9 Convergence in probability

The idea is to extricate a simple deterministic component out of a random situation. This is typically possible when a large number of random effects cancel each other out, so some limit is involved. The general situation, then, is the following: given a sequence of random variables, $Y_1, Y_2, \ldots$, we want to show that, when n is large, Y_n is approximately $f(n)$ for some simple deterministic function $f(n)$. The meaning of "approximately" is what we now make clear.

A sequence, $Y_1, Y_2, \ldots$, of random variables converges to a number a *in probability* if, as $n \to \infty$, $P(|Y_n - a| \le \epsilon)$ converges to 1, for any fixed $\epsilon > 0$. This is equivalent to $P(|Y_n - a| > \epsilon) \to 0$ as $n \to \infty$, again for any fixed $\epsilon > 0$.

Example 9.1. Toss a fair coin n times, independently. Let R_n be the "longest run of heads," i.e., the longest sequence of consecutive tosses of Heads. For example, if $n = 15$ and the tosses come out

HHTTHHHTHTHTHHH.

then $R_n = 3$. We will show that, as $n \to \infty$,

$$\frac{R_n}{\log_2 n} \to 1,$$

in probability. This means that, to a first approximation, one should expect about 20 consecutive heads somewhere in a million tosses.

To solve a problem such as this, we need to find upper bounds on probabilities that R_n is large and that it is small, i.e., on $P(R_n \ge k)$ and $P(R_n \le k)$, for appropriately chosen k. Now, for arbitrary k,

$$\begin{aligned}
P(R_n \ge k) &= P\left(k \text{ consecutive Heads start at some } i,\ 0 \le i \le n-k+1\right) \\
&= P\Big(\bigcup_{i=1}^{n-k+1} \{i \text{ is the first Heads in a succession of at least } k \text{ Heads}\}\Big) \\
&\le n \cdot \frac{1}{2^k}
\end{aligned}$$

For the lower bound, divide the string of size n into disjoint blocks of size k. There is $\lfloor \frac{n}{k} \rfloor$ such blocks (if n is not divisible by k, simply throw away the leftover smaller block at the end). Then $R_n \ge k$ as soon as one of the blocks consists of Heads only, and different blocks are independendent. Therefore,

$$P(R_n < k) \le \left(1 - \frac{1}{2^k}\right)^{\lfloor \frac{n}{k} \rfloor} \le \exp\left(-\frac{1}{2^k} \left\lfloor \frac{n}{k} \right\rfloor\right),$$

using the famous inequality $1 - x \le e^{-x}$, valid for all x.

Below, we will use these trivial inequalities, valid for any real number $x \ge 2$: $\lfloor x \rfloor \ge x - 1$, $\lceil x \rceil \le x + 1$, $x - 1 \ge \frac{x}{2}$, and $x + 1 \le 2x$.

To demonstrate that $\frac{R_n}{\log_2 n} \to 1$, in probability, we need to show that, for any $\epsilon > 0$,

$$\text{(1)} \qquad P(R_n \ge (1+\epsilon)\log_2 n) \ \to \ 0,$$

$$\text{(2)} \qquad P(R_n \le (1-\epsilon)\log_2 n) \ \to \ 0,$$

as

$$\begin{aligned} P\left(\left|\frac{R_n}{\log_2 n} - 1\right| \ge \epsilon\right) &= P\left(\frac{R_n}{\log_2 n} \ge 1+\epsilon \text{ or } \frac{R_n}{\log_2 n} \le 1-\epsilon\right) \\ &= P\left(\frac{R_n}{\log_2 n} \ge 1+\epsilon\right) + P\left(\frac{R_n}{\log_2 n} \le 1-\epsilon\right) \\ &= P\left(R_n \ge (1+\epsilon)\log_2 n\right) + P\left(R_n \le (1-\epsilon)\log_2 n\right). \end{aligned}$$

A little bit of fussing in the proof comes from the fact that $(1 \pm e)\log_2 n$ are not integers. This common in problems such as this. To prove (1), we plug $k = \lfloor (1+\epsilon)\log_2 n \rfloor$ into the upper bound to get

$$\begin{aligned} P(R_n \ge (1+\epsilon)\log_2 n) &\le n \cdot \frac{1}{2^{(1+\epsilon)\log_2 n - 1}} \\ &= n \cdot \frac{2}{n^{1+\epsilon}} \\ &= \frac{2}{n^{\epsilon}} \to 0 \end{aligned}$$

as $n \to \infty$. One the other hand, to prove (2) we need to plug $k = \lceil (1-\epsilon)\log_2 n \rceil + 1$ into the lower bound,

$$\begin{aligned} P(R_n \le (1-\epsilon)\log_2 n) &\le P(R_n < k) \\ &\le \exp\left(-\frac{1}{2^k}\left\lfloor \frac{n}{k} \right\rfloor\right) \\ &\le \exp\left(-\frac{1}{2^k}\left(\frac{n}{k} - 1\right)\right) \\ &\le \exp\left(-\frac{1}{32} \cdot \frac{1}{n^{1-\epsilon}} \cdot \frac{n}{(1-\epsilon)\log_2 n}\right) \\ &= \exp\left(-\frac{1}{32}\frac{n^{\epsilon}}{(1-\epsilon)\log_2 n}\right) \\ &\to 0, \end{aligned}$$

as $n \to \infty$, as n^{ϵ} is much larger than $\log_2 n$.

