# 8. Limit Theorems



## Modes of convergence

For real numbers, there's only one mode (i.e. only one meaning for "converges"):

$a_n$ converges to $a$ means

$$\forall \varepsilon > 0, \exists N \in \mathbb{N}, \forall n \geq N, \ |a_n - a| < \varepsilon$$

Notation: $\lim_{n\to\infty} a_n = a$

or $a_n \to a$ as $n \to \infty$

or $a_n \xrightarrow{n\to\infty} a$

---

For random variables, there are several modes.

- $X_n \to X$ pointwise means:

$$\forall \omega \in \Omega, \ \lim_{n\to\infty} X_n(\omega) = X(\omega)$$

- $X_n \to X$ almost surely (or a.s.) means:

$$P\left(\lim_{n\to\infty} X_n = X\right) = 1$$

- $X_n \to X$ in probability means:

$$\forall \varepsilon > 0, \quad \lim_{n\to\infty} P(|X_n - X| > \varepsilon) = 0$$

- $X_n \to X$ in distribution means:

$\forall x$ at which $F_X$ is continuous,

$$\lim_{n\to\infty} F_{X_n}(x) = F_X(x) \text{ whenever } F$$

---

Thm: Consider the 4 possibilites:

(a) $X_n \to X$ pointwise

(b) $X_n \to X$ a.s.

(c) $X_n \to X$ in probability

(d) $X_n \to X$ in distribution

Then (a) $\Rightarrow$ (b) $\Rightarrow$ (c) $\Rightarrow$ (d)

---

Thm If $M_{X_n}(t) \to M_X(t)$, then $X_n \to X$ in distribution.

---

Pfs are beyond us.

2nd used to prove central limit theorem.

# 8.2 Chebyshev's Inequality and the Weak Law of Large Numbers

(LLN)

**Prop. 2.1** (Markov's inequality)

If $X \geq 0$ and $a > 0$, then

$$P(X \geq a) \leq \frac{E[X]}{a}$$

**Pf:** Let $A = \{X \geq a\}$. Then

$$E[X] \geq E[X 1_A] \geq E[a 1_A] = a \overset{P(A)}{\cancel{E[1_A]}}$$

$$\Rightarrow P(X \geq a) = P(A) \leq \frac{E[X]}{a}. \quad \square$$

**Prop. 2.2** (Chebyshev's inequality)

means the expected value exists (and is finite)

Suppose $X$ has a finite mean. Let $\mu = E[X]$. Then for all $k > 0$,

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

**Pf:** Let $Y = |X - \mu|^2$. Then $E[Y] = Var(X)$. So

$$P(|X-\mu| \geq k) = P(|X-\mu|^2 \geq k^2)$$

$$= P(Y \geq k^2) \underset{\text{Markov's inequ.}}{\leq} \frac{E[Y]}{k^2} = \frac{Var(X)}{k^2}. \quad \square$$

**Example 2a** Suppose that it is known that the number of items produced in a factory during a week is a random variable with mean 50.

**(a)** What can be said about the probability that this week's production will exceed 75?

**(b)** If the variance of a week's production is known to equal 25, then what can be said about the probability that this week's production will be between 40 and 60?

$X$ = # of items produced this week

$E[X] = 50$

(a) $P(X \geq 75) \leq \underbrace{\frac{\overset{50}{\cancel{E[X]}}}{75}}_{\text{Markov}} = \frac{2}{3}$

$$\boxed{P(X \geq 75) \leq \frac{2}{3}}$$

(b) $Var(X) = 25$

$$P(40 \leq X \leq 60) = P(|X - 50| \leq 10)$$

$$= 1 - P(|X - \underset{E[X]}{50}| \geq 10) \geq 1 - \frac{Var(X)}{10^2}$$

Chebyshev

$$= 1 - \frac{25}{100} = 1 - \frac{1}{4} = \frac{3}{4}$$

$$\boxed{P(40 \leq X \leq 60) \geq \frac{3}{4}}$$

These are terrible estimates. Not close at all.

These inequalities are typically useful for proving things.

Prop 2.3

If $Var(X)=0$, then $P(X=E[X])=1$.

