Section 22

Sums of Independent Random Variables

Po-Ning Chen, Professor
Institute of Communications Engineering
National Chiao Tung University
Hsin Chu, Taiwan 30010, R.O.C.
Law of large numbers revisited

**Theorem 22.1 (advanced version of strong law of large numbers)** If \( X_1, X_2, \ldots \) are pair-wise independent with common marginal distribution and finite mean, then

\[
\frac{S_n}{n} \to E[X_1] \quad \text{with probability 1,}
\]

where \( S_n = X_1 + X_2 + \cdots + X_n \).

**Proof** (due to Etemadi): Assume without loss of generality that \( X_i \) is non-negative.

If the theorem holds for non-negative random variables, then

\[
\frac{S_n}{n} = \frac{1}{n} \sum_{k=1}^{n} X_k^+ - \frac{1}{n} \sum_{k=1}^{n} X_k^- \overset{w.p.}{\longrightarrow} E[X_1^+] - E[X_1^-] = E[X_1].
\]

- Consider the truncated random variable \( Y_k = X_k I_{[X_k \leq k]} \), and denote \( S_n^* = \sum_{k=1}^{n} Y_k \). (Notably, \( Y_1, Y_2, \ldots \) is not identically distributed, but only pair-wise independent.)

Then for \( k \leq n \),

\[
E[Y_k^2] = E[X_k^2 I_{[X_k \leq k]}] = E[X_1^2 I_{[X_1 \leq k]}] \leq E[X_1^2 I_{[X_1 \leq n]}] = E[Y_n^2].
\]
Law of large numbers revisited

The reason of introducing a truncated version of $X_n$ is because $E[X_n^2]$ may be infinity!
This is the key technique used in this proof.

- **Claim**: For $\alpha > 1$ fixed, and $u_n = [\alpha^n],$

$$\sum_{n=1}^{\infty} \Pr \left[ \left| \frac{S^{*}_{u_n} - E[S^{*}_{u_n}]}{u_n} \right| > \varepsilon \right] < \infty$$
for any $\varepsilon > 0.$

**Theorem 4.3 (First Borel-Cantelli lemma)**

$$\sum_{n=1}^{\infty} P(A_n) < \infty \Rightarrow P \left( \limsup_{n \to \infty} A_n \right) = P(A_n \text{ i.o.}) = 0.$$

**Proof of the claim**: By Chebyshev’s inequality,

$$\sum_{n=1}^{\infty} \Pr \left[ \left| \frac{S^{*}_{u_n} - E[S^{*}_{u_n}]}{u_n} \right| > \varepsilon \right] \leq \sum_{n=1}^{\infty} \frac{\text{Var}[S^{*}_{u_n}]}{u_n^2 \varepsilon^2} \leq \frac{1}{\varepsilon^2} \sum_{n=1}^{\infty} \frac{E[Y^2_{u_n}]}{u_n},$$

where by pair-wise independence,

$$\text{Var}[S^{*}_{u_n}] = \sum_{k=1}^{u_n} \text{Var}[Y_k] \leq u_n E[Y^2_{u_n}].$$
Law of large numbers revisited

Hence,

\[
\sum_{n=1}^{\infty} \Pr \left[ \left| \frac{S_{u_n}^* - E[S_{u_n}^*]}{u_n} \right| > \varepsilon \right] \leq \frac{1}{\varepsilon^2} \sum_{n=1}^{\infty} \frac{E[Y_{u_n}^2]}{u_n} = \frac{1}{\varepsilon^2} \sum_{n=1}^{\infty} \frac{E[X_{u_n}^2 I[X_{u_n} \leq u_n]]}{u_n}
\]

\[
= \frac{1}{\varepsilon^2} \sum_{n=1}^{\infty} \frac{E[X_1^2 I[X_1 \leq u_n]]}{u_n} = \frac{1}{\varepsilon^2} \lim_{m \to \infty} \sum_{n=1}^{m} \frac{E[X_1^2 I[X_1 \leq u_n]]}{u_n}
\]

\[
= \frac{1}{\varepsilon^2} \lim_{m \to \infty} E \left[ X_1^2 \sum_{n=1}^{m} \frac{1}{u_n} I[X_1 \leq u_n] \right]
\]

\[
= \frac{1}{\varepsilon^2} \lim_{m \to \infty} \sum_{n=1}^{m} \frac{1}{u_n} I[X_1 \leq u_n]
\]

\[
= \frac{1}{\varepsilon^2} \sum_{n=1}^{\infty} \frac{1}{u_n} I[X_1 \leq u_n]
\]

**Monotone convergence theorem**: If for every positive integer \( m \) and every \( x \) in the support \( \mathcal{X} \) of random variable \( X \), \( 0 \leq f_m(x) \leq f_{m+1}(x) \), then

\[
\lim_{m \to \infty} E[f_m(X)] = \lim_{m \to \infty} \int_{\mathcal{X}} f_m(x) dP_X(x) = \int_{\mathcal{X}} \lim_{m \to \infty} f_m(x) dP_X(x) = E \left[ \lim_{m \to \infty} f_m(X) \right].
\]
Law of large numbers revisited

