Section 6

The Law of Large Numbers

Po-Ning Chen, Professor

Institute of Communications Engineering

National Chiao Tung University

Hsin Chu, Taiwan 30010, R.O.C.

The law of large numbers

- The noise and interference are in fact an **aggregated** phenomenon of a big quantity, possibly independent (or dependent).
- The load of a communication system is an **aggregated** result of quite a lot of user behavior.
- Such an **aggregation** is obtained through "summing" all the small quantities.
- As a consequence, understanding of aggregated statistic phenomenon of a big population helps the system design.
- This directs us to investigate the *law of large numbers*.
- In short, the law of large numbers is a simplified statistical model for the aggregated statistical phenomenon of a big quantity. Such a simplification makes easy the theoretical study as well as empirical study of a subject like communications.

Mean

- What is the first statistical quantity that is of general interest for a sequence of i.i.d. random variables, X_1, X_2, X_3, $\ldots$?
- Answer: Mean, i.e., $E[X_n]$.
- Question: Can we estimate the value of $E[X_n]$ in terms of $(X_1 + X_2 + \cdots + X_n)/n$?

The answer to the above question can be used to estimate any function value of X_n, i.e., $Y_n = f(X_n)$, as long as some properties hold for function $f(\cdot)$.

- What is the first statistical quantity that we are interested in for a sequence of i.i.d. random variables, $Y_1 = f(X_1)$, $Y_2 = f(X_2)$, $Y_3 = f(X_3)$, $\ldots$?
- Answer: Mean, i.e., $E[Y_n] = E[f(X_n)]$.
- Question: Can we estimate the value of $E[Y_n] = E[f(X_n)]$ in terms of

$$\frac{Y_1 + Y_2 + \cdots + Y_n}{n} = \frac{f(X_1) + f(X_2) + \cdots + f(X_n)}{n}?$$

So it suffices to have a theorem on "mean" on $X_1, X_2, X_3, \cdots$?

The Strong and Weak Laws: Theorems on Mean

- What is the key difference between the *strong law* and *weak law* of large numbers?

  (Strong Law) "limit" is placed inside braces.

$$\Pr\left[\lim_{n\to\infty}\frac{1}{n}(X_1+X_2+\cdots+X_n)=m\right]=1.$$

  (Weak Law) "limit" is placed outside braces.

$$\lim_{n\to\infty}\Pr\left[\left|\frac{1}{n}(X_1+X_2+\cdots+X_n)-m\right|<\varepsilon\right]=1 \text{ for any } \varepsilon>0.$$

  We are interested in the conditions under which the *strong law* holds, and under which the *weak law* is valid.

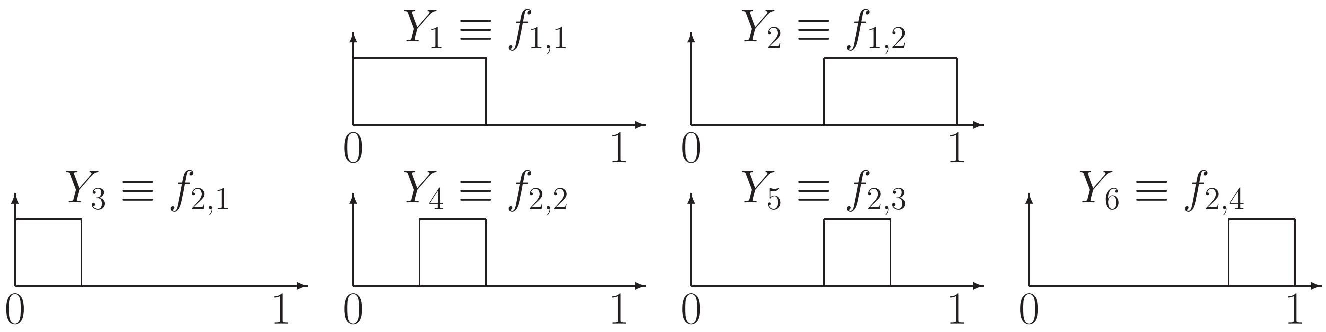
  Variants of theorems on law of large numbers basically provide *different* conditions under which these laws hold.

† In notations, we will use [ ] to represent an event (for a random variable X), such as $[X>0]$. Pr[ ] will be used to denote the probability of the concerned event, e.g., $\Pr[X>0]$, where the probability is defined through the random variable. Braces are reserved for sets, e.g., $\{x\in\mathcal{X}:x>0\}$. The probability of the concerned set under a probability measure P or P_X will be denoted by $P(\{x\in\mathcal{X}:x>0\})$ or $P_X(\{x\in\mathcal{X}:x>0\})$.

What is the difference between "placing limit inside" and "placing limit outside"?

For $i, j \in \mathbb{N}$, define a function $f_{i,j}(\cdot)$ over $[0, 1)$ as:

$$f_{i,j}(\omega) = \begin{cases} 1, & \text{if } (j-1)2^{-i} \le \omega < j2^{-i}; \\ 0, & \text{otherwise.} \end{cases}$$



Let Z be uniformly distributed over $[0, 1)$.

Define a sequence of binary random variables Y_1, Y_2, Y_3, $\ldots$ as

$$Y_n = f_{i,j}(Z),$$

where $i = \lfloor \log_2(n+1) \rfloor$ and $j = n + 2 - 2^i$.

