Convergence in Distribution

Let $X_1, X_2, \ldots$ be a sequence of random variables and let X be another random variable. We say that $\{X_n\}$ converges in distribution to X, if

$$\lim_{n\to\infty} F_{X_n}(x) = F_X(x) \tag{1}$$

at all points x on which F_X is continuous. This is denoted by

$$X_n \xrightarrow{D} X. \tag{2}$$

Note: The notation $X_n \xrightarrow{D} X$ might be a little misleading, since there is no pointwise convergence between the sequence X_n and X; what matter are their distributions. So we may write $X_n \xrightarrow{D} \mathsf{No}(0,1)$ or $F_n \xrightarrow{D} \mathsf{No}(0,1)$ as an abbreviated way of saying that "the distributions of X_n's or the distributions with CDFs F_n's converge to a standard normal distribution."

- $X_n \xrightarrow{P} b \iff X_n \xrightarrow{D} b$. Convergence in distribution is the same as convergence in probability, if the target is a constant.

  Let's show that $X_n \xrightarrow{P} b$ implies $X_n \xrightarrow{D} b$. Let $F_n(x)$ denote the CDF of X_n, and $F(x)$ denote the CDF of a constant b. Here a constant b stands for a discrete distribution with probability 1 to be b, so

  $$F(x) = 0, \quad \text{if } x < b; \quad F(x) = 1, \quad \text{if } x \geq b.$$

  We need to show:

  - For any fixed x value that is less than b, $F_n(x) \to 0$:
    $$F_n(x) = \mathbb{P}(X \leq x) = \mathbb{P}(|X - b| \geq |b - x|) \to 0;$$
  - For any fixed x value that is bigger than b, $F_n(x) \to 1$:
    $$F_n(x) = \mathbb{P}(X \leq x) = 1 - \mathbb{P}(X \geq x) = 1 - \mathbb{P}(|X - b| \geq |b - x|) \to 1.$$
  - When $x = b$, the limit of $F_n(b)$ may not equal 1 and the limit may not even exist, but it doesn't matter since convergence in distribution does not require the convergence of $F_n(b)$ since $F(x)$ is not continuous at b.

- $X_n \xrightarrow{P} X \implies X_n \xrightarrow{D} X$. When the target is a random variable X, convergence in distribution is usually not the same as convergence in probability; the latter is stronger.

  - Consider a standard normal rv Z. Define $X_n = (-1)^n \cdot Z$. Then X_n's and Z have exactly the same distribution, i.e., $X_n \xrightarrow{D} Z$ but X_n doesn't converge to Z in probability.

- Consider a standard normal rv Z. Define $X_n = -Z$. Then $X_n \xrightarrow{D} Z$, but X_n does not converge in probability to Z, instead $X_n \xrightarrow{P} -Z$.

- When talking about $X_n \xrightarrow{P} Z$, the distribution of Z need to be somehow related to X_n's; when talking about $X_n \xrightarrow{D} Z$, the symbol Z here just denotes a standard normal distribution, i.e., $X_n \xrightarrow{D} Z$ just means $X_n \xrightarrow{D} \mathsf{N}(0,1)$ (the marginal distribution of X_n's looks more and more like a standard normal when n gets large), so X_n's and the target distribution Z do not need to be dependent.

- <u>Example 5*</u> (Piazza Notes): Let

$$\mathbb{P}(X_n = \frac{i}{n}) = \frac{1}{n}, \quad i = 1, \ldots, n.$$

Show that $X_n \xrightarrow{D} \mathsf{Unif}(0,1)$.

- <u>Example 10(a)</u> (Piazza Notes): Suppose $U \sim \mathsf{Unif}(0,1)$. Let

$$X_n = \begin{cases} 1, & U \in \left(0, \frac{1}{3} + \frac{1}{n}\right) \\ 2, & U \in \left[\frac{1}{3} + \frac{1}{n}, \frac{2}{3} - \frac{1}{n}\right) \\ 3, & U \in \left[\frac{2}{3} - \frac{1}{n}, 1\right) \end{cases} \qquad X = \begin{cases} 1, & U \in \left(0, \frac{1}{3}\right) \\ 2, & U \in \left[\frac{1}{3}, \frac{2}{3}\right) \\ 3, & U \in \left[\frac{2}{3}, 1\right) \end{cases}$$

To evaluate $\mathbb{P}(|X_n - X| \geq \epsilon)$, let's first find out the joint distribution of (X_n, X), which is given by the following 3×3 table

X_n / X	1	2	3
1	$\left(\frac{1}{3}\right)$	$\left(\frac{1}{n}\right)$	0
2	0	$\left(\frac{1}{3} - \frac{2}{n}\right)$	0
3	0	$\left(\frac{1}{n}\right)$	$\left(\frac{1}{3}\right)$

It is easy to check that $\mathbb{P}(|X_n - X| \geq \epsilon) = \frac{2}{n}$ for any small ϵ. So $X_n \xrightarrow{P} X$.