Observe that for any $x > 0$ fixed,

\[
\sum_{n=1}^{\infty} \frac{1}{u_n} I_{[x \leq u_n]} = \sum_{\{n \in \mathbb{N} : u_n \geq x\}} \frac{1}{u_n}
\]

\[
= \sum_{n \geq N} \frac{1}{u_n}, \text{ where } N = \min\{n \in \mathbb{N} : u_n \geq x\}
\]

\[
\leq \sum_{n \geq N} \frac{2}{\alpha^n}, \text{ (since } u_n = \lfloor \alpha^n \rfloor \text{ and } \lfloor y \rfloor \geq \frac{1}{2}y \text{ for } y \geq 1)\n\]

\[
= \left( \frac{2}{1 - \alpha^{-1}} \right) \frac{1}{\alpha^N}
\]

\[
\leq \left( \frac{2\alpha}{\alpha - 1} \right) \frac{1}{x}. \text{ (by } \alpha^N \geq \lfloor \alpha^N \rfloor = u_N \geq x)\n\]

This concludes that:

\[
\sum_{n=1}^{\infty} \text{Pr} \left[ \left| \frac{S_{u_n}^* - E[S_{u_n}^*]}{u_n} \right| > \varepsilon \right] \leq \frac{1}{\varepsilon^2} E \left[ X_1^2 \sum_{n=1}^{\infty} \frac{1}{u_n} I_{[X_1 \leq u_n]} \right] \leq \frac{1}{\varepsilon^2} \left( \frac{2\alpha}{\alpha - 1} \right) E[X_1] < \infty.
\]
Law of large numbers revisited

- By the above claim and the first Borel-Cantelli lemma,
  \[
  \frac{S_{u_n}^* - E[S_{u_n}^*]}{u_n} \to 0 \text{ with probability } 1.
  \]

- By the Cesáro-mean theorem (cf. the next slide),
  \[
  \frac{1}{u_n} E[S_{u_n}^*] = \frac{1}{u_n} \sum_{k=1}^{u_n} E[Y_{u_n}]
  \]

should have the same limit as
\[
\lim_{n \to \infty} E[Y_{u_n}] = \lim_{n \to \infty} E[X_1 I[X_1 \leq u_n]] = E[X_1].
\]

Thus,
\[
\frac{S_{u_n}^*}{u_n} \to E[X_1] \text{ with probability } 1.
\]
Law of large numbers revisited

**Theorem (Cesáro-mean theorem)** If \( a_n \to a \) and \( b_n = (1/n) \sum_{i=1}^{n} a_i \), then \( b_n \to a \) as \( n \to \infty \).

**Proof:** \( a_n \to a \) implies that for any \( \varepsilon > 0 \), there exists \( N \) such that for all \( n > N \), \( |a_n - a| < \varepsilon \). Then

\[
|b_n - a| = \left| \frac{1}{n} \sum_{i=1}^{n} (a_i - a) \right|
\]

\[
\leq \frac{1}{n} \sum_{i=1}^{n} |a_i - a|
\]

\[
= \frac{1}{n} \sum_{i=1}^{N} |a_i - a| + \frac{1}{n} \sum_{i=N+1}^{n} |a_i - a|
\]

\[
\leq \frac{1}{n} \sum_{i=1}^{N} |a_i - a| + \frac{n - N}{n} \varepsilon.
\]

Hence, \( \lim_{n \to \infty} |b_n - a| \leq \varepsilon \). Since \( \varepsilon \) can be made arbitrarily small, the lemma holds. \( \square \)
Law of large numbers revisited

- Claim: \( \frac{S_n - S_n^*}{n} \to 0 \) with probability 1.

Proof of the claim:

\[
\sum_{n=1}^{\infty} \Pr[X_n \neq Y_n] = \sum_{n=1}^{\infty} \Pr[X_n \neq X_n I_{X_n \leq n}]
\]

\[
= \sum_{n=1}^{\infty} \Pr[X_n > n]
\]

\[
= \sum_{n=1}^{\infty} \Pr[X_1 > n] \quad \text{(by “identical distributed” assumption)}
\]

\[
\leq \int_0^{\infty} \Pr[X_1 > t]dt
\]

\[
= E[X_1] \quad \text{(by non-negativity assumption of } X_1)\]

\[
< \infty.
\]

Hence, the first Bore-Cantelli lemma gives that

\[
\Pr[(X_n \neq Y_n) \text{ is true infinitely often in } n] = 0,
\]
equivalently,

\[
\Pr[(X_n \neq Y_n) \text{ is true finitely many in } n] = 1.
\]
Law of large numbers revisited

This implies that
\[
\Pr\left[ (\exists N_n = \{n_1, n_2, \ldots, n_M\}) \ X_n \neq Y_n \text{ only for } n \in N_n \right] = 1.
\]

The above result, together with the fact that
\[
\Pr[(X_n - Y_n) < \infty] = \Pr\left[ X_n I_{[X_n > n]} < \infty \right] = \Pr\left[ X_1 I_{[X_1 > n]} < \infty \right] = 1
\]
because \( E[X_1] < \infty \), leads to:
\[
\Pr\left[ \lim_{n \to \infty} \frac{(X_1 - Y_1) + \cdots + (X_n - Y_n)}{n} = 0 \right] = 1.
\]

Now we have

- \( S_{u_n}^*/u_n \to E[X_1] \) with probability 1, where \( u_n = \lfloor \alpha^n \rfloor \) for some \( \alpha > 1 \) fixed, and
- \( (S_n - S_n^*)/n \to 0 \) with probability 1.

The above two results directly imply \( S_{u_n}/u_n \to E[X_1] \) (as \( n \) goes to infinity) with probability 1.