(I.e., $Y_1, Y_2, Y_3, Y_4, Y_5, Y_6, \cdots = f_{1,1}(Z), f_{1,2}(Z), f_{2,1}(Z), f_{2,2}(Z), f_{2,3}(Z), f_{2,4}(Z), \cdots$.)

– Then as $\lim_{n\to\infty} f_{\lfloor \log_2(n+1)\rfloor,\, n+2-2^{\lfloor \log_2(n+1)\rfloor}}(z)$ does not exist for any $z \in [0,1)$, $\lim_{n\to\infty} Y_n$ does not exist; it is therefore meaningless to calculate $\Pr\left[\lim_{n\to\infty} Y_n = 0\right]$.

– However, for $0 < \varepsilon < 1$,

$$\begin{aligned}
\lim_{n\to\infty} \Pr[|Y_n - 0| < \varepsilon] &= \lim_{n\to\infty} \Pr\left[\left|f_{\lfloor \log_2(n+1)\rfloor,\, n+2-2^{\lfloor \log_2(n+1)\rfloor}}(Z)\right| < \varepsilon\right] \\
&= \lim_{n\to\infty} \Pr\left[\left|f_{\lfloor \log_2(n+1)\rfloor,\, n+2-2^{\lfloor \log_2(n+1)\rfloor}}(Z)\right| = 0\right] \\
&= \lim_{n\to\infty} \left(1 - 2^{-\lfloor \log_2(n+1)\rfloor}\right) \\
&= 1.
\end{aligned}$$

– In terminology, we say that Y_n converges to 0 in probability (limit outside), but does not converge to 0 with probability one (limit inside).

– From this example, you learn that the *strong law* is really strong in a sense that the limit of $(X_1 + X_2 + \cdots + X_n)/n$ has to exist first, while the *weak law* only requires the resultant probability "value" for each n to converge.

– The question that remains is how to validate the strong law, as the limit is placed inside the squared braces? Answer: **Borel-Cantelli lemma**.

Borel-Cantelli Lemmas

Theorem 4.3 (The First Borel-Cantelli Lemma) If $\sum_{k=1}^{\infty} P(A_k)$ converges (i.e., $\sum_{k=1}^{\infty} P(A_k) < \infty$), then

$$P\left(\limsup_{n\to\infty} A_n\right) = P(A_n \text{ i.o.}) = 0.$$

- $\limsup_{n\to\infty} A_n \equiv \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k$, named *limit superior* of the sequence of sets $\{A_n\}_{n=1}^{\infty}$.
- $\omega \in \limsup_{n\to\infty} A_n \quad \equiv \quad \omega$ belongs to A_n infinitely often (i.o.)
- **Example (Dyadic Expansion)** A dyadic expansion of $\omega \in [0,1)$ is a binary representation $.d_1 d_2 d_3 \ldots$ of it, where $\omega = \sum_{n=1}^{\infty} d_n 2^{-n}$. Notably, $d_n = d_n(\omega)$ is a function of ω for each n. E.g.,

$$d_1(\omega) = \begin{cases} 0, & \text{if } 0 \le \omega < 1/2; \\ 1, & \text{if } 1/2 \le \omega < 1. \end{cases}$$

Let $A_n = \{\omega \in [0,1) : d_n(\omega) = 0\}$. Then what is $\limsup_{n\to\infty} A_n$?

Answer:

$$\begin{aligned}
A_1 &= \{\omega \in [0,1) : d_1(\omega) = 0\} = [0, 1/2) \\
A_2 &= \{\omega \in [0,1) : d_2(\omega) = 0\} = [0, 1/4) \cup [1/2, 3/4) \\
A_3 &= \{\omega \in [0,1) : d_3(\omega) = 0\} = [0, 1/8) \cup [1/4, 3/8) \cup [1/2, 5/8) \cup [3/4, 7/8) \\
&\vdots
\end{aligned}$$

As it turns out, $\bigcup_{k=n}^{\infty} A_k = [0,1)$ for any n. Consequently,

$$\begin{aligned}
\limsup_{n\to\infty} A_n &= \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k \\
&= \bigcap_{n=1}^{\infty} [0,1) \\
&= [0,1).
\end{aligned}$$

In other words, any number in $[0,1)$ lies in $\{A_n\}_{n=1}^{\infty}$ infinitely often.

Proof of The First Borel-Cantelli Lemma:

- $\limsup_{n\to\infty} A_n = \bigcap_{m=1}^{\infty}\bigcup_{k=m}^{\infty} A_k \subset \bigcup_{k=m}^{\infty} A_k$

  $\Rightarrow P\left(\limsup_{n\to\infty} A_n\right) \le P\left(\bigcup_{k=m}^{\infty} A_k\right)$ for any m.

  $A \subset B \Rightarrow P(A) \le P(B)$

- $P\left(\bigcup_{k=m}^{\infty} A_k\right) \le \sum_{k=m}^{\infty} P(A_k)$.

