(\frac{k}{n}\right)\binom{n}{k}p^k(1-p)^{n-k}$. We show that as $n \to \infty$, $\|Q_{f,n} - f\| \to 0$ which is clearly enough. To achieve this, we observe that $Q_{f,n}(p) = \mathbf{E}[f(n^{-1}S_n)]$, where S_n has Binomial(n,p) distribution. Law of large numbers enters, because Binomial may be thought of as a sum of i.i.d Bernoullis.

For $p \in [0,1]$, consider $X_1, X_2, \ldots$ i.i.d Ber(p) random variables. For any $p \in [0,1]$, we have

$$\begin{aligned}
\left|\mathbf{E}_p\left[f\left(\frac{S_n}{n}\right)\right] - f(p)\right| &\le \mathbf{E}_p\left[\left|f\left(\frac{S_n}{n}\right) - f(p)\right|\right] \\
&= \mathbf{E}_p\left[\left|f\left(\frac{S_n}{n}\right) - f(p)\right|\mathbf{1}_{\left|\frac{S_n}{n}-p\right|\le\delta}\right] + \mathbf{E}_p\left[\left|f\left(\frac{S_n}{n}\right) - f(p)\right|\mathbf{1}_{\left|\frac{S_n}{n}-p\right|>\delta}\right] \\
&\le \omega_f(\delta) + 2\|f\|\mathbf{P}_p\left(\left|\frac{S_n}{n} - p\right| > \delta\right) \qquad (2.9)
\end{aligned}$$

where $\|f\|$ is the sup-norm of f and $\omega_f(\delta) := \sup_{|x-y|<\delta}|f(x) - f(y)|$ is the modulus of continuity of f. Observe that $\text{Var}_p(X_1) = p(1-p)$ to write

$$\mathbf{P}_p\left(\left|\frac{S_n}{n} - p\right| > \delta\right) \le \frac{p(1-p)}{n\delta^2} \le \frac{1}{4\delta^2 n}.$$

Plugging this into (2.9) and recalling that $Q_{f,n}(p) = \mathbf{E}_p\left[f\left(\frac{S_n}{n}\right)\right]$, we get

$$\sup_{p\in[0,1]}\left|Q_{f,n}(p) - f(p)\right| \le \omega_f(\delta) + \frac{\|f\|}{2\delta^2 n}$$

Since f is uniformly continuous (which is the same as saying that $\omega_f(\delta) \downarrow 0$ as $\delta \downarrow 0$), given any $\epsilon > 0$, we can take $\delta > 0$ small enough that $\omega_f(\delta) < \epsilon$. With that choice of δ, we can choose n large enough so that the second term becomes smaller than ϵ. With this choice of δ and n, we get $\|Q_{f,n} - f\| < 2\epsilon$. ■

Remark 2.21. It is possible t write the proof without invoking WLLN. In fact, we did not use WLLN, but the Chebyshev bound. The main point is that the Binomial(n,p) probability measure puts almost all its mass between $np(1-\delta)$ and $np(1+\delta)$. Nevertheless, WLLN makes it transparent why this is so.

Application 2: Monte Carlo method for evaluating integrals. Consider a continuous function $f:[a,b]\to\mathbb{R}$ whose integral we would like to compute. Quite often, the form of the function may be sufficiently complicated that we cannot analytically compute it, but is explicit enough that we can numerically evaluate (on a computer) $f(x)$ for any specified x. Here is how one can evaluate the integral by use of random numbers.

Suppose $X_1, X_2, \ldots$ are i.i.d uniform($[a,b]$). Then, $Y_k := f(X_k)$ are also i.i.d with $\mathbf{E}[Y_1] = \int_a^b f(x)dx$. Therefore, by WLLN,

$$\mathbf{P}\left(\Big| \frac{1}{n}\sum_{k=1}^{n} f(X_k) \; - \; \int_a^b f(x)dx \; \Big| > \delta \right) \to 0.$$

Hence if we can sample uniform random numbers from $[a,b]$, then we can evaluate $\frac{1}{n}\sum_{k=1}^{n} f(X_k)$, and present it as an approximate value of the desired integral!

In numerical analysis one uses the same idea, but with deterministic points. The advantage of random samples is that it works irrespective of the niceness of the function. The accuracy is not great, as the standard deviation of $\frac{1}{n}\sum_{k=1}^{n} f(X_k)$ is $Cn^{-1/2}$, so to decrease the error by half, one needs to sample four times as many points.

Exercise 2.22. Since $\pi = \int_0^1 \frac{4}{1+x^2}dx$, by sampling uniform random numbers X_k and evaluating $\frac{1}{n}\sum_{k=1}^{n} \frac{4}{1+X_k^2}$ we can estimate the value of π! Carry this out on the computer to see how many samples you need to get the right value to three decimal places.

Application 3: Accuracy in sample surveys Quite often we read about sample surveys or polls, such as "do you support the war in Iraq?". The poll may be conducted across continents, and one is sometimes dismayed to see that the pollsters asked a 1000 people in France and about 1800 people in India (a much much larger population). Should the sample sizes have been proportional to the size of the population?

Behind the survey is the simple hypothesis that each person is a Bernoulli random variable (1='yes', 0='no'), and that there is a probability p_i (or p_f) for an Indian (or a French person) to have the opinion yes. Are different peoples' opinions independent? Definitely not, but let us make that hypothesis. Then, if we sample n people, we estimate p by $\overline{X_n}$ where X_i are i.i.d Ber(p). The accuracy of the estimate is measured by its mean-squared deviation $\sqrt{\text{Var}(\overline{X_n})} = \sqrt{p(1-p)}n^{-\frac{1}{2}}$. Note that this does not depend on the population size, which means that the estimate is about as accurate in India as in France, with the same sample size! This is all correct, provided that the sample size is much smaller than the total population. Even if not satisfied with the assumption of independence, you must concede that the vague feeling of unease about relative sample sizes has no basis in fact...