It remains to show \( S_k/k \to E[X_1] \) (as \( k \) goes to infinity) with probability 1.
Law of large numbers revisited

- For \( u_n \leq k < u_{n+1} \),

\[
\begin{align*}
\frac{u_n S_{u_n}}{u_{n+1} u_n} &= \frac{S_{u_n}}{u_{n+1}} \\
&= \frac{X_1 + \cdots + X_{u_n}}{u_{n+1}} \\
&\leq \frac{X_1 + \cdots + X_{u_n}}{k} \\
&\leq \frac{X_1 + \cdots + X_{u_n} + \cdots + X_k}{k} = \frac{S_k}{k} \\
&\leq \frac{X_1 + \cdots + X_{u_n} + \cdots + X_k + \cdots + X_{u_{n+1}}}{u_n} \\
&= \frac{S_{u_{n+1}}}{u_n} \\
&= \frac{u_{n+1} S_{u_{n+1}}}{u_n u_{n+1}}
\end{align*}
\]

since \( X_n \) is assumed non-negative.
Law of large numbers revisited

Because
\[
\frac{u_n}{u_{n+1}} \frac{S_{u_n}}{u_n} \to \frac{1}{\alpha} E[X_1] \text{ with probability 1,}
\]
and
\[
\frac{u_{n+1}}{u_n} \frac{S_{u_{n+1}}}{u_{n+1}} \to \alpha E[X_1] \text{ with probability 1,}
\]
we obtain:
\[
\frac{1}{\alpha} E[X_1] \leq \lim \inf_{k \to \infty} \frac{S_k}{k} \leq \frac{1}{\alpha} E[X_1] \leq \lim \sup_{k \to \infty} \frac{S_k}{k} \leq \alpha E[X_1] \text{ with probability 1.}
\]

As the above statement is valid for any \( \alpha > 1 \), we conclude that
\[
\frac{S_k}{k} \to E[X_1] \text{ with probability 1.}
\]
Law of large numbers revisited

**Theorem** If $X_1, X_2, \ldots$ are pair-wise independent with common marginal distribution whose mean exists (could be infinity as defined in slide 21-1), then

$$\frac{1}{n} \sum_{k=1}^{n} X_k \rightarrow E[X_1] \text{ with probability 1.}$$

**Proof:** Now, based on the previous theorem, we only need to prove the current theorem for the case of $E[X_1] = \infty$.

- Suppose without loss of generality that $E[X_1^-] < \infty$ and $E[X_1^+] = \infty$. Then

  $$\frac{1}{n} \sum_{k=1}^{n} X_k^- \rightarrow E[X_1^-] \text{ with probability 1.}$$

- Let $Y_n(u) = X_n^+ I_{[X_n \leq u]}$, and observe that

  $$\frac{1}{n} \sum_{k=1}^{n} X_k^+ \geq \frac{1}{n} \sum_{k=1}^{n} Y_k(u), \text{ and } \frac{1}{n} \sum_{k=1}^{n} Y_k(u) \rightarrow E[Y_k(u)] \text{ with probability 1.}$$

  Hence,

  $$\frac{1}{n} \sum_{k=1}^{n} X_k^+ \geq E[Y_k(u)] \text{ (as } n \text{ goes to infinity) with probability 1.}$$
Law of large numbers revisited

- Since the above statement is valid for any $u$, and $E[Y_k(u)] \to \infty$ as $u \to \infty$,

$$\frac{1}{n} \sum_{k=1}^{n} X_k^+ \to \infty \text{ with probability 1.}$$

- Finally,

$$\frac{1}{n} \sum_{k=1}^{n} X_k = \frac{1}{n} \sum_{k=1}^{n} X_k^+ - \frac{1}{n} \sum_{k=1}^{n} X_k^- \to \infty \text{ with probability 1.}$$
Limit of normalized Poisson

Next, we introduce a famous result for Possion distribution, whose validity can be proved by weak-law or Chebyshev’s-inequality argument.

**Lemma (degeneration of normalized Poisson)** Let $Y_\lambda$ be a Poisson random variable with parameter $\lambda$, and let $G_\lambda(\cdot)$ be the cdf of a $Y_\lambda/\lambda$. Then

$$
\lim_{\lambda \to \infty} G_\lambda(t) = \begin{cases} 
1, & \text{if } t > 1; \\
0, & \text{if } t < 1.
\end{cases}
$$

**Proof:** By Chebyshev’s inequality,

$$
\Pr \left[ \frac{|Y_\lambda - \lambda|}{\lambda} \geq \varepsilon \right] = \Pr \left[ |Y_\lambda - \lambda| \geq \varepsilon \lambda \right] \leq \frac{\text{Var}[Y_\lambda]}{\lambda^2 \varepsilon^2} = \frac{\lambda}{\lambda^2 \varepsilon^2} = \frac{1}{\lambda \varepsilon^2} \to 0
$$
as $\lambda \to \infty$. \qed
Limit of normalized Poisson

Let $X$ be a non-negative random variable.

Derive the one-sided Laplace transform of the distribution of $X$ as:

$$ M_X(s)_+ = \int_{0}^{\infty} e^{-sx} dF_X(x) \text{ for } s \geq 0. $$

Notably, $M_X(s)_+ = \int_{0}^{\infty} e^{-sx} dF_X(x) \leq \int_{0}^{\infty} dF_X(x) = 1$ is finite for all $s \geq 0$, but may be infinity for $s < 0$.

Here, we are only interested in those $s$ with $s \geq 0$; hence, it is named the one-sided Laplace transform.