- So if $\sum_{k=1}^{\infty} P(A_k)$ converges (i.e., finite), then $\lim_{m\to\infty}\sum_{k=m}^{\infty} P(A_k) = 0$, which in turns proves the theorem. □

Definition (Probability Measure) A set function P on a measurable space $(S, \mathcal{F})$ is a *probability measure*, if it satisfies the three Kolmogorov axioms (1936):

1. **(non-negativity)** $P(\mathcal{A}) \ge 0$ for $\mathcal{A} \in \mathcal{F}$.
2. **(unit measure)** $P(S) = 1$.
3. **(countable additivity or σ-additivity)** if $\mathcal{A}_1$, $\mathcal{A}_2$, ... is a disjoint sequence of sets in $\mathcal{F}$, then

$$P\left(\bigcup_{k=1}^{\infty} \mathcal{A}_k\right) = \sum_{k=1}^{\infty} P(\mathcal{A}_k).$$

Theorem 4.4 (The Second Borel-Cantelli Lemma) If $\{A_n\}_{n=1}^{\infty}$ forms an independent sequence of events for a probability measure P, and $\{\sum_{k=n}^{\infty} P(A_k)\}_{n=1}^{\infty}$ diverges (i.e., $\sum_{k=1}^{\infty} P(A_k) = \infty$), then $P\left(\limsup_{n\to\infty} A_n\right) = 1$.

Proof: For any m,

$$\begin{aligned} P\left(\bigcap_{k=m}^{\infty} A_k^c\right) &= \prod_{k=m}^{\infty} P\left(A_k^c\right) \quad \text{(by independence)} \\ &= \prod_{k=m}^{\infty} \left[1 - P\left(A_k\right)\right] \\ &\leq \exp\left[-\sum_{k=m}^{\infty} P\left(A_k\right)\right] \quad (\text{since } 1 - x \leq e^{-x})\ \forall x \in \Re \\ &= 0 \quad \text{(by divergence of sum)} \end{aligned}$$

Hence,

$$\begin{aligned} P\left(\limsup_{n\to\infty} A_n\right) &= P\left(\bigcap_{m=1}^{\infty}\bigcup_{k=m}^{\infty} A_k\right) \\ &= 1 - P\left(\bigcup_{m=1}^{\infty}\bigcap_{k=m}^{\infty} A_k^c\right) \quad \text{(by De Morgan's law)} \\ &\geq 1 - \sum_{m=1}^{\infty} P\left(\bigcap_{k=m}^{\infty} A_k^c\right) \\ &= 1 \end{aligned}$$

□

Example (Dyadic Expansion, cont.) Let d_n be the outcome of the nth toss of a fair coin, and each toss is independent of all the other tosses; hence, $\omega = .d_1d_2d_3\ldots$ is uniformly distributed over $[0, 1)$. Then

$$A_n = \{\omega \in [0, 1) : d_n(\omega) = 0\}$$

form independent events under such a probability measure.

From the two Borel-Cantelli lemmas, we found that since $\sum_{k=1}^{\infty} P(A_k)$ either converges or diverges, so $P\big(\limsup_{n\to\infty} A_n\big)$ is either 1 or 0, and cannot be any value inbetween!

Theorem (A simplified theorem of Theorem 4.5: Kolmogrov's zero-one law) If A_1, A_2, $\ldots$ are independent events under a probability measure P, then $P\big(\limsup_{n\to\infty} A_n\big)$ is either 1 or 0.

In other words, for a sequence of independent events, set of all "outcomes" that occur infinitely often is either with probability 1 (certainty) or with probability 0 (impossible)!

Strong Law of Large Numbers: Revisited

Why introducing Borel-Cantelli Lemma?
Answer: In order to prove the strong law.

Notably, for the strong law, the "limit" is inside the squared braces instead of outside the squared braces.

Theorem 6.1 If $X_1, X_2, \ldots$ are i.i.d. with bounded fourth central moment, and $E[X_n] = m$ (for some finite m), then the **strong law** holds.

Proof:

- $\lim_{n\to\infty} a_n = a$ (for some finite a) if, and only if,

$$(\forall\ \varepsilon > 0)(\exists\ N)(\forall\ n > N)|a_n - a| < \varepsilon.$$

ε is real and hence is not good for an induction proof. Fortunately, we can change the statement to:

- $\lim_{n\to\infty} a_n = a$ (for some finite a) if, and only if,

$$(\forall \text{ integer } j > 0)(\exists\ N)(\forall\ n > N)|a_n - a| < 1/j.$$

Strong Law of Large Numbers: Revisited

- So the set of $\left\{\underline{x} = (x_1, x_2, \ldots) \in \Re^\infty : \lim_{n\to\infty} \frac{1}{n}(x_1 + x_2 + \cdots + x_n) = m\right\}$ is equivalent to:

$$\left\{\underline{x} \in \Re^\infty : (\forall\ j > 0)(\exists\ N)(\forall\ n > N) \left|\frac{1}{n}(x_1 + x_2 + \cdots + x_n) - m\right| < \frac{1}{j}\right\}$$

$$= \bigcap_{j=1}^{\infty} \bigcup_{N=1}^{\infty} \bigcap_{n=N+1}^{\infty} \left\{\underline{x} \in \Re^\infty : \left|\frac{1}{n}(x_1 + x_2 + \cdots + x_n) - m\right| < \frac{1}{j}\right\}$$

$$= \bigcap_{j=1}^{\infty} \bigcup_{N=1}^{\infty} \bigcap_{n=N+1}^{\infty} A_n^c(1/j),$$

where $A_n(\varepsilon) \triangleq \left\{\underline{x} \in \Re^\infty : \left|\frac{1}{n}(x_1 + x_2 + \cdots + x_n) - m\right| \geq \varepsilon\right\}$.

