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

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

- Consider the truncated random variable $Y_k = X_k I_{[X_k \le k]}$, 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 $k \le n$,

$$E[Y_k^2] = E[X_k^2 I_{[X_k \le k]}] = E[X_1^2 I_{[X_1 \le k]}] \le E[X_1^2 I_{[X_1 \le n]}] = E[Y_n^2].$$

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 $u_n \triangleq \lfloor \alpha^n \rfloor$ with $\alpha > 1$ fixed,

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

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,

$$\sum_{n=1}^{\infty} \Pr\left[\left|\frac{S_{u_n}^* - E[S_{u_n}^*]}{u_n}\right| > \varepsilon\right] \le \sum_{n=1}^{\infty} \frac{\mathrm{Var}[S_{u_n}^*]}{u_n^2 \varepsilon^2} \le \frac{1}{\varepsilon^2} \sum_{n=1}^{\infty} \frac{E[Y_{u_n}^2]}{u_n},$$

where by pair-wise independence,

$$\mathrm{Var}[S_{u_n}^*] = \sum_{k=1}^{u_n} \mathrm{Var}[Y_k] \le u_n E[Y_{u_n}^2].$$

Law of large numbers revisited

Hence,

$$\begin{aligned}
\sum_{n=1}^{\infty}\Pr\left[\left|\frac{S^*_{u_n}-E[S^*_{u_n}]}{u_n}\right|>\varepsilon\right] &\le \frac{1}{\varepsilon^2}\sum_{n=1}^{\infty}\frac{E[Y^2_{u_n}]}{u_n}=\frac{1}{\varepsilon^2}\sum_{n=1}^{\infty}\frac{E[X^2_{u_n}I_{[X_{u_n}\le u_n]}]}{u_n}\\
&= \frac{1}{\varepsilon^2}\sum_{n=1}^{\infty}\frac{E[X_1^2 I_{[X_1\le u_n]}]}{u_n}=\frac{1}{\varepsilon^2}\lim_{m\to\infty}\sum_{n=1}^{m}\frac{E[X_1^2 I_{[X_1\le u_n]}]}{u_n}\\
&= \frac{1}{\varepsilon^2}\lim_{m\to\infty}E\left[X_1^2\sum_{n=1}^{m}\frac{1}{u_n}I_{[X_1\le u_n]}\right] \quad \left(f_m(x)\triangleq x^2\sum_{n=1}^{m}\frac{1}{u_n}I_{[x\le u_n]}\right)\\
&= \frac{1}{\varepsilon^2}E\left[X_1^2\lim_{m\to\infty}\sum_{n=1}^{m}\frac{1}{u_n}I_{[X_1\le u_n]}\right] \quad \text{(by monotone conv. thm.)}\\
&= \frac{1}{\varepsilon^2}E\left[X_1^2\sum_{n=1}^{\infty}\frac{1}{u_n}I_{[X_1\le u_n]}\right]
\end{aligned}$$

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

$$\begin{aligned}
\sum_{n=1}^{\infty} \frac{1}{u_n} I_{[x \le u_n]} &= \sum_{\{n \in \mathbb{N} \,:\, u_n \ge x\}} \frac{1}{u_n} \\
&= \sum_{n \ge N} \frac{1}{u_n}, \text{ where } N = \min\{n \in \mathbb{N} : u_n \ge x\} \\
&\le \sum_{n \ge N} \frac{2}{\alpha^n}, \text{ (since } u_n = \lfloor \alpha^n \rfloor \text{ and } \lfloor y \rfloor \ge \frac{1}{2} y \text{ for } y \ge 1) \\
&= \left(\frac{2}{1 - \alpha^{-1}} \right) \frac{1}{\alpha^N} \\
&\le \left(\frac{2\alpha}{\alpha - 1} \right) \frac{1}{x}. \quad \text{(by } \alpha^N \ge \lfloor \alpha^N \rfloor = u_N \ge x)
\end{aligned}$$