In addition, $M_X(s)_+ = M_X(-s)$, where $M_X(\cdot)$ is the moment generating function of $X$. 
Application of degeneration of $G_\lambda(t)$ as $\lambda \to \infty$

**Proposition** Fix a non-negative random variable $X$. For $y > 0$,

$$\Pr[X \leq y] = \lim_{s \to \infty} \sum_{k=0}^{\lfloor sy \rfloor} \frac{(-1)^k}{k!} s^k M_X^{(k)}(s)_+.$$

**Proof:** For $s > 0$,

$$M_X^{(k)}(s)_+ = (-1)^k \int_0^\infty x^k e^{-sx} dF_X(x).$$
Application of degeneration of $G_\lambda(t)$ as $\lambda \to \infty$

Hence, for $s > 0$,

$$\sum_{k=0}^{\lfloor sy \rfloor} \frac{(-1)^k}{k!} s^k M_X^{(k)}(s)_+ = \sum_{k=0}^{\lfloor sy \rfloor} \frac{(-1)^k}{k!} s^k \left( (-1)^k \int_0^\infty x^k e^{-sx} dF_X(x) \right)$$

$$= \int_0^\infty \sum_{k=0}^{\lfloor sy \rfloor} e^{-sx} \frac{(sx)^k}{k!} dF_X(x)$$

$$= \int_0^\infty \Pr [Y_{sx} \leq \lfloor sy \rfloor] dF_X(x)$$

$$= \int_0^\infty \Pr [Y_{sx} \leq sy] dF_X(x)$$

$$= \int_0^\infty \Pr \left[ \frac{Y_{sx}}{sx} \leq \frac{y}{x} \right] dF_X(x)$$

$$= \int_0^\infty G_{sx} \left( \frac{y}{x} \right) dF_X(x).$$
Application of degeneration of $G_\lambda(t)$ as $\lambda \to \infty$

As a result,

$$\lim_{s \to \infty} \sum_{k=0}^{\lfloor sy \rfloor} \frac{(-1)^k}{k!} s^k M^{(k)}_X(s)_+ = \lim_{s \to \infty} \int_0^\infty G_{sx} \left( \frac{y}{x} \right) dF_X(x) = \int_0^\infty \lim_{s \to \infty} G_{sx} \left( \frac{y}{x} \right) dF_X(x),$$

since by dominated convergence theorem, $f_n(x) = G_{nx}(y/x) \leq 1 = g(x)$ for every $n$, and $\int_0^\infty g(x)dF_X(x) = 1 < \infty$.

Give a sequence of non-negative $\mu$-measurable function $f_n$ with $\lim_{n \to \infty} f_n(x) = f(x)$ for all $x \in \mathcal{X}$, except on a subset of $\mathcal{X}$ with $\mu$-measure zero.

**Lemma (Fatou’s lemma)**

$$\int_{\mathcal{X}} \left[ \lim_{n \to \infty} f_n(x) \right] \mu(dx) \leq \liminf_{n \to \infty} \int_{\mathcal{X}} f_n(x) \mu(dx).$$

Fatou’s lemma indicates that in general, we cannot interchange the order of integration and limit operation.

**Theorem (Lebesgue convergence theorem or dominated convergence theorem)**

If, in addition to non-negativity, $f_n(x) \leq g(x)$ for all $x \in \mathcal{X}$, except on a subset of $\mathcal{X}$ with $\mu$-measure zero, and $g(\cdot)$ is $\mu$-integrable in $\mathcal{X}$ (namely, $\int_{\mathcal{X}} g(x) \mu(dx) < \infty$), then

$$\int_{\mathcal{X}} \left[ \lim_{n \to \infty} f_n(x) \right] \mu(dx) = \lim_{n \to \infty} \int_{\mathcal{X}} f_n(x) \mu(dx).$$
Application of degeneration of $G_{\lambda}(t)$ as $\lambda \to \infty$

Consequently, (for $y$ that has no point mass),

$$\lim_{s \to \infty} \sum_{k=0}^{[sy]} \frac{(-1)^k}{k!} s^k M_X^{(k)}(s) = \int_0^{\infty} \lim_{s \to \infty} G_{sx} \left( \frac{y}{x} \right) dF_X(x)$$

$$= \int_0^y dF_X(x)$$

$$= \Pr[X \leq y].$$

(How to determine $\Pr[X \leq y]$ when $X$ has point mass at $y$?)

**Corollary** The distribution of a non-negative random variable is uniquely determined by its moment generating function $M_X(s)$ at $s < 0$.

**Proof:** For $y > 0$,

$$\Pr[X \leq y] = \lim_{s \to \infty} \sum_{k=0}^{[sy]} \frac{(-1)^k}{k!} s^k \frac{\partial^k M_X(-s)}{\partial s^k}.$$  

Determining $\Pr[X = 0]$ by the right-continuity of cdf gives the desired result.

**Final comment:** In fact, to determined the cdf of a non-negative random variable $X$, we only need to know $M_X(s)$ for $s < -s_0$ for any $s_0 > 0$. 
Maximal inequalities

The maximal inequalities concern the maxima of partial sums.

**Theorem 22.4 (due to Kolmogorov)** Suppose that $X_1, X_2, \ldots$ are independent with zero mean and finite variances (not necessarily identically distributed). Then for $\alpha > 0$,

$$
\Pr \left[ \max_{1 \leq k \leq n} |S_k| \geq \alpha \right] \leq \frac{1}{\alpha^2} \text{Var}[S_n],
$$

where $S_n = X_1 + \cdots + X_n$.

Chebyshev’s inequality said that

$$
\Pr[|S_n| \geq \alpha] \leq \alpha^{-2} \text{Var}[S_n].
$$

This theorem strengthens the result that $\alpha^{-2} \text{Var}[S_n]$ not only bounds $\Pr[|S_n| \geq \alpha]$, but also bounds $\Pr \left[ \max_{1 \leq k \leq n} |S_k| \geq \alpha \right]$.