- Accordingly,

$$\left\{\underline{x} \in \Re^\infty : \lim_{n\to\infty} \frac{1}{n}(x_1 + x_2 + \cdots + x_n) = m\right\}^c = \bigcup_{j=1}^{\infty} \bigcap_{N=1}^{\infty} \bigcup_{n=N+1}^{\infty} A_n(1/j) = \bigcup_{j=1}^{\infty} \limsup_{n\to\infty} A_n(1/j)$$

- Hence, by the first Borel-Cantelli lemma, if $\sum_{n=1}^{\infty} P\left(A_n(1/j)\right)$ converges, then $P\left(\limsup_{n\to\infty} A_n(1/j)\right) = 0$, which implies that

$$P\left(\left\{\underline{x} \in \Re^\infty : \lim_{n\to\infty} \frac{1}{n}(x_1 + x_2 + \cdots + x_n) = m\right\}^c\right) = P\left(\bigcup_{j=1}^{\infty} \limsup_{n\to\infty} A_n(1/j)\right) \le \sum_{j=1}^{\infty} P\left(\limsup_{n\to\infty} A_n(1/j)\right) = 0.$$

(This may not be true for sum of uncountably many zeros.)

Strong Law of Large Numbers: Revisited

In summary, any probability space that gives a bounded $\sum_{n=1}^{\infty} P\Big(A_n(\varepsilon)\Big)$ satisfies the strong law!

- **(Markov inequality)** $\Pr[\,|Z| \geq \alpha] \leq \dfrac{1}{\alpha^k} E[\,|Z|^k]$

  (Lyapounov's inequality) $E^{1/\alpha}[\,|Z|^\alpha] \leq E^{1/\beta}[\,|Z|^\beta]$ for $0 < \alpha \leq \beta$

$$
\begin{aligned}
P\left(A_n(\varepsilon)\right) &= \Pr\left[\,|(X_1 - m) + (X_2 - m) + \cdots + (X_n - m)| \geq n\varepsilon\,\right] \\
&\leq \frac{1}{n^4\varepsilon^4} E\left[((X_1 - m) + (X_2 - m) + \cdots + (X_n - m))^4\right] \quad \text{(by Markov's ineq.)} \\
&= \frac{1}{n^4\varepsilon^4}\left(\sum_{i=1}^{n} E[(X_i - m)^4] + \binom{4}{2}\sum_{i=1}^{n}\sum_{j=i+1}^{n} E[(X_i - m)^2]E[(X_j - m)^2]\right) \\
&\leq \frac{1}{n^4\varepsilon^4}\left(\sum_{i=1}^{n} 1 + \binom{4}{2}\sum_{i=1}^{n}\sum_{j=i+1}^{n} 1\right) E[(X - m)^4] \quad \text{(by Lyapounov's ineq.)} \\
&= \frac{(3n-2)}{n^3\varepsilon^4} E[(X - m)^4] \quad \text{so it's summable!}
\end{aligned}
$$

□

Strong Law of Large Numbers: Revisited

Corollary If $\sum_{n=1}^{\infty} \Pr\left[\left|\frac{1}{n}(X_1 + X_2 + \cdots + X_n) - m\right| \geq \varepsilon\right] < \infty$ for any $\varepsilon > 0$ arbitrarily small, then the strong law holds.

After we know how to validate the strong law, the question that naturally follows is how to validate the weak law.

Strong Law and Weak Law

Lemma If the strong law holds, then the weak law holds.

Proof:

$$
\begin{aligned}
&\Pr\left[\lim_{n\to\infty}\frac{1}{n}(X_1+X_2+\cdots+X_n)=m\right]\\
&= \Pr\left[(\forall\ \varepsilon>0)(\exists\ N)(\forall\ n>N)\left|\frac{1}{n}(X_1+X_2+\cdots+X_n)-m\right|<\varepsilon\right]\\
&= P\left(\bigcap_{\varepsilon>0}\bigcup_{N=1}^{\infty}\bigcap_{n=N}^{\infty}\left\{\underline{x}\in\Re^{\infty}:\left|\frac{1}{n}(x_1+x_2+\cdots+x_n)-m\right|<\varepsilon\right\}\right)\\
&\le P\left(\bigcup_{N=1}^{\infty}\bigcap_{n=N}^{\infty}\left\{\underline{x}\in\Re^{\infty}:\left|\frac{1}{n}(x_1+x_2+\cdots+x_n)-m\right|<\varepsilon\right\}\right)\\
&= P\left(\bigcup_{N=1}^{\infty}B_N\right),
\end{aligned}
$$

where $B_N=\bigcap_{n=N}^{\infty}\left\{\underline{x}\in\Re^{\infty}:\left|\frac{1}{n}(x_1+x_2+\cdots+x_n)-m\right|<\varepsilon\right\}$.

It can be easily seen that $B_1\subset B_2\subset\cdots\subset B_N\subset B_{N+1}\subset\cdots$

Hence, $B_N\uparrow\bigcup_{N=1}^{\infty}B_N$.