This concludes that:

$$\sum_{n=1}^{\infty} \Pr\left[\left| \frac{S^*_{u_n} - E[S^*_{u_n}]}{u_n} \right| > \varepsilon \right] \le \frac{1}{\varepsilon^2} E\left[X_1^2 \sum_{n=1}^{\infty} \frac{1}{u_n} I_{[X_1 \le u_n]} \right] \le \frac{1}{\varepsilon^2} \left(\frac{2\alpha}{\alpha - 1} \right) E\left[X_1 \right] < \infty.$$

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),

$$\lim_{u_n\to\infty} E[Y_{u_n}] \quad \left(= \lim_{n\to\infty} E[Y_{u_n}] = \lim_{n\to\infty} E[X_1 I_{[X_1 \le u_n]}] \right) \quad = E[X_1] < \infty$$

implies

$$\frac{1}{u_n} E[S^*_{u_n}] = \frac{1}{u_n} \sum_{k=1}^{u_n} E[Y_k] \to E[X_1] \quad \text{as } n \to \infty.$$

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 $\lim_{n\to\infty} a_n = a$ and $b_n = (1/n)\sum_{i=1}^{n} a_i$, where a is finite, then $\lim_{n\to\infty} b_n = a$.

Proof: $\lim_{n\to\infty} a_n = a$ implies that for any $\varepsilon > 0$, there exists N such that for all $n > N$, $|a_n - a| < \varepsilon$. Then

$$\begin{aligned} |b_n - a| &= \left|\frac{1}{n}\sum_{i=1}^{n}(a_i - a)\right| \le \frac{1}{n}\sum_{i=1}^{n}|a_i - a| \\ &= \frac{1}{n}\sum_{i=1}^{N}|a_i - a| + \frac{1}{n}\sum_{i=N+1}^{n}|a_i - a| \\ &\le \frac{1}{n}\sum_{i=1}^{N}|a_i - a| + \frac{n-N}{n}\varepsilon. \end{aligned}$$

Hence, $\lim_{n\to\infty} |b_n - a| \le \varepsilon$. Since ε can be made arbitrarily small, $\lim_{n\to\infty} b_n = a$. □

Law of large numbers revisited

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

  Proof of the claim:

$$\begin{aligned}
\sum_{n=1}^{\infty} \Pr[X_n \neq Y_n] &= \sum_{n=1}^{\infty} \Pr[X_n \neq X_n I_{[X_n \leq n]}] \\
&= \sum_{n=1}^{\infty} \Pr[X_n > n] \\
&= \sum_{n=1}^{\infty} \Pr[X_1 > n] \text{ (by ``identical distributed'' assumption)} \\
&\leq \int_0^{\infty} \Pr[X_1 > t] dt \\
&= E[X_1] \text{ (by non-negativity assumption of } X_1) \\
&< \infty.
\end{aligned}$$

  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

$$\Pr\left[(\exists\ \mathbb{U} = \{n_1, n_2, \ldots, n_M\})\ X_n \neq Y_n \text{ only for } n \in \mathbb{U}\right] = 1.$$

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:

$$\begin{aligned} &\Pr\left[\lim_{n\to\infty} \frac{(X_1 - Y_1) + \cdots + (X_n - Y_n)}{n} = 0\right] \\ = &\ \Pr\left[\lim_{n\to\infty} \frac{(X_{n_1} - Y_{n_1}) + \cdots + (X_{n_M} - Y_{n_M})}{n} = 0\right] \\ = &\ 1. \end{aligned}$$

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 $u_n \le k < u_{n+1}$,

$$
\begin{aligned}
\boxed{\frac{u_n}{u_{n+1}}\frac{S_{u_n}}{u_n}} &= \frac{S_{u_n}}{u_{n+1}} \\
&= \frac{X_1 + \cdots + X_{u_n}}{u_{n+1}} \\
&\le \frac{X_1 + \cdots + X_{u_n}}{k} \\
&\le \frac{X_1 + \cdots + X_{u_n} + \cdots + X_k}{k} = \boxed{\frac{S_k}{k}} \\
&\le \frac{X_1 + \cdots + X_{u_n} + \cdots + X_k}{u_n} \\
&\le \frac{X_1 + \cdots + X_{u_n} + \cdots + X_k + \cdots + X_{u_{n+1}}}{u_n} \\
&= \frac{S_{u_{n+1}}}{u_n} \\
&= \boxed{\frac{u_{n+1}}{u_n}\frac{S_{u_{n+1}}}{u_{n+1}}},
\end{aligned}
$$