**Proof:** Define the event

$$
A_k = [|S_1| < \alpha \land |S_2| < \alpha \land \cdots \land |S_{k-1}| < \alpha \land |S_k| \geq \alpha].
$$
Maximal inequalities

Since there is exactly one of \( \{A_k\}_{k=1}^\infty \) is true,

\[
E[S_n^2] = E \left[ S_n^2 \left( I_{A_1} + I_{A_2} + \cdots + I_{A_n} + I_{A_{n+1}} + \cdots \right) \right] \\
\geq E \left[ S_n^2 (I_{A_1} + I_{A_2} + \cdots + I_{A_n}) \right] \\
= \sum_{k=1}^n E \left[ S_n^2 I_{A_k} \right] \\
= \sum_{k=1}^n E \left[ (S_k^2 + 2S_k(S_n - S_k) + (S_n - S_k)^2) I_{A_k} \right] \\
\geq \sum_{k=1}^n E \left[ (S_k^2 + 2S_k(S_n - S_k)) I_{A_k} \right] \\
= \sum_{k=1}^n E \left[ S_k^2 I_{A_k} + 2S_k I_{A_k} (S_n - S_k) \right] \\
= \sum_{k=1}^n \left( E \left[ S_k^2 I_{A_k} \right] + 2E \left[ S_k I_{A_k} (S_n - S_k) \right] \right) \\
= \sum_{k=1}^n \left( E \left[ S_k^2 I_{A_k} \right] + 2E \left[ S_k I_{A_k} \right] E \left[ S_n - S_k \right] \right),
\]

where the last step follows from the independence between \( S_k I_{A_k} \) and \( S_n - S_k \).
Maximal inequalities

Continue the previous derivation:

\[
E[S_n^2] \geq \sum_{k=1}^{n} \left( E[S_k^2 I_{A_k}] + 2E[S_k I_{A_k}] E[S_n - S_k] \right)
\]

\[
= \sum_{k=1}^{n} E[S_k^2 I_{A_k}] \quad \text{(by the zero mean assumption, } E[S_n - S_k] = 0 \text{)}
\]

\[
\geq \sum_{k=1}^{n} E[\alpha^2 I_{A_k}] \quad (I_{A_k} = 1 \text{ only when } |S_k| \geq \alpha)
\]

\[
= \alpha^2 \sum_{k=1}^{n} \Pr[A_k]
\]

\[
= \alpha^2 \Pr\left[ \max_{1 \leq k \leq n} |S_k| \geq \alpha \right].
\]
Maximal inequalities

The previous theorem provide a bound for the cdf of \( \max_{1 \leq k \leq n} |S_k| \) using the second moment.

We can also bound the cdf of \( \max_{1 \leq k \leq n} |S_k| \) by the cdf of \( |S_k| \) for \( 1 \leq k \leq n \).

**Theorem 22.5 (due to Etemadi)** Suppose that \( X_1, X_2, \ldots \) are independent. For \( \alpha \geq 0 \),

\[
\Pr \left[ \max_{1 \leq k \leq n} |S_k| \geq 3\alpha \right] \leq 3 \max_{1 \leq k \leq n} \Pr[|S_k| \geq \alpha].
\]

**Proof:** Define the event

\[
A_k = [ |S_1| < 3\alpha \land |S_2| < 3\alpha \land \cdots \land |S_{k-1}| < 3\alpha \land |S_k| \geq 3\alpha ].
\]

Then

\[
\Pr \left[ \max_{1 \leq k \leq n} |S_k| \geq 3\alpha \right] = \Pr \left[ \left( \max_{1 \leq k \leq n} |S_k| \geq 3\alpha \right) \land (|S_n| \geq \alpha) \right]
= \Pr \left[ \left( \max_{1 \leq k \leq n} |S_k| \geq 3\alpha \right) \land (|S_n| < \alpha) \right]
\leq \Pr [|S_n| \geq \alpha] + \Pr \left[ \left( \max_{1 \leq k \leq n} |S_k| \geq 3\alpha \right) \land (|S_n| < \alpha) \right].
\]
Maximal inequalities

(Continue from the previous slide)

\[
\Pr \left[ \max_{1 \leq k \leq n} |S_k| \geq 3\alpha \right] \leq \Pr [|S_n| \geq \alpha] + \Pr \left[ \left( \max_{1 \leq k \leq n} |S_k| \geq 3\alpha \right) \land (|S_n| < \alpha) \right]
\]

\[
= \Pr [|S_n| \geq \alpha] + \Pr \left[ (A_1 \lor A_2 \lor \cdots \lor A_n) \land (|S_n| < \alpha) \right]
\]

\[
= \Pr [|S_n| \geq \alpha] + \sum_{k=1}^{n} \Pr[A_k \land (|S_n| < \alpha)] \quad (\{A_k\}_{k=1}^{n} \text{ are disjoint events.})
\]

\[
= \Pr [|S_n| \geq \alpha] + \sum_{k=1}^{n-1} \Pr[A_k \land (|S_n| < \alpha)] \quad (\Pr[A_n \land (|S_n| < \alpha)] = 0)
\]

\[
\leq \Pr [|S_n| \geq \alpha] + \sum_{k=1}^{n-1} \Pr[A_k \land (|S_n - S_k| > 2\alpha)]
\]

| \begin{align*}
|S_n| &< \alpha \land |S_k| \geq 3\alpha \\
\Rightarrow & \quad (-\alpha < S_n < \alpha \land S_k \geq 3\alpha) \lor (-\alpha < S_n < \alpha \land S_k \leq -3\alpha) \\
\Rightarrow & \quad (S_n < \alpha \land -S_k \leq -3\alpha) \lor (S_n > -\alpha \land -S_k \geq 3\alpha) \\
\Rightarrow & \quad (S_n - S_k < -2\alpha) \lor (S_n - S_k > 2\alpha) \\
\Rightarrow & \quad |S_n - S_k| > 2\alpha.
\end{align*} |
Maximal inequalities