Strong Law and Weak Law

Let $C_1 = B_1$, and $C_N = B_N \setminus B_{N-1}$ for $N > 1$. Then $\{C_N\}_{N=1}^{\infty}$ are disjoint. We finally obtain:

$$
\begin{aligned}
P\left(\bigcup_{N=1}^{\infty} B_N\right) &= P\left(\bigcup_{N=1}^{\infty} C_N\right) \\
&= \sum_{N=1}^{\infty} P(C_N) \\
&= \lim_{N\to\infty} \sum_{n=1}^{N} P(C_n) \\
&= \lim_{N\to\infty} P(B_N) \\
&= \lim_{N\to\infty} P\left(\bigcap_{n=N}^{\infty} \left\{ \underline{x} \in \Re^\infty : \left| \frac{1}{n}(x_1 + x_2 + \cdots + x_n) - m \right| < \varepsilon \right\}\right) \\
&\le \lim_{N\to\infty} P\left(\left\{ \underline{x} \in \Re^\infty : \left| \frac{1}{N}(x_1 + x_2 + \cdots + x_N) - m \right| < \varepsilon \right\}\right) \\
&= \lim_{N\to\infty} P\left[\left| \frac{1}{N}(X_1 + X_2 + \cdots + X_N) - m \right| < \varepsilon \right].
\end{aligned}
$$

Strong Law and Weak Law

In summary, we prove that:

$$\Pr\left[\lim_{n\to\infty}\frac{1}{n}(X_1+X_2+\cdots+X_n)=m\right]\le\lim_{n\to\infty}P\left[\left|\frac{1}{n}(X_1+X_2+\cdots+X_n)-m\right|<\varepsilon\right].$$

So if the left-hand-side equals one, so does the right-hand-side.
This completes the proof that strong law implies weak law. □

Note: An alternative statement for this lemma is that *convergence with probability 1* implies *convergence in probability*.

Lemma If $X_1, X_2, \ldots$ are i.i.d. with bounded variance, and $E[X_n] = m$, then the weak law holds.

Proof: By Chebyshev's inequality,

$$P\left[\left|\frac{1}{n}(X_1 + X_2 + \cdots + X_n) - m\right| \geq \varepsilon\right] \leq \frac{\text{Var}[X_1]}{n\varepsilon^2} \to 0.$$

□

Question: Is the aforementioned condition also *necessary* under an i.i.d. assumption? (In other words, can we also say "For an i.i.d. random variables $X_1, X_2, \ldots$, if the weak law holds, then the variance is bounded.")

Answer: No. In fact, the bounded-variance condition can be further weakened.

Example 6.3: Generalization of Laws

Can we generalize the weak law (and the strong law) to situations that concern not necessarily "sample-sum-divided-by-sample-number?"
Here is an example.

- Let the *sample space*, Ω_n, be a set consisting of $n!$ permutations of numbers $1, 2, 3, \ldots, n$.
- Let 2^{Ω_n}, named the power set of Ω_n, be the *event space*. (An event is a subset of the sample space, whose probability can be evaluated. An event space contains all the probabilistically evaluable events.)

  Interpretation of Event: Given a probability space $(\{0,1\}, \{\emptyset, \{0,1\}\}, P)$, in which $\{0,1\}$ is the sample space, and $\{\emptyset, \{0,1\}\}$ is the event space. We cannot evaluate $P(\{0\})$ since $\{0\}$ is not an **event**.

- Let the probability measure $P_{\boldsymbol{\omega}_n}$ be equally probable over Ω_n.

Denote the random variable defined over the above probability space $(\Omega_n, 2^{\Omega_n}, P_{\boldsymbol{\omega}_n})$ by $\boldsymbol{\omega}_n$.

Example 6.3: Generalization of Laws

Equivalent representation of a permutation : Transform a permutation of $1, 2, \ldots, n$ to its equivalent "product-of-cycle" format. For example, permutation $\omega = (5, 1, 7, 4, 6, 2, 3)$ can be re-written as:

$$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 \\ 5 & 1 & 7 & 4 & 6 & 2 & 3 \end{pmatrix}$$

So we first observe that 1 maps to 5, which maps to 6, which maps to 2, which maps back to 1. We then form the first cycle, $(1, 5, 6, 2)$.

The next smallest number that does not appear in the previous cycle is 3, which maps to 7, which maps back to 3. So we form another cycle, $(3, 7)$.

The next smallest number that does not appear in the previous two cycles is the self-mapped 4, which gives us the last cycle (4).

Consequently, the equivalent "product-of-cycle" format of $\omega = (5, 1, 7, 4, 6, 2, 3)$ is $(1, 5, 6, 2)(3, 7)(4)$.

Example 6.3: Generalization of Laws

Number of cycles : Define $f_{n,k}(\omega) = 1$, if the k-position of the equivalent "product-of-cycle" format of $\omega \in \Omega_n$ completes a cycle; otherwise, $f_{n,k}(\omega) = 0$. In the previous example, only $f_{7,4} = f_{7,6} = f_{7,7} = 1$ (i.e., $\mathbf{2}, \mathbf{7}, \mathbf{4}$ in $(1, 5, 6, \mathbf{2})(3, \mathbf{7})(\mathbf{4})$), and $f_{7,1} = f_{7,2} = f_{7,3} = f_{7,5} = 0$.