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:

$$\frac{1}{\alpha}E[X_1] \le \liminf_{k\to\infty}\frac{S_k}{k} \le \limsup_{k\to\infty}\frac{S_k}{k} \le \alpha E[X_1] \text{ with probability 1.}$$

As the above statement is valid for any $\alpha > 1$, we conclude that

$$\frac{S_k}{k} \to E[X_1] \text{ with probability 1.}$$

$\square$

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

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

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

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

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

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

  Hence,

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

Law of large numbers revisited

$$\begin{cases} \Pr[A_n \geq B_n] = 1 \\ \Pr\left[\lim_{n\to\infty} B_n = b\right] = 1 \end{cases} \quad \Rightarrow \begin{cases} \Pr\left[\liminf_{n\to\infty} A_n \geq b\right] \\ \geq \Pr\left[\liminf_{n\to\infty} A_n \geq b \ \wedge \ \liminf_{n\to\infty} B_n = b\right] \\ \geq \Pr\left[\liminf_{n\to\infty} A_n \geq \liminf_{n\to\infty} B_n \ \wedge \ \liminf_{n\to\infty} B_n = b\right] \\ = 1 \end{cases}$$

- 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_λ be a Poisson random variable with parameter λ, and let $G_\lambda(\cdot)$ be the cdf of a Y_λ/λ. 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,

$$\Pr\left[\left|\frac{Y_\lambda - \lambda}{\lambda}\right| \geq \varepsilon\right] = \Pr\left[|Y_\lambda - \lambda| \geq \varepsilon\lambda\right] \leq \frac{\mathrm{Var}[Y_\lambda]}{\lambda^2\varepsilon^2} = \frac{\lambda}{\lambda^2\varepsilon^2} = \frac{1}{\lambda\varepsilon^2} \to 0$$

as $\lambda \to \infty$. □

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 $y > 0$,

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

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 $s > 0$,

$$\begin{aligned}
\sum_{k=0}^{\lfloor sy \rfloor} \frac{(-1)^k}{k!} s^k M_X^{(k)}(s)_+ &= \sum_{k=0}^{\lfloor sy \rfloor} \frac{(-1)^k}{k!} s^k \left((-1)^k \int_0^\infty x^k e^{-sx} dF_X(x) \right) \\
&= \int_0^\infty \sum_{k=0}^{\lfloor sy \rfloor} e^{-sx} \frac{(sx)^k}{k!} dF_X(x) \\
&= \int_0^\infty \Pr\left[Y_{sx} \le \lfloor sy \rfloor \right] dF_X(x) \\
&= \int_0^\infty \Pr\left[Y_{sx} \le sy \right] dF_X(x) \\
&= \int_0^\infty \Pr\left[\frac{Y_{sx}}{sx} \le \frac{y}{x} \right] dF_X(x) \\
&= \int_0^\infty G_{sx}\left(\frac{y}{x} \right) dF_X(x).
\end{aligned}$$

Application of degeneration of $G_\lambda(t)$ as $\lambda \to \infty$

As a result,

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

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

Give a sequence of non-negative μ-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 μ-measure zero.