The most basic tool in proving convergence in probability is *Chebyshev's inequality*: if X is a random variable with $EX = \mu$ and $\text{Var}(X) = \sigma^2$, then

$$P(|X - \mu| \geq k) \leq \frac{\sigma^2}{k^2},$$

for any $k > 0$. We proved this inequality in the previous chapter, and we will use it to prove the next theorem.

Theorem 9.1. *Connection between variance and convergence in probability.*

> *Assume that Y_n are random variables and a is a constant such that*
>
> $$EY_n \to a,$$
> $$\text{Var}(Y_n) \to 0,$$
>
> *as $n \to \infty$. Then*
>
> $$Y_n \to a,$$
>
> *as $n \to \infty$, in probability.*

Proof. Fix an $\epsilon > 0$. If n is so large that

$$|EY_n - a| < \epsilon/2,$$

then

$$\begin{aligned} P(|Y_n - a| > \epsilon) &\leq P(|Y_n - EY_n| > \epsilon/2) \\ &\leq 4\,\frac{\text{Var}(Y_n)}{\epsilon^2} \\ &\to 0, \end{aligned}$$

as $n \to \infty$. Note that the second inequality in the computation is Chebyshev's inequality. □

This is most often applied to sums of random variables. Let

$$S_n = X_1 + \ldots + X_n,$$

where X_i are random variables with finite expectation and variance. Then, *without any independence assumption*,

$$ES_n = EX_1 + \ldots + EX_n$$

and

$$E(S_n^2) = \sum_{i=1}^{n} EX_i^2 + \sum_{i \neq j} E(X_i X_j),$$

$$\text{Var}(S_n) = \sum_{i=1}^{n} \text{Var}(X_i) + \sum_{i \neq j} \text{Cov}(X_i, X_j).$$

You should recall that

$$\mathrm{Cov}(X_1, X_j) = E(X_i X_j) - EX_i EX_j$$

and

$$\mathrm{Var}(aX) = a^2 \mathrm{Var}(X).$$

Moreover, if X_i are independent,

$$\mathrm{Var}(X_1 + \ldots + X_n) = \mathrm{Var}(X_1) + \ldots + \mathrm{Var}(X_n).$$

Continuing with the review, let's reformulate and reprove the most famous convergence in probability theorem. We will use the common abbreviation i. i. d. for independent identically distributed random variables.

Theorem 9.2. *Weak law of large numbers. Let $X, X_1, X_2, \ldots$ be i. i. d. random variables with with $EX_1 = \mu$ and $\mathrm{Var}(X_1) = \sigma^2 < \infty$. Let $S_n = X_1 + \ldots + X_n$. Then, as $n \to \infty$,*

$$\frac{S_n}{n} \to \mu$$

in probability.

Proof. Let $Y_n = \frac{S_n}{n}$. We have $EY_n = \mu$, and

$$\mathrm{Var}(Y_n) = \frac{1}{n^2}\,\mathrm{Var}(S_n) = \frac{1}{n^2}\, n\,\sigma^2 = \frac{\sigma^2}{n}.$$

Thus we can simply apply the previous theorem. □

Example 9.2. We analyze a typical "investment" (the accepted euphemism for gambling on financial markets) problem. Assume you have two investment choices at the beginning of each year:

- a risk-free "bond" which returns 6% per year; and
- a risky "stock" which increases your investment by 50% with probability 0.8 and wipes it away with probability 0.2.

Putting an amount s in the bond, then, gives you $1.06s$ after a year. The same amount in the stock gives you $1.5s$ with probability 0.8 and 0 with probabiliy 0.2; note that the expected value is $0.8 \cdot 1.5s = 1.2s > 1.06s$. We will assume year-to-year independence of the stock's return.

We will try to maximize the return to our investment by "hedging." That is, we invest, at the beginning of each year, a fixed proportion x of our current capital into the stock and the remaining proportion $1 - x$ into the bond. We collect the resulting capital at the end of the year, which is simultaneously the beginning of next year, and reinvest with the same proportion x. Assume that our initial capital is x_0.