Define a random variable $X_{n,k} = f_{n,k}(\boldsymbol{\omega}_n)$ for $\boldsymbol{\omega}_n$ defined over $(\Omega_n, 2^{\Omega_n}, P_{\boldsymbol{\omega}_n})$.

- Billingsley's book, as most porbabilitists do, just saves the effort to directly define $X_{n,k} = X_{n,k}(\omega)$.
- My lengthy introduction here is just to show to you that "*a random variable X is a real-valued function on sample space Ω, which maps from Ω to real line $\Re$, satisfying that $\{\omega \in \Omega : X(\omega) = x\}$ is an event for each real x,*" for which the definition can be found in any fundamental probability books.
- So, the function $X_{n,k}(\cdot)$ maps each permutation ω in Ω_n to a real number, either 0 or 1, as $f_{n,k}(\cdot)$ did. The set of all permutations that causes $X_{n,k}(\omega) = 1$, and the set of all permutations that yields $X_{n,k}(\omega) = 0$ are certainly subsets of Ω_n, and hence they are events in our power-set event space. Since events are probabilistically evaluable, we can now safely talk about $\Pr[X_{n,k} = 1] = P_{\boldsymbol{\omega}_n}\{\omega \in \Omega_n : f_{n,k}(\omega) = 1\}$ and $\Pr[X_{n,k} = 0]$.

Accordingly, the number of cycles equals $S_n = X_{n,1} + X_{n,2} + \cdots + X_{n,n}$.

Example 6.3: Generalization of Laws

Distributions of $\{X_{n,k}\}_{k=1}^{n}$: It can be shown that $\{X_{n,k}\}_{k=1}^{n}$ are independent, and $\Pr[X_{n,k} = 1] = \dfrac{1}{n-k+1}$ (cf. Example 5.6).

Mean of S_n :

$$\begin{aligned} E[S_n] &= \sum_{k=1}^{n} E[X_{n,k}] \\ &= \sum_{k=1}^{n} \Pr[X_{n,k} = 1] \\ &= \sum_{k=1}^{n} \frac{1}{n-k+1} \\ &= \sum_{k=1}^{n} \frac{1}{k}. \end{aligned}$$

Example 6.3: Generalization of Laws

Generalization of Weak Law :

$$
\begin{aligned}
&\Pr\left[\left|\frac{X_{n,1}+X_{n,2}+\cdots+X_{n,n}}{E[X_{n,1}]+E[X_{n,2}]+\cdots+E[X_{n,n}]}-1\right|\geq\varepsilon\right] \quad \text{(Is this a nice generalization?)}\\
&= \Pr\left[\left|\frac{X_{n,1}+X_{n,2}+\cdots+X_{n,n}}{E[S_n]}-1\right|\geq\varepsilon\right]\\
&= \Pr\left[\,|(X_{n,1}+X_{n,2}+\cdots+X_{n,n})-E[S_n]|\geq\varepsilon\cdot|E[S_n]|\,\right]\\
&\leq \frac{E\left[((X_{n,1}+X_{n,2}+\cdots+X_{n,n})-E[S_n])^2\right]}{\varepsilon^2E^2[S_n]} \quad \text{(by Markov's ineq.)}\\
&= \frac{E\left[((X_{n,1}-E[X_{n,1}])+(X_{n,2}-E[X_{n,2}])+\cdots+(X_{n,n}-E[X_{n,n}]))^2\right]}{\varepsilon^2E^2[S_n]}\\
&= \frac{\sum_{j=1}^n \mathrm{Var}[X_{n,j}]}{\varepsilon^2E^2[S_n]} \quad \text{(by independence)} \quad \leq \frac{\sum_{j=1}^n E[X_{n,j}^2]}{\varepsilon^2E^2[S_n]} \quad \text{(by variance} \leq \text{second moment)}\\
&= \frac{\sum_{j=1}^n E[X_{n,j}]}{\varepsilon^2E^2[S_n]} \quad \left(\text{by } E[X_{n,j}]=E[X_{n,j}^2]\right) \quad = \frac{E[S_n]}{\varepsilon^2E^2[S_n]}\\
&= \frac{1}{\varepsilon^2E[S_n]} = \frac{1}{\varepsilon^2\sum_{k=1}^n(1/k)} \leq \frac{1}{\varepsilon^2\log(n+1)}.
\end{aligned}
$$

Example 6.3: Generalization of Laws

Question : Does the previous derivation suffice to prove that

$$\Pr\left[\lim_{n\to\infty}\frac{S_n}{E[S_n]}=1\right]=1,$$

by the first Borel-Cantelli lemma? Answer by yourself.

Thinking : $\Pr\left[\left|\frac{S_n}{E[S_n]}-1\right|\geq\varepsilon\right]\leq\frac{1}{\varepsilon^2\log(n)}$ indicates that

$$(1-\varepsilon)E[S_n]\leq S_n\leq E[S_n](1+\varepsilon)\quad\text{with probability at least } 1-\frac{1}{\varepsilon^2\log(n)}.$$

Remember that all the permutations are equally probable.