Lemma (Fatou's lemma) $\int_{\mathcal{X}}\left[\lim_{n\to\infty} f_n(x)\right]\mu(dx) \le \liminf_{n\to\infty}\int_{\mathcal{X}} f_n(x)\mu(dx).$

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, $f_n(x) \le g(x)$ for all $x \in \mathcal{X}$, except on a subset of $\mathcal{X}$ with μ-measure zero, and $g(\cdot)$ is μ-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),

$$\begin{aligned}\lim_{s\to\infty}\sum_{k=0}^{\lfloor sy\rfloor}\frac{(-1)^k}{k!}s^k M_X^{(k)}(s)_+ &= \int_0^\infty \lim_{s\to\infty} G_{sx}\left(\frac{y}{x}\right)dF_X(x)\\ &= \int_0^y dF_X(x)\\ &= \Pr[X\le y].\end{aligned}$$

(How to determine $\Pr[X \le y]$ when X has point mass at y? Hint: Right-continuity)

□

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 $y > 0$,

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

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

Final comment: In fact, to determine 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 $\alpha > 0$,

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

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

Chebyshev's inequality said that

$$\Pr[|S_n| \ge \alpha] \le \frac{1}{\alpha^2}\text{Var}[S_n].$$

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

Proof: Define the event

$$A_k = \left[|S_1| < \alpha \wedge |S_2| < \alpha \wedge \cdots \wedge |S_{k-1}| < \alpha \wedge |S_k| \ge \alpha\right].$$

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

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

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:

$$
\begin{aligned}
E[S_n^2] &\geq \sum_{k=1}^{n} \left(E\left[S_k^2 I_{A_k}\right] + 2E\left[S_k I_{A_k}\right] E\left[S_n - S_k\right] \right) \\
&= \sum_{k=1}^{n} E\left[S_k^2 I_{A_k}\right] \quad \text{(by the zero mean assumption, } E[S_n - S_k] = 0\text{)} \\
&\geq \sum_{k=1}^{n} E\left[\alpha^2 I_{A_k}\right] \quad (I_{A_k} = 1 \text{ only when } |S_k| \geq \alpha) \\
&= \alpha^2 \sum_{k=1}^{n} \Pr[A_k] \\
&= \alpha^2 \Pr\left[\max_{1\leq k\leq n} |S_k| \geq \alpha\right].
\end{aligned}
$$

□

Maximal inequalities

The previous theorem provides a bound for the cdf of $\max_{1\le k\le n} |S_k|$ using the second moment.

We can also bound the cdf of $\max_{1\le k\le n} |S_k|$ by the cdf of $|S_k|$ for $1 \le k \le n$.

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

$$\Pr\left[\max_{1\le k\le n} |S_k| \ge 3\alpha\right] \le 3 \max_{1\le k\le n} \Pr[|S_k| \ge \alpha].$$

Proof: Define the event

$$A_k = \big[|S_1| < 3\alpha \wedge |S_2| < 3\alpha \wedge \cdots \wedge |S_{k-1}| < 3\alpha \wedge |S_k| \ge 3\alpha\big].$$

Then

$$\begin{aligned}
\Pr\left[\max_{1\le k\le n} |S_k| \ge 3\alpha\right] &= \Pr\left[\left(\max_{1\le k\le n} |S_k| \ge 3\alpha\right) \wedge (|S_n| \ge \alpha)\right] \\
&\quad + \Pr\left[\left(\max_{1\le k\le n} |S_k| \ge 3\alpha\right) \wedge (|S_n| < \alpha)\right] \\
&\le \Pr[|S_n| \ge \alpha] + \Pr\left[\left(\max_{1\le k\le n} |S_k| \ge 3\alpha\right) \wedge (|S_n| < \alpha)\right].
\end{aligned}$$

Maximal inequalities

(Continue from the previous slide)

$$\begin{aligned}
\Pr\left[\max_{1\le k\le n}|S_k|\ge 3\alpha\right] &\le \Pr\left[|S_n|\ge\alpha\right]+\Pr\left[\left(\max_{1\le k\le n}|S_k|\ge 3\alpha\right)\wedge(|S_n|<\alpha)\right]\\
&= \Pr\left[|S_n|\ge\alpha\right]+\Pr\left[(A_1\vee A_2\vee\cdots\vee A_n)\wedge(|S_n|<\alpha)\right]\\
&= \Pr\left[|S_n|\ge\alpha\right]+\sum_{k=1}^{n}\Pr\left[A_k\wedge(|S_n|<\alpha)\right]\quad(\{A_k\}_{k=1}^n\text{ are disjoint events.})\\
&= \Pr\left[|S_n|\ge\alpha\right]+\sum_{k=1}^{n-1}\Pr\left[A_k\wedge(|S_n|<\alpha)\right]\quad(\Pr[A_n\wedge(|S_n|<\alpha)]=0)\\
&\le \Pr\left[|S_n|\ge\alpha\right]+\sum_{k=1}^{n-1}\Pr\left[A_k\wedge(|S_n-S_k|>2\alpha)\right]
\end{aligned}$$