Pf: Let $\mu = E[X]$. Then

$$P\left(|X-\mu| \geq \tfrac{1}{n}\right) \leq \frac{\overset{0}{\cancel{Var(X)}}}{\left(\frac{1}{n}\right)^2} = 0$$

$$\Rightarrow P\left(|X-\mu| \geq \tfrac{1}{n}\right) = 0$$

$$\{X = E[X]\} = \bigcap_{n=1}^{\infty} \underbrace{\left\{|X - E[X]| < \tfrac{1}{n}\right\}}_{A_n}$$

$$A_1 \supseteq A_2 \supseteq A_3 \supseteq \cdots$$

$$\therefore P(X = E[X]) = \lim_{n\to\infty} P(A_n)$$

$$= \lim_{n\to\infty} \left(1 - \overset{0}{\cancel{P\left(|X-E[X]| \geq \tfrac{1}{n}\right)}}\right) = 1. \quad \square$$

Thm 2.1 (The weak law of large numbers)

Suppose $X_1, X_2, \ldots$ are i.i.d. and have a finite mean. Let $\mu = E[X_n]$. Then

$$\frac{X_1 + \cdots + X_n}{n} \longrightarrow \mu \text{ in probability.}$$

Pf: Proof is beyond us. But if we assume the $X_n$'s have a finite variance, we can do it:

$$\sigma^2 = \mathrm{Var}(X_n)$$

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

$$E[S_n] = \sum_{j=1}^n \overset{\mu}{\cancel{E[X_j]}} = n\mu$$

$$\mathrm{Var}(S_n) \underset{\substack{\uparrow \\ \text{b/c } X_1, X_2, \ldots \text{ indep.}}}{=} \sum_{j=1}^n \overset{\sigma^2}{\cancel{\mathrm{Var}(X_j)}} = n\sigma^2$$

Need to show $\frac{S_n}{n} \to \mu$ in probability, which means that $\forall \varepsilon > 0$,

$$P\left(\left|\frac{S_n}{n} - \mu\right| \geq \varepsilon\right) \xrightarrow{n\to\infty} 0 .$$

Let $\varepsilon > 0$. Then

$$P\left(\left|\frac{S_n}{n} - \mu\right| \geq \varepsilon\right) = P\left(\left|\frac{S_n - \overset{E[S_n]}{\cancel{n\mu}}}{n}\right| \geq \varepsilon\right)$$

$$= P\left(|S_n - E[S_n]| \geq n\varepsilon\right) \underset{\substack{\uparrow \\ \text{Cheby}}}{\leq} \frac{\mathrm{Var}(S_n)}{(n\varepsilon)^2}$$

$$= \frac{n\sigma^2}{n^2\varepsilon^2} = \frac{\sigma^2}{n\varepsilon^2} \xrightarrow{n\to\infty} 0 . \ \square$$

→ from Section 8.4

Thm 4.1 (The strong law of large numbers)

Suppose $X_1, X_2, \ldots$ are i.i.d. and have a finite mean. Let $\mu = E[X_n]$. Then

$$\frac{X_1 + \cdots + X_n}{n} \longrightarrow \mu \quad \text{almost surely}$$

- The same as the weak law, but a stronger conclusion. At this level, the weak law is redundant. At higher levels, there are many weak and strong laws with different hypotheses.
- The pf is beyond us. If we assume $E[X_n^4] < \infty$, we can do it. See the book for details. The idea is this. First assume $\mu = 0$. Prove

$$E\left[\sum_{n=1}^{\infty} \frac{S_n^4}{n^4}\right] < \infty \Rightarrow P\left(\sum_{n=1}^{\infty} \frac{S_n^4}{n^4} < \infty\right) = 1$$

$$\Rightarrow P\left(\frac{S_n^4}{n^4} \xrightarrow{n\to\infty} 0\right) = 1.$$

  Take 4th roots. Then use this to prove it when $\mu \neq 0$.
- The LLN justifies maximizing expectation. But be sure the hypotheses of LLN hold. Some counterexamples: Kelly criterion, St. Petersburg paradox

# 8.3 The central limit theorem (CLT)

**Thm 3.1 (CLT)**

Suppose $X_1, X_2, \ldots$ are i.i.d., have a finite mean, and a finite positive variance. Let $\mu = E[X_n]$ and $\sigma^2 = \mathrm{Var}(X_n)$. Let $Z \sim N(0,1)$. Then

$$\sqrt{n}\left(\frac{S_n}{n} - \mu\right) \longrightarrow \sigma Z \quad \text{in distribution.}$$

---

Alternate form: $\dfrac{S_n - n\mu}{\sigma\sqrt{n}} \longrightarrow Z$ in distribution

LLN says $\dfrac{S_n}{n} \approx \mu \Rightarrow \boxed{S_n \approx n\mu \quad (n \text{ large})}$

CLT says $\dfrac{S_n - n\mu}{\sqrt{n}\,\sigma} \overset{d}{\approx} Z \Rightarrow \boxed{S_n \overset{d}{\approx} n\mu + \sqrt{n}\,\sigma Z}$

CLT is a higher order approximation.

- Can now use normal approx. on more than binomial.
- Only use cont. correction when approximating discrete r.v.s
- Pf is in the book. Uses MGFs.