So we can say most of the permutations (random interleavers) contain approximately $E[S_n]\approx\log(n)$ cycles!

Applications of weak-law argument (Chebyshev's ineq)

Weak-law argument (or Chebyshev's ineq) has many applications, such as Shannon's coding theory that is introduced in the course of *Information Theory*.

Here, we introduce two examples: Bernstein's Theorem, and a refinement of second Borel-Cantelli lemma.

Theorem 6.2 (Bernstein's Theorem) If function $f(\cdot)$ is continuous on $[0, 1]$, then

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

converges to $f(x)$ uniformly on $[0, 1]$.

Note 1: $B_n(x)$ is called the *Bernstein polynomial* of degree n associated with $f(\cdot)$.

Note 2: The difference between a continuous function on $[0, 1]$ and a continuous function on $(0, 1)$ is that the former also implies boundedness on $[0, 1]$.

Note 3: $g_n(x)$ is said to converge in n to $f(x)$ uniformly on domain $\mathcal{X}$ if

$$\lim_{n\to\infty} \sup_{x\in\mathcal{X}} |g_n(x) - f(x)| = 0.$$

Applications of weak-law argument (Chebyshev's ineq)

Note 4: Bernstein's result goes further than the Weierstrass approximation theorem does (which states that every compact-support continuous function can be uniformly approximated by polynomials) by specifically specifying (one of) the approximation polynomials.

Proof:

- Let $M = \sup_{x\in[0,1]} |f(x)|$, and $\delta(\varepsilon) = \sup_{\{(x,y)\in[0,1]^2:|x-y|<\varepsilon\}} |f(x) - f(y)|$.
- By the continuity of $f(\cdot)$ over a closed and bounded (hence, compact) set $[0, 1]$,

$$\lim_{\varepsilon\downarrow 0} \delta(\varepsilon) = 0.$$

- Let $X_1, X_2, \ldots, X_n$ be independent random variables with

$$\Pr[X_i = 1] = x \quad \text{and} \quad \Pr[X_i = 0] = 1 - x.$$

Denote $S_n = X_1 + X_2 + \cdots + X_n$. Then

$$E\left[f\left(\frac{S_n}{n}\right)\right] = \sum_{k=0}^{n} \binom{n}{k} x^k (1-x)^{n-k} f\left(\frac{k}{n}\right) = B_n(x).$$

Applications of weak-law argument (Chebyshev's ineq)

- Denoting $\mathbb{N}_n = \{0, 1, 2, \ldots, n\}$, we obtain:

$$\begin{aligned}
|B_n(x) - f(x)| &= \left| E\left[f\left(\frac{S_n}{n}\right)\right] - f(x)\right| \\
&= \left| E\left[f\left(\frac{S_n}{n}\right) - f(x)\right]\right| \\
&\leq E\left[\left| f\left(\frac{S_n}{n}\right) - f(x)\right|\right] \\
&= \sum_{s_n \in \mathbb{N}_n} \left| f\left(\frac{s_n}{n}\right) - f(x)\right| \Pr[S_n = s_n] \\
&= \sum_{\{s_n \in \mathbb{N}_n : |s_n/n - x| < \varepsilon\}} \left| f\left(\frac{s_n}{n}\right) - f(x)\right| \Pr[S_n = s_n] \\
&\quad + \sum_{\{s_n \in \mathbb{N}_n : |s_n/n - x| \geq \varepsilon\}} \left| f\left(\frac{s_n}{n}\right) - f(x)\right| \Pr[S_n = s_n]
\end{aligned}$$

Applications of weak-law argument (Chebyshev's ineq)

$$
\begin{aligned}
&\leq \sum_{\{s_n \in \mathbb{N}_n : |s_n/n - x| < \varepsilon\}} \delta(\varepsilon) \Pr[S_n = s_n] + \sum_{\{s_n \in \mathbb{N}_n : |s_n/n - x| \geq \varepsilon\}} (2M) \Pr[S_n = s_n] \\
&= \delta(\varepsilon) \Pr\left[\left|\frac{S_n}{n} - x\right| < \varepsilon\right] + (2M) \Pr\left[\left|\frac{S_n}{n} - x\right| \geq \varepsilon\right] \\
&\leq \delta(\varepsilon) + (2M)\frac{\operatorname{Var}[X_1]}{n\varepsilon^2} \quad \text{(by Chebyshev's ineq.)} \\
&= \delta(\varepsilon) + (2M)\frac{x(1-x)}{n\varepsilon^2} \\
&\leq \delta(\varepsilon) + (2M)\frac{1}{4n\varepsilon^2} \\
&= \delta(\varepsilon) + \frac{M}{2n\varepsilon^2},
\end{aligned}
$$

for which the upper bound is independent of x (and hence, is a uniform bound). $\square$

Refinement of Second Bore-Cantelli Lemma

We may replace the "independence assumption" in the second Borel-Cantelli lemma by another condition as follows.

This can also be smartly proved by weak-law (or Chebyshev's ineq) argument.

Theorem 4.4 (Second Borel-Cantelli Lemma) If $\{A_n\}_{n=1}^{\infty}$ forms an independent sequence of events for a probability measure P, and $\sum_{n=1}^{\infty} P(A_n)$ diverges, then $P\left(\limsup_{n\to\infty} A_n\right) = 1$.