It is important to note that the *expected value* of the capital at the end of the year is maximized when $x = 1$, but using this strategy you will eventually *lose everything*. Let X_n be your capital at the end of year n. Define the *average growth rate* of your investment as

$$\lambda = \lim_{n\to\infty} \frac{1}{n} \log \frac{X_n}{x_0},$$

so that

$$X_n \approx x_0 e^{\lambda n}.$$

We will express λ in terms of x; in particular, we will show it is a nonrandom quantity.

Let $I_i = I_{\{\text{stock goes up in year } i\}}$. These are independent indicators with $EI_i = 0.8$.

$$\begin{aligned} X_n &= X_{n-1}(1-x) \cdot 1.06 + X_{n-1} \cdot x \cdot 1.5 \cdot I_n \\ &= X_{n-1}(1.06(1-x) + 1.5x \cdot I_n) \end{aligned}$$

and so we can unroll the recurrence to get

$$X_n = x_0(1.06(1-x) + 1.5x)^{S_n}((1-x)1.06)^{n-S_n},$$

where $S_n = I_1 + \ldots + I_n$. Therefore,

$$\begin{aligned} \frac{1}{n} \log \frac{X_n}{x_0} &= \frac{S_n}{n} \log(1.06 + 0.44x) + \left(1 - \frac{S_n}{n}\right) \log(1.06(1-x)) \\ &\to 0.8 \log(1.06 + .44x) + 0.2 \log(1.06(1-x)), \end{aligned}$$

in probability, as $n \to \infty$. The last expression defines λ as a function of x. To maximize this, we set $\frac{d\lambda}{dx} = 0$ to get

$$\frac{0.8 \cdot 0.44}{1.06 + 0.44x} = \frac{0.2}{1-x}.$$

The solution is $x = \frac{7}{22}$, which gives $\lambda \approx 8.1\%$.

Example 9.3. Distribute n balls independently at random into n boxes. Let N_n be the number of empty boxes. Show that $\frac{1}{n} N_n$ converges in probability and identify the limit.

Note that

$$N_n = I_1 + \ldots + I_n,$$

where $I_i = I_{\{i\text{'th box is empty}\}}$, but you cannot use the weak law of large numbers as I_i are not independent. Nevertheless,

$$EI_i = \left(\frac{n-1}{n}\right)^n = \left(1 - \frac{1}{n}\right)^n,$$

and so

$$EN_n = n \cdot \left(1 - \frac{1}{n}\right)^n.$$

Moreover,

$$E(N_n^2) = EN_n + \sum_{i \neq j} E(I_i I_j)$$

with

$$E(I_i I_j) = P(\text{box } i \text{ and } j \text{ are both empty}) = \left(\frac{n-2}{n}\right)^n,$$

so that

$$\text{Var}(N_n) = E(N_n^2) - (EN_n)^2 = n\left(1 - \frac{1}{n}\right)^n + n(n-1)\left(1 - \frac{2}{n}\right)^n - n^2\left(1 - \frac{1}{n}\right)^{2n}.$$

Now let $Y_n = \frac{1}{n} N_n$. We have

$$EY_n \to e^{-1}$$

as $n \to \infty$, and

$$\begin{aligned}
\text{Var}(Y_n) &= \frac{1}{n}\left(1 - \frac{1}{n}\right)^n + \frac{n-1}{n}\left(1 - \frac{2}{n}\right)^n - \left(1 - \frac{1}{n}\right)^{2n} \\
&\to 0 + e^{-2} - e^{-2} = 0,
\end{aligned}$$

as $n \to \infty$. Therefore

$$Y_n = \frac{N_n}{n} \to e^{-1},$$

as $n \to \infty$, in probability.

Problems

1. Assume that n married couples (amounting to $2n$ people) are seated at random on $2n$ seats around a table. Let T be the number of couples that sit together. Determine ET and $\text{Var}(T)$.

2. There are n birds that sit in a row on a wire. Each bird looks left or right with equal probability. Let N be the number of birds not seen by any neighboring bird. Determine, with proof, the constant c so that, as $n \to \infty$, $\frac{1}{n} N \to c$ in probability.

3. Recall the coupon collector problem: sample from n cards, with replacement, indefinitely, and let N be the number of cards you need to get each of n different cards are represented. Find a sequence a_n so that, as $n \to \infty$, N/a_n converges to 1 in probability.

4. Kings and Lakers are playing a "best of seven" playoff series, which means they play until one team wins four games. Assume Kings win every game independently with probability p.

(There is no difference between home and away games.) Let N be the number of games played. Compute EN and $\text{Var}(N)$.

5. An urn contains n red and m black balls. Pull balls from the urn one by one without replacement. Let X be the number of red balls you pull before any black one, and Y the number of red balls between the first and the second black one. Compute EX and EY.