**Example 3a** An astronomer is interested in measuring the distance, in light-years, from his observatory to a distant star. Although the astronomer has a measuring technique, he knows that because of changing atmospheric conditions and normal error, each time a measurement is made, it will not yield the exact distance, but merely an estimate. As a result, the astronomer plans to make a series of measurements and then use the average value of these measurements as his estimated value of the actual distance. If the astronomer believes that the values of the measurements are independent and identically distributed random variables having a common mean $d$ (the actual distance) and a common variance of 4 (light-years), how many measurements need he make to be reasonably sure that his estimated distance is accurate to within ±.5 light-year?

$d$ = actual distance to the star

$n$ = # of measurements the astronomer makes

$\underbrace{X_1, X_2, \ldots, X_n}_{\text{independent}}$ : results of his measurements

for all $j$, $EX_j = d$, $Var(X_j) = 4$

$D$ = estimated distance to the star

$$D = \frac{X_1 + X_2 + \cdots + X_n}{n}$$

$P(|D - d| < 0.5)$ : probability that estimated distance is accurate to within 0.5 light-years

What does "reasonably sure" mean? The problem doesn't say. But in the book's solution, they say the above probability should be at least 95%

So...

The question:

How big does $n$ need to be so that

$P(|D - d| < 0.5) \geq 0.95$?

---

Use CLT

$$S_n = X_1 + X_2 + \cdots + X_n = \sum_{j=1}^{n} X_j$$

$$D = \frac{S_n}{n}$$

CLT: $S_n$ is approx. normal

$$E[S_n] = \sum_{j=1}^{n} \underbrace{\cancel{E[X_j]}}_{d} = dn$$

$$Var(S_n) = \sum_{j=1}^{n} \underbrace{\cancel{Var(X_j)}}_{4} = 4n$$

b/c the $X_j$'s are indep.

$$S_n \overset{d}{\approx} N(dn, 4n)$$

$$S_n \overset{d}{\approx} dn + \sqrt{4n}\, Z \quad N(0,1)$$

$$= dn + 2\sqrt{n}\, Z$$

$$P(|D-d|<0.5) = P\left(\left|\frac{S_n}{n} - d\right| < 0.5\right)$$

$$\approx P\left(\left|\frac{dn + 2\sqrt{n}\,Z}{n} - d\right| < 0.5\right)$$

$$= P\left(\left|\cancel{d} + \frac{2}{\sqrt{n}} Z \cancel{- d}\right| < 0.5\right)$$

$$= P\left(\frac{2}{\sqrt{n}} |Z| < \frac{1}{2}\right)$$

$$= P\left(|Z| < \frac{\sqrt{n}}{4}\right)$$

$$= 2\Phi\left(\frac{\sqrt{n}}{4}\right) - 1$$

We want this to equal 0.95

$$2\Phi\left(\frac{\sqrt{n}}{4}\right) - 1 = 0.95$$

$$\Phi\left(\frac{\sqrt{n}}{4}\right) = \frac{1.95}{2} = 0.975$$

From the table of $\Phi$-values, $\Phi(1.96) \approx 0.9750$,

so...

$$\frac{\sqrt{n}}{4} \approx 1.96$$

$$\sqrt{n} \approx 7.84$$

$$n \approx (7.84)^2 = 61.4656$$

He needs to make an integer # of measurements, so

$$\boxed{n = 62}$$

**Example 3b** The number of students who enroll in a psychology course is a Poisson random variable with mean 100. The professor in charge of the course has decided that if the number enrolling is 120 or more, he will teach the course in two separate sections, whereas if fewer than 120 students enroll, he will teach all of the students together in a single section. What is the probability that the professor will have to teach two sections?

$X$ = # of students enrolling, $X \sim \text{Poisson}(100)$

$A$ = "Prof. teaches two sections"

$A = \{X \geq 120\}$

$P(A) = ?$

exact answer:

$$P(A) = P(X \geq 120) = \underbrace{\sum_{j=120}^{\infty} e^{-100} \cdot \frac{100^j}{j!}}_{\text{cannot simplify any more}}$$

approximate answer:

let $X_1, X_2, \ldots, X_{100}$ be independent Poisson (1)

Then $X_1 + X_2 + \cdots + X_{100} \sim \text{Poisson}(100)$

So $X \overset{d}{=} X_1 + X_2 + \cdots + X_{100} \overset{d}{\approx}$ normal by CLT

(In other words, when $\lambda$ is large, Poisson $\approx$ normal)

$$E[X] = \lambda = 100$$

$$\text{Var}(X) = \lambda = 100$$

$$X \overset{d}{\approx} \lambda + \sqrt{\lambda}\, \underbrace{Z}_{N(0,1)} = 100 + 10Z$$

$$P(X \geq 120) \approx P(100 + 10Z \geq 119.5)$$

discrete, so need continuity correction

$$= P(Z \geq 1.95)$$

$$= 1 - \Phi(1.95)$$

from table

$$\approx 1 - 0.9744 = \boxed{0.0256}$$

**Example 3c** If 10 fair dice are rolled, find the approximate probability that the sum obtained is between 30 and 40, inclusive.