Theorem 6.3 (Refinement of Second Borel-Cantelli Lemma) For a probability measure P, if $\{A_n\}_{n=1}^{\infty}$ satisfies

$$\liminf_{n\to\infty} \frac{\sum_{j=1}^{n}\sum_{k=1}^{n} P(A_j \cap A_k)}{\left(\sum_{k=1}^{n} P(A_k)\right)^2} \le 1,$$

and $\sum_{n=1}^{\infty} P(A_n)$ diverges, then $P\left(\limsup_{n\to\infty} A_n\right) = 1$.

Refinement of Second Bore-Cantelli Lemma

Proof: Let Z be the random variable having probability measure P.

$$X_n(Z) = \begin{cases} 1, & \text{if } Z \in A_n; \\ 0, & \text{if } Z \notin A_n. \end{cases}$$

Also, let $S_n = X_1 + X_2 + \cdots + X_n$, and denote $m_n = E[S_n] = \sum_{k=1}^{n} P(A_k)$, which approaches infinity as n goes to infinity.

For convenience, denote

$$\theta_n = \frac{\sum_{j=1}^{n} \sum_{k=1}^{n} P(A_j \cap A_k)}{\left(\sum_{k=1}^{n} P(A_k)\right)^2}.$$

Observe that $P\left(\limsup_{n\to\infty} A_n\right) = P(A_n \text{ i.o.}) = \Pr\left[\sup_{k>1} S_k = \infty\right]$. Hence, it suffices to show that $\Pr\left[\sup_{k>1} S_k < \infty\right] = 0$.

m_n nondecreasing, $m_n \uparrow \infty$,
$m_n \le n$, $\frac{m_n^2}{(m_n-s)^2} \to 1$

For $s < m_n$, derive

$$\begin{aligned} \Pr[S_n \le s] &\le \Pr[S_n \le s \text{ or } S_n \ge 2m_n - s] \\ &= \Pr[S_n - m_n \le -(m_n - s) \text{ or } S_n - m_n \ge m_n - s] \\ &= \Pr\left[\,|S_n - m_n| \ge m_n - s\right] \\ &\le \frac{E[|S_n - m_n|^2]}{(m_n - s)^2} \quad \text{(By Markov ineq.)} \quad = \frac{\text{Var}[S_n]}{(m_n - s)^2}, \end{aligned}$$

and

$$\begin{aligned}
\text{Var}[S_n] &= E[S_n^2] - E^2[S_n] \\
&= \sum_{j=1}^n \sum_{k=1}^n E[X_j X_k] - \left(\sum_{k=1}^n E[X_k] \right)^2 \\
&= \sum_{j=1}^n \sum_{k=1}^n P(A_j \cap A_k) - \left(\sum_{k=1}^n P(A_k) \right)^2 \\
&= \left(\sum_{k=1}^n P(A_k) \right)^2 (\theta_n - 1) = m_n^2(\theta_n - 1),
\end{aligned}$$

which implies $\theta_n \geq 1$. We therefore obtain that for $s < m_n$,

$$\Pr[S_n \leq s] \leq (\theta_n - 1)\frac{m_n^2}{(m_n - s)^2}.$$

The above inequality implies that if $\liminf_{n\to\infty} \theta_n \leq 1$ (or equivalently, $\liminf_{n\to\infty}(\theta_n - 1) = 0$), then

$$\begin{aligned}
\liminf_{n\to\infty} \Pr[S_n \leq s] &\leq \liminf_{n\to\infty} \left[(\theta_n - 1)\frac{m_n^2}{(m_n - s)^2} \right] \\
&\leq \left(\liminf_{n\to\infty}(\theta_n - 1) \right) \left(\limsup_{n\to\infty} \frac{m_n^2}{(m_n - s)^2} \right) = 0.
\end{aligned}$$

Refinement of Second Bore-Cantelli Lemma

Consequently, as

$$\Pr\left[\sup_{k\geq 1} S_k < s\right] \;\leq\; \Pr\left[S_n < s\right] \text{ for any } n,$$

we conclude:

$$\Pr\left[\sup_{k\geq 1} S_k < s\right] \;\leq\; \liminf_{n\to\infty} \Pr\left[S_n < s\right] = 0,$$

which immediately gives:

$$\begin{aligned}
\Pr\left[\sup_{k\geq 1} S_k < \infty\right] &= \Pr\left[\sup_{k\geq 1} S_k < 1 \text{ or } \sup_{k\geq 1} S_k < 2 \text{ or } \sup_{k\geq 1} S_k < 3 \text{ or } \cdots\right] \\
&\leq \sum_{s=1}^{\infty} \Pr\left[\sup_{k\geq 1} S_k < s\right] \\
&= 0.
\end{aligned}$$

□

Refinement of Second Bore-Cantelli Lemma

Theorem 4.4′ (Refinement of the Second Borel-Cantelli Lemma) If $\{A_n\}_{n=1}^{\infty}$ forms a **pair-wise** independent sequence of events for a probability measure P, and $\sum_{n=1}^{\infty} P(A_n)$ diverges, then $P\left(\limsup_{n\to\infty} A_n\right) = 1$.

Proof: By $P(A_j \cap A_k) = P(A_j)P(A_k)$ and Theorem 6.3. □