(Continue from the previous slide)

\[
\Pr \left[ \max_{1 \leq k \leq n} |S_k| \geq 3\alpha \right] \leq \Pr \left[ |S_n| \geq \alpha \right] + \sum_{k=1}^{n-1} \Pr [A_k \wedge (|S_n - S_k| > 2\alpha)] \\
= \Pr \left[ |S_n| \geq \alpha \right] + \sum_{k=1}^{n-1} \Pr [A_k] \Pr [|S_n - S_k| > 2\alpha] \\
\quad \text{(by the independence of } A_k \text{ and } |S_n - S_k|) \\
\leq \Pr \left[ |S_n| \geq \alpha \right] + \max_{1 \leq k \leq n} \Pr [|S_n - S_k| \geq 2\alpha] \\
\leq \Pr \left[ |S_n| \geq \alpha \right] + \max_{1 \leq k \leq n} \left( \Pr [|S_n| \geq \alpha] + \Pr [|S_k| \geq \alpha] \right) \\
\leq \max_{1 \leq k \leq n} \Pr [|S_k| \geq \alpha] + \max_{1 \leq k \leq n} \Pr [|S_k| \geq \alpha] + \max_{1 \leq k \leq n} \Pr [|S_k| \geq \alpha] \\
= 3 \max_{1 \leq k \leq n} \Pr [|S_k| \geq \alpha].
\]

\[\square\]
Convergence of $X_1 + X_2 + \cdots + X_n$  

**Theorem (Implication of Kolmogorov’s zero-one law)** If $X_1, X_2, \ldots$ are independent binary 0-1 random variables, then $\Pr \left[ \sum_{k=1}^{\infty} X_k < \infty \right]$ is either 1 or 0.

**Proof:**

- Define the event $A_k = [X_k = 1]$. Then $A_1, A_2, \ldots$ are independent events. By the two Borel-Cantelli lemmas, $\Pr[A_n \text{ i.o}]$ is either 1 or 0.

**Theorem 4.3 (First Borel-Cantelli lemma)**

\[
\sum_{n=1}^{\infty} P(A_n) < \infty \Rightarrow \Pr\left(\limsup_{n\to\infty} A_n\right) = \Pr(A_n \text{ i.o.}) = 0.
\]

**Theorem 4.4 (Second Borel-Cantelli Lemma)** If $\{A_n\}_{n=1}^{\infty}$ forms an independent sequence of events,

\[
\sum_{n=1}^{\infty} P(A_n) = \infty \Rightarrow \Pr\left(\limsup_{n\to\infty} A_n\right) = \Pr(A_n \text{ i.o.}) = 1.
\]

- Apparently, if $A_1, A_2, \ldots$ are valid infinitely often in $n$ with probability 1, $\sum_{k=1}^{n} X_k = \infty$ with probability 1.
Convergence of $X_1 + X_2 + \cdots + X_n$

- On the contrary, if $A_1, A_2, \ldots$ are valid finitely many times in $n$ with probability 1, $\sum_{k=1}^{\infty} X_k < \infty$ with probability 1. $\square$

**Theorem (general version)** If $X_1, X_2, \ldots$ are independent random variables, then $\Pr\left[\sum_{k=1}^{\infty} X_k < \infty\right]$ is either 1 or 0.

- In general, to determine whether $\sum_{k=1}^{\infty} X_k$ converge or diverge is hard.

- In what follows, we provide theorems that can tell whether $\sum_{k=1}^{\infty} X_k$ converges by their moments.
**Convergence of $X_1 + X_2 + \cdots + X_n$**

**Theorem 22.6** Suppose that $X_1, X_2, \ldots$ are pair-wise independent with zero mean. Then, if $\sum_{k=1}^{\infty} \text{Var}[X_k] < \infty$, $\sum_{k=1}^{\infty} X_k < \infty$ with probability 1.

**Proof:**

Again, I use a different proof from Billingsley’s book, which is easier to understand for engineering-major students. It suffices to prove that $\Pr[\max_{k \geq 1} |S_{n+k}| < \infty] = 1$.

- First, for any $n$ fixed, $|S_n| < \infty$ with probability 1 because it were not true, we have $\Pr[|S_n| = \infty] > 0$. Derive

  $$\Pr[|S_n| \geq L] \leq \frac{1}{L^2} \sum_{k=1}^{n} \text{Var}[X_k]. \quad \text{(by zero mean and Chebyshev’s ineq)}$$

  As $\Pr[|S_n| \geq L]$ is non-increasing in $L$, its limit exists, and

  $$\lim_{L \to \infty} \Pr[|S_n| \geq L] = 0,$$

  a contradiction to $\Pr[|S_n| = \infty] > 0$. 