$\underbrace{X_1, X_2, \ldots, X_{10}}_{\text{indep.}}$ : result of dice

$S_{10} = X_1 + X_2 + \cdots + X_{10}$ : sum of dice

$$E[S_{10}] = \sum_{j=1}^{10} \underbrace{\cancel{E[X_j]}}_{\frac{7}{2}} = 35$$

$$\mathrm{Var}(S_{10}) = \sum_{j=1}^{10} \underbrace{\cancel{\mathrm{Var}(X_j)}}_{\frac{35}{12}} = \frac{5(35)}{6} = \frac{175}{6}$$

worked out in an expl many weeks ago

$$S_{10} \overset{d}{\approx} 35 + \sqrt{\frac{175}{6}}\, Z \quad \leftarrow N(0,1)$$

$$= 35 + 5\sqrt{\frac{7}{6}}\, Z$$

$$P(30 \le S_{10} \le 40) \approx P\left(29.5 \le 35 + 5\sqrt{\frac{7}{6}}\, Z \le 40.5\right)$$

$$= P\left(-5.5 \le 5\sqrt{\frac{7}{6}}\, Z \le 5.5\right)$$

$$= P\left(5\sqrt{\frac{7}{6}}\,|Z| \le 5.5\right)$$

$$= P\left(|Z| \le \frac{11}{10}\sqrt{\frac{6}{7}}\right)$$

$$= 2\Phi\left(\frac{11}{10}\sqrt{\frac{6}{7}}\right) - 1 \approx 2\Phi(1.02) - 1$$

$$\approx 2(0.8461) - 1 = 1.6922 - 1 = \boxed{0.6922}$$

**Example 3d** Let $X_i, i = 1, \ldots, 10$, be independent random variables, each uniformly distributed over (0, 1). Calculate an approximation to $P\left\{\sum_{i=1}^{10} X_i > 6\right\}$.

$$S_{10} = \sum_{i=1}^{10} X_i$$

$$E[S_{10}] = \sum_{i=1}^{10} \underbrace{\cancel{E[X_i]}}_{\frac{1}{2}} = 5$$

$$Var(S_{10}) = \sum_{i=1}^{10} \underbrace{\cancel{Var(X_i)}}_{\frac{1}{12}} = \frac{5}{6}$$

cont. r.v., so don't need a continuity correction

$$S_{10} \overset{d}{\approx} 5 + \sqrt{\frac{5}{6}}\, Z \leftarrow N(0,1)$$

$$P(S_{10} > 6) \approx P\left(5 + \sqrt{\tfrac{5}{6}}\, Z > 6\right)$$

$$= P\left(Z > \sqrt{\tfrac{6}{5}}\right)$$

$$= 1 - \Phi\left(\sqrt{\tfrac{6}{5}}\right) \approx 1 - \Phi(1.10)$$

$$\approx 1 - 0.8643 = \boxed{0.1357}$$

**Example 3e** An instructor has 50 exams that will be graded in sequence. The times required to grade the 50 exams are independent, with a common distribution that has mean 20 minutes and standard deviation 4 minutes. Approximate the probability that the instructor will grade at least 25 of the exams in the first 450 minutes of work.

so variance is 16

$X_j$ = time to grade $j^{th}$ exam (min)

$E[X_j] = 20$, $Var(X_j) = 16$

$X_1, X_2, X_3, \ldots, X_{50}$ indep

$S_{25} = X_1 + X_2 + \cdots + X_{25}$ : time to grade $1^{st}$ 25 exams

$P(S_{25} \leq 450) \approx ?$

---

$E[S_{25}] = 25 \cdot 20 = 500$

$Var(S_{25}) = 25 \cdot 16 = 400$

$$S_{25} \overset{d}{\approx} 500 + \sqrt{\cancel{400}}^{\,20}\, Z \quad \leftarrow N(0,1)$$

cont. r.v.
no cont. corr.

$$P(S_{25} \le 450) \approx P(500 + 20Z \le 450)$$

$$= P\left(Z \le -\frac{5}{2}\right) = 1 - \Phi(2.5)$$

$$\approx 1 - 0.9938 = \boxed{0.0062}$$

HW: Ch. 8: 1, 2, 5–7, 9, 10, 13, 14–16