$$\begin{aligned}
&|S_n|<\alpha\wedge|S_k|\ge 3\alpha\\
&\Rightarrow\ (-\alpha<S_n<\alpha\wedge S_k\ge 3\alpha)\vee(-\alpha<S_n<\alpha\wedge S_k\le -3\alpha)\\
&\Rightarrow\ (S_n<\alpha\wedge -S_k\le -3\alpha)\vee(S_n>-\alpha\wedge -S_k\ge 3\alpha)\\
&\Rightarrow\ (S_n-S_k<-2\alpha)\vee(S_n-S_k>2\alpha)\\
&\Rightarrow\ |S_n-S_k|>2\alpha.
\end{aligned}$$

Maximal inequalities

(Continue from the previous slide)

$$
\begin{aligned}
\Pr\left[\max_{1\le k\le n}|S_k|\ge 3\alpha\right] &\le \Pr[|S_n|\ge\alpha]+\sum_{k=1}^{n-1}\Pr[A_k\wedge(|S_n-S_k|>2\alpha)]\\
&= \Pr[|S_n|\ge\alpha]+\sum_{k=1}^{n-1}\Pr[A_k]\Pr[|S_n-S_k|>2\alpha]\\
&\qquad \text{(by the independence of } A_k \text{ and } |S_n-S_k|)\\
&\le \Pr[|S_n|\ge\alpha]+\max_{1\le k\le n}\Pr[|S_n-S_k|\ge 2\alpha]\\
&\le \Pr[|S_n|\ge\alpha]+\max_{1\le k\le n}\Pr[|S_n|\ge\alpha\vee|S_k|\ge\alpha]\\
&\quad \text{(Notably, } |x|<\alpha \text{ and } |y|<\alpha \text{ imply } |x-y|<2\alpha.)\\
&\le \Pr[|S_n|\ge\alpha]+\max_{1\le k\le n}\left(\Pr[|S_n|\ge\alpha]+\Pr[|S_k|\ge\alpha]\right)\\
&\le \max_{1\le k\le n}\Pr[|S_k|\ge\alpha]+\max_{1\le k\le n}\Pr[|S_k|\ge\alpha]+\max_{1\le k\le n}\Pr[|S_k|\ge\alpha]\\
&= 3\max_{1\le k\le n}\Pr[|S_k|\ge\alpha].
\end{aligned}
$$

□

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)

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

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

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

- 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. □

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.

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

Proof:

Again, I use a different proof from that in 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

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

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

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

$$\begin{aligned}
\Pr\left[\max_{1\leq k\leq r}|S_{n+k}-S_n| \geq L\right] &\leq \frac{1}{L^2}\text{Var}\left[S_{n+r}-S_n\right] \text{ (by Theorem 22.4 on Slide 22-19)}\\
&= \frac{1}{L^2}\text{Var}\left[X_{n+1}+\cdots+X_{n+r}\right]\\
&= \frac{1}{L^2}\sum_{k=1}^{r}\text{Var}[X_{n+k}] \text{ (by pair-wise independence)}\\
&\leq \frac{1}{L^2}\sum_{k=1}^{\infty}\text{Var}[X_{n+k}].
\end{aligned}$$

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

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

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

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

$$\begin{aligned}\Pr(A) = 1 \text{ and } \Pr(B) = 1 \;&\Rightarrow\; \Pr(A \cup B) = 1\\ &\Rightarrow\; \Pr(A \cap B) = \Pr(A) + \Pr(B) - \Pr(A \cup B) = 1.\end{aligned}$$

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:

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

$\square$

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 $\omega = .d_1 d_2 d_3 \ldots$ is a number lying in $[0, 1)$.