Convergence of $X_1 + X_2 + \cdots + X_n$

- Secondly, for any $n$ fixed, $\max_{k \geq 1} |S_{n+k} - S_n| < \infty$ with probability 1. Because if $\Pr[\max_{k \geq 1} |S_{n+k} - S_n| = \infty] > 0$, then a contradiction can be obtained as follows.

$$
\Pr \left[ \max_{1 \leq k \leq r} |S_{n+k} - S_n| \geq L \right] \leq \frac{1}{L^2} \Var [S_{n+r} - S_n] \quad \text{(by Theorem 22.4 on slide 22-19)}
$$

$$
= \frac{1}{L^2} \Var [X_{n+1} + \cdots + X_{n+r}]
$$

$$
= \frac{1}{L^2} \sum_{k=1}^{r} \Var [X_{n+k}] \quad \text{(by pair-wise independence)}
$$

$$
\leq \frac{1}{L^2} \sum_{k=1}^{\infty} \Var [X_{n+k}].
$$

Since $\Pr[\max_{1 \leq k \leq r} |S_{n+k} - S_n| \geq L]$ is non-decreasing in $r$, its limit exists by the monotone convergence theorem. Thus,

$$
\lim_{r \to \infty} \Pr \left[ \max_{1 \leq k \leq r} |S_{n+k} - S_n| \geq L \right] = \Pr \left[ \max_{k \geq 1} |S_{n+k} - S_n| \geq L \right] \leq \frac{1}{L^2} \sum_{k=1}^{\infty} \Var [X_{n+k}].
$$

Then by taking $L$ to infinity, we obtain the same contradiction as the previous item.
Convergence of $X_1 + X_2 + \cdots + X_n$

- Thirdly,

$$\Pr[|S_n| < \infty] = 1 \quad \text{and} \quad \Pr\left[\max_{k \geq 1} |S_{n+k} - S_n| < \infty\right] = 1$$

imply

$$\Pr\left[|S_n| < \infty \land \max_{k \geq 1} |S_{n+k} - S_n| < \infty\right] = 1.$$

Pr($A$) = 1 and Pr($B$) = 1 $\Rightarrow$ Pr($A \cup B$) = 1

$\Rightarrow$ Pr($A \cap B$) = Pr($A$) + Pr($B$) – Pr($A \cup B$) = 1.

By

$$\max_{k \geq 1} |S_{n+k}| \leq \max_{k \geq 1} (|S_{n+k} - S_n| + |S_n|) \leq \max_{k \geq 1} (|S_{n+k} - S_n|) + |S_n|,$$

we get:

$$\Pr\left[\max_{k \geq 1} |S_{n+k}| < \infty\right] \geq \Pr\left[|S_n| < \infty \land \max_{k \geq 1} |S_{n+k} - S_n| < \infty\right] = 1.$$
Convergence of $X_1 + X_2 + \cdots + X_n$

Example 22.2 The Rademacher functions $\{r_n(\omega)\}_{n=1}^{\infty}$ on a unit interval are defined as:

$$r_n(\omega) = \begin{cases} +1, & \text{if } d_n = 1; \\ -1, & \text{if } d_n = 0, \end{cases}$$

where $\omega = .d_1d_2d_3\ldots$ is a number lying in $[0, 1)$.

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

Define $R_n = r_n(W)$. Then $\{R_n\}_{n=1}^{\infty}$ is i.i.d. with uniform marginal.

Also, define $X_n = a_nR_n$, where $\{a_n\}_{n=1}^{\infty}$ is a constant sequence.

As a result,

$$\text{Var}[X_n] = a_n^2 \text{Var}[R_n] = a_n^2.$$

By Theorem 22.6,

$$\sum_{n=1}^{\infty} \text{Var}[X_n] = \sum_{n=1}^{\infty} a_n^2 < \infty \quad \Rightarrow \quad \sum_{n=1}^{\infty} X_n < \infty \text{ with probability 1.}$$
Convergence of $X_1 + X_2 + \cdots + X_n$

A small note on $S_n = \sum_{k=1}^{n} X_k$:

- If $S_n$ converges with probability 1, then $S_n$ converges to some finite random variable $S$ with probability 1.
When convergence in prob. $\iff$ convergence w.p. 1?

**Theorem 22.7** For an independent sequence $\{X_n\}$,

$$\sum_{k=1}^{\infty} X_k$$ converges with probability 1

if, and only if,

$$\sum_{k=1}^{\infty} X_k$$ converges in probability.

**Proof:**

1. $\sum_{n=1}^{\infty} X_n$ converges with probability 1

implies

$$\sum_{n=1}^{\infty} X_n$$ converges in probability

is a already known result. So its proof is omitted.
When convergence in prob. \( \Leftrightarrow \) convergence w.p. 1?

1. \( S_n \) converges with probability 1 if
   \[
   \lim_{n \to \infty} \Pr \left[ \max_{k \geq 1} |S_{n+k} - S_n| > \varepsilon \right] = 0.
   \]

2. That \( S_n \) converges to \( S \) in probability implies
   \[
   \limsup_{n \to \infty} \Pr \left[ |S_n - S| > \varepsilon \right] = 0.
   \]

2. Suppose \( S_n \) converges to \( S \) in probability.

Then from Theorem 22.5 (cf. slide 22-22),

\[
\Pr \left[ \max_{1 \leq k \leq r} |S_{n+k} - S_n| > 3\varepsilon \right] \leq 3 \max_{1 \leq k \leq r} \Pr \left[ |S_{n+k} - S_n| \geq \varepsilon \right]
\]

\[
\leq 3 \max_{1 \leq k \leq r} \left( \Pr \left[ |S_{n+k} - S| \geq \frac{\varepsilon}{2} \right] + \Pr \left[ |S_n - S| \geq \frac{\varepsilon}{2} \right] \right)
\]

\[
= 3 \max_{1 \leq k \leq r} \Pr \left[ |S_{n+k} - S| \geq \frac{\varepsilon}{2} \right] + 3 \Pr \left[ |S_n - S| \geq \frac{\varepsilon}{2} \right]
\]

\[
\leq 3 \max_{k \geq 1} \Pr \left[ |S_{n+k} - S| \geq \frac{\varepsilon}{2} \right] + 3 \Pr \left[ |S_n - S| \geq \frac{\varepsilon}{2} \right].
\]
When convergence in prob. \(\iff\) convergence w.p. 1?