Solutions to problems

1. Let I_i be the indicator of the event that the i'th couple sits together. Then $T = I_1 + \cdots + I_n$. Moreover,

$$EI_i = \frac{2}{2n-1}, \quad E(I_i I_j) = \frac{2^2(2n-3)!}{(2n-1)!} = \frac{4}{(2n-1)(2n-2)},$$

for any i and $j \neq i$. Thus

$$ET = \frac{2n}{2n-1}$$

and

$$E(T^2) = ET + n(n-1)\,\frac{4}{(2n-1)(2n-2)} = \frac{4n}{2n-1},$$

so

$$\text{Var}(T) = \frac{4n}{2n-1} - \frac{4n^2}{(2n-1)^2} = \frac{4n(n-1)}{(2n-1)^2}.$$

2. Let I_i indicate the event that bird i is not seen by any other bird. Then EI_i is $\frac{1}{2}$ if $i = 1$ or $i = n$ and $\frac{1}{4}$ otherwise. It follows that

$$EN = 1 + \frac{n-2}{4} = \frac{n+2}{4}.$$

Furthermore I_i and I_j are independent if $|i-j| \geq 3$ (two birds that have two or more birds between them are observed independently). Thus $\text{Cov}(I_i, I_j) = 0$ if $|i-j| \geq 3$. As I_i and I_j are indicators, $\text{Cov}(I_i, I_j) \leq 1$ for any i and j. For the same reason $\text{Var}(I_i) \leq 1$. Therefore

$$\text{Var}(N) = \sum_i \text{Var}(I_i) + \sum_{i \neq j} \text{Cov}(I_i, I_j) \leq n + 4n = 5n.$$

Clearly, if $M = \frac{1}{n}N$, then $EM = \frac{1}{n}EN \to \frac{1}{4}$ and $\text{Var}(M) = \frac{1}{n^2}\text{Var}(N) \to 0$. It follows that $c = \frac{1}{4}$.

3. Let N_i be the number of coupons needed to get i different coupons after having $i-1$ different ones. Then $N = N_1 + \ldots + N_n$, and N_i are independent Geometric with success probability

$\frac{n-i+1}{n}$. So

$$EN_i = \frac{n}{n-i+1}, \quad \mathrm{Var}(N_i) = \frac{n(i-1)}{(n-i+1)^2},$$

and therefore

$$EN = n\left(1 + \frac{1}{2} + \ldots + \frac{1}{n}\right),$$
$$\mathrm{Var}(N) = \sum_{i=1}^{n} \frac{n(i-1)}{(n-i+1)^2} \le n^2\left(1 + \frac{1}{2^2} + \ldots + \frac{1}{n^2}\right) \le n^2\,\frac{\pi^2}{6} < 2n^2.$$

If $a_n = n\log_2 n$, then

$$\frac{1}{a_n}\,EN \to 1, \quad \frac{1}{a_n^2}\,EN \to 0,$$

as $n \to \infty$, so that

$$\frac{1}{a_n}\,N \to 1$$

in probability.

4. Let I_i be the indicator of the event that the i'th game is played. Then $EI_1 = EI_2 = EI_3 = EI_4 = 1$,

$$EI_5 = 1 - p^4 - (1-p)^4,$$
$$EI_6 = 1 - p^5 - 5p^4(1-p) - 5p(1-p)^4 - (1-p)^5,$$
$$EI_7 = \binom{6}{3} p^3(1-p)^3.$$

Add the seven expectations to get EN. Now to compute $E(N^2)$ we use the fact that, if $i > j$, $I_i I_j = I_i$, so that $E(I_i I_j) = EI_i$. So

$$EN^2 = \sum_i EI_i + 2\sum_{i>j} E(I_i I_j) = \sum_i EI_i + 2\sum_i (i-1)EI_i = \sum_{i=1}^{7}(2i-1)EI_i,$$

and the final result can be obtained by plugging in EI_i and finally by the standard formula

$$\mathrm{Var}(N) = E(N^2) - (EN)^2.$$

5. Imagine the balls ordered in a row, and the ordering specifies the sequence in which they are pulled. Let I_i be the indicator of the event that the i'th red ball is pulled before any black ones. Then $EI_i = \frac{1}{m+1}$, simply the probability that in a random ordering of the i'th red balls and all m black ones, the red comes first. As $X = I_1 + \ldots + I_n$, $EX = \frac{n}{m+1}$.

Now let J_i be the indicator of the event that the i'th red ball is pulled between the first and the second black one. Then EJ_i is the probability that the red ball is second in the ordering of $m+1$ balls as above, so $EJ_i = EI_i$, and $EY = EX$.