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

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

Also, define $X_n = a_n R_n$, 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. $\Leftrightarrow$ convergence w.p. 1?

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

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

if, and only if,

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

Proof:

1.
$$\sum_{n=1}^{\infty} X_n \text{ converges with probability 1}$$
   implies
$$\sum_{n=1}^{\infty} X_n \text{ converges in probability}$$
   is a known result.

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),

$$\begin{aligned}
\Pr\left[\max_{1\leq k\leq r} |S_{n+k} - S_n| > 3\varepsilon\right] &\leq 3 \max_{1\leq k\leq r} \Pr\left[|S_{n+k} - S_n| \geq \varepsilon\right] \\
&\leq 3 \max_{1\leq k\leq r} \left(\Pr\left[|S_{n+k} - S| \geq \frac{\varepsilon}{2}\right] + \Pr\left[|S_n - S| \geq \frac{\varepsilon}{2}\right]\right) \\
&= 3 \max_{1\leq k\leq r} \Pr\left[|S_{n+k} - S| \geq \frac{\varepsilon}{2}\right] + 3\Pr\left[|S_n - S| \geq \frac{\varepsilon}{2}\right] \\
&\leq 3 \max_{k\geq 1} \Pr\left[|S_{n+k} - S| \geq \frac{\varepsilon}{2}\right] + 3\Pr\left[|S_n - S| \geq \frac{\varepsilon}{2}\right].
\end{aligned}$$

When convergence in prob. $\Leftrightarrow$ convergence w.p. 1?

So,

$$\begin{aligned}\Pr\left[\max_{k\geq 1}|S_{n+k}-S_n|>3\varepsilon\right] &= \lim_{r\to\infty}\Pr\left[\max_{1\leq k\leq r}|S_{n+k}-S_n|>3\varepsilon\right]\\ &\leq 3\max_{k\geq 1}\Pr\left[|S_{n+k}-S|\geq\frac{\varepsilon}{2}\right]+3\Pr\left[|S_n-S|\geq\frac{\varepsilon}{2}\right],\end{aligned}$$

which implies

$$\begin{aligned}&\limsup_{n\to\infty}\Pr\left[\max_{k\geq 1}|S_{n+k}-S_n|>3\varepsilon\right]\\ &\leq 3\limsup_{n\to\infty}\max_{k\geq 1}\Pr\left[|S_{n+k}-S|\geq\frac{\varepsilon}{2}\right]+3\limsup_{n\to\infty}\Pr\left[|S_n-S|\geq\frac{\varepsilon}{2}\right]\\ &= 3\lim_{n\to\infty}\sup_{\ell\geq n}\max_{k\geq 1}\Pr\left[|S_{\ell+k}-S|\geq\frac{\varepsilon}{2}\right]+3\limsup_{n\to\infty}\Pr\left[|S_n-S|\geq\frac{\varepsilon}{2}\right]\\ &= 3\lim_{n\to\infty}\sup_{k'\geq n+1}\Pr\left[|S_{k'}-S|\geq\frac{\varepsilon}{2}\right]+3\limsup_{n\to\infty}\Pr\left[|S_n-S|\geq\frac{\varepsilon}{2}\right]\\ &= 3\limsup_{n\to\infty}\Pr\left[|S_n-S|\geq\frac{\varepsilon}{2}\right]+3\limsup_{n\to\infty}\Pr\left[|S_n-S|\geq\frac{\varepsilon}{2}\right]=0.\end{aligned}$$

$\square$

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

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

Proof: Omitted. □

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,

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

By Theorem 22.8,

$$\sum_{n=1}^{\infty} X_n \text{ converges with probability 1} \quad \Rightarrow \quad \sum_{n=1}^{\infty} \mathrm{Var}\left[a_n R_n\right] = \sum_{n=1}^{\infty} a_n^2 < \infty.$$

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,

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

$$\Rightarrow \quad a_n \text{ is bounded infinitely often in } n. \quad \Box$$