So,

\[
\Pr \left[ \max_{k \geq 1} |S_{n+k} - S_n| > 3\varepsilon \right] = \lim_{r \to \infty} \Pr \left[ \max_{1 \leq k \leq r} |S_{n+k} - S_n| > 3\varepsilon \right]
\leq 3 \max_{k \geq 1} \Pr \left[ |S_{n+k} - S| \geq \frac{\varepsilon}{2} \right] + 3 \Pr \left[ |S_n - S| \geq \frac{\varepsilon}{2} \right],
\]
When convergence in prob. $\iff$ convergence w.p. 1?

which implies

$$\limsup_{n \to \infty} \Pr \left[ \max_{k \geq 1} |S_{n+k} - S_n| > 3\varepsilon \right] \leq 3 \limsup_{n \to \infty} \max_{k \geq 1} \Pr \left[ |S_{n+k} - S| \geq \frac{\varepsilon}{2} \right]$$

$$+ 3 \limsup_{n \to \infty} \Pr \left[ |S_n - S| \geq \frac{\varepsilon}{2} \right]$$

$$= 3 \limsup_{n \to \infty} \max_{l \geq n, k \geq 1} \Pr \left[ |S_{l+k} - S| \geq \frac{\varepsilon}{2} \right]$$

$$+ 3 \limsup_{n \to \infty} \Pr \left[ |S_n - S| \geq \frac{\varepsilon}{2} \right]$$

$$= 3 \limsup_{n \to \infty} \Pr \left[ |S_{k'} - S| \geq \frac{\varepsilon}{2} \right]$$

$$+ 3 \limsup_{n \to \infty} \Pr \left[ |S_n - S| \geq \frac{\varepsilon}{2} \right]$$

$$= 3 \limsup_{n \to \infty} \max_{k' \geq n+1} \Pr \left[ |S_{k'} - S| \geq \frac{\varepsilon}{2} \right]$$

$$+ 3 \limsup_{n \to \infty} \Pr \left[ |S_n - S| \geq \frac{\varepsilon}{2} \right]$$

$$= 3 \limsup_{n \to \infty} \Pr \left[ |S_n - S| \geq \frac{\varepsilon}{2} \right]$$

$$+ 3 \limsup_{n \to \infty} \Pr \left[ |S_n - S| \geq \frac{\varepsilon}{2} \right] = 0.$$
Three-series theorem

• Alternative conditions for convergence with probability 1.

**Theorem 22.8 (Three-series theorem)** Suppose that \( \{X_n\}_{n=1}^{\infty} \) is independent. Then

1. If
   \[
   \sum_{n=1}^{\infty} \Pr[|X_n| > c], \quad \sum_{n=1}^{\infty} E[X_n I_{|X_n| \leq c}], \quad \text{and} \quad \sum_{n=1}^{\infty} \text{Var}[X_n I_{|X_n| \leq c}]
   \]
   converges for some positive \( c \), then \( \sum_{n=1}^{\infty} X_n \) converges with probability 1.

2. If \( \sum_{n=1}^{\infty} X_n \) converges with probability 1, then
   \[
   \sum_{n=1}^{\infty} \Pr[|X_n| > c], \quad \sum_{n=1}^{\infty} E[X_n I_{|X_n| \leq c}], \quad \text{and} \quad \sum_{n=1}^{\infty} \text{Var}[X_n I_{|X_n| \leq c}]
   \]
   converge for all positive \( c \).

**Proof:** Omitted.
Three-series theorem

Example 22.3 Continue from Example 22.2.
Define \( X_n = a_n R_n \), where \( \{a_n\}_{n=1}^{\infty} \) is a constant sequence, and \( \{R_n\}_{n=1}^{\infty} \) is i.i.d. with \( \Pr[R_n = 1] = \Pr[R_n = -1] = 1/2 \).

By Theorem 22.6,

\[
\sum_{n=1}^{\infty} \text{Var}[X_n] = \sum_{n=1}^{\infty} a_n^2 < \infty \implies \sum_{n=1}^{\infty} X_n \text{ converges with probability 1.}
\]

By Theorem 22.8,

\[
\sum_{n=1}^{\infty} X_n \text{ converges with probability 1} \implies \sum_{n=1}^{\infty} \text{Var}[a_n R_n] = \sum_{n=1}^{\infty} a_n^2 < \infty.
\]

So \( \sum_{n=1}^{\infty} X_n \) converges with probability 1 if, and only if, \( \sum_{n=1}^{\infty} a_n^2 < \infty \).
Three-series theorem

By Theorem 22.8,

\[ \sum_{n=1}^{\infty} X_n \text{ converges with probability 1} \Rightarrow \sum_{n=1}^{\infty} \text{Pr}[|a_n R_n| > c] = \sum_{n=1}^{\infty} I_{[|a_n| > c]} < \infty. \]

\[ \Rightarrow a_n \text{ is bounded infinitely often in } n. \quad \square \]