- <u>Example 10(b)</u> (Piazza Notes): Suppose $U, U_1, U_2, \ldots$ iid $\sim \mathsf{Unif}(0,1)$. Let

$$X_n = \begin{cases} 1, & U_n \in \left(0, \frac{1}{3} + \frac{1}{n}\right) \\ 2, & U_n \in \left[\frac{1}{3} + \frac{1}{n}, \frac{2}{3} - \frac{1}{n}\right) \\ 3, & U_n \in \left[\frac{2}{3} - \frac{1}{n}, 1\right) \end{cases} \qquad X = \begin{cases} 1, & U \in \left(0, \frac{1}{3}\right) \\ 2, & U \in \left[\frac{1}{3}, \frac{2}{3}\right) \\ 3, & U \in \left[\frac{2}{3}, 1\right) \end{cases}$$

Then $X_n \xrightarrow{D} X$ but X_n does not converge in prob to X. This is because for a small ϵ, we have

$$\begin{aligned}\mathbb{P}(|X_n - X| \geq \epsilon) &= \mathbb{P}(X_n \neq X) = 1 - \mathbb{P}(X_n = X) \\ &= 1 - \Big[\Big(\frac{1}{3}+\frac{1}{n}\Big)\frac{1}{3} + \Big(\frac{1}{3}-\frac{2}{n}\Big)\frac{1}{3} + \Big(\frac{1}{3}+\frac{1}{n}\Big)\frac{1}{3}\Big] \\ &\to 1 - \frac{1}{3} = \frac{2}{3}.\end{aligned}$$

- Suppose $X_n \sim \mathsf{N}(0, n)$, do X_n's converge to some distribution? The CDF of X_n does converge to a constant function $F(x) = 0.5$:

$$F_n(x) = \mathbb{P}(X_n \leq x) = \mathbb{P}(X_n/\sqrt{n} \leq x/\sqrt{n}) = \Phi(x/\sqrt{n}) \to 0.5.$$

However, the function $F(x) = 0.5$ is not a valid CDF (Why). So X_n's do not converge in distribution.

Note that in the definition (1), we implicitly assume that the target function $F(x)$ is a valid CDF.

- Can we check $\xrightarrow{D}$ by checking the limit of the pdf f_n or the limit of the pmf p_n? Although the cdf is determined by the pdf/pmf

$$F_X(x) = \begin{cases} \int_{-\infty}^x f(u)du, & \text{if } X \text{ continuous;} \\ \sum_{u\leq x} p(u), & \text{if } X \text{ discrete,}\end{cases}$$

in general, we cannot determine the limiting distributions by considering the limits of pdfs/pmfs. Assume X_n's are continuous,

$$F_X(x) = \lim_n F_{X_n}(x) = \lim_n \int_{-\infty}^x f_n(u)du \quad \neq \quad \int_{-\infty}^x \lim f_n(u)du.$$

In fact, Fatou's Lemma states that $\neq$ should be $\geq$. That is, $\lim f_n$ only gives us a lower bound of $\lim F_n$. This explains why in the example ($X_n \sim \mathsf{N}(0, n)$), we loss 1/2 mass when considering $\lim_n f_n$: for example, assume any $x > 0$,

$$\begin{aligned}\lim_n F_{X_n}(x) = \lim_n \int_{-\infty}^x \frac{1}{\sqrt{2n}} e^{-u^2/n} du &= \frac{1}{2} + \lim_n \int_0^x \frac{1}{\sqrt{2n}} e^{-u^2/n} du \\ &\geq \frac{1}{2} + \int_0^x \lim_n \Big(\frac{1}{\sqrt{2n}} e^{-u^2/n}\Big) du = \frac{1}{2}.\end{aligned}$$

(*) But $\neq$ could be replaced by an equality, such as in Dominated Convergence Theorem where f_n's are upper bounded by some integrable function, i.e.,

$$|f_n| \leq g \quad \text{and} \quad \int |g(u)|du < \infty.$$

(*) Example (**4.3.3**): Let T_n have a t-distribution with n degrees of freedom with a pdf

$$f_n(t) = \frac{\Gamma(\frac{n+1}{2})}{\sqrt{\pi n}\Gamma(\frac{n}{2})} \frac{1}{(1+t^2/n)^{(n+1)/2}}.$$

$f_n(t) \le 10 f_1(t)$ and $10 f_1(t)$ is integrable, so

$$\lim_n f_n(t) = \lim_n \left\{ \frac{\Gamma(\frac{n+1}{2})}{\sqrt{\pi n}\Gamma(\frac{n}{2})} \right\} \lim_n \left\{ \frac{1}{(1+t^2/n)^{(n+1)/2}} \right\} = \frac{1}{\sqrt{2\pi}} e^{-t^2/2}$$

shows that $T_n \xrightarrow{D} \mathsf{No}(0,1)$.

- Convergence in the r-th mean ($r > 0$) if

$$\mathbb{E}|X_n - X|^r \to 0,$$

which implies $X_n \xrightarrow{P} X$. This can be shown by using Markov/Chebyshev inequality: for any $\epsilon > 0$,

$$\begin{aligned} \mathbb{E}|X_n - X|^r &= \int |x_n - x|^r f(x_n, x) dx_n dx \\ &\ge \int_{|x_n - x| \ge \epsilon} |x_n - x|^r f(x_n, x) dx_n dx \\ &\ge \int_{|x_n - x| \ge \epsilon} \epsilon^r f(x_n, x) dx_n dx \\ &= \epsilon^r \mathbb{P}(|X_n - X| \ge \epsilon) \to 0, \end{aligned}$$

since the upper bound (left side) goes to zero.

- Convergence in dist/prob is different from $\mathbb{E}|X_n|^r \to \mathbb{E}|X|^r$. See P10 in HW5.

- Recall that for real numbers $a_n \to a$, we have $g(a_n) \to g(a)$ if g is continuous at a. We can have the same result for convergence in dist/prob, which I refer to as the **Continuous Mapping Theorem**: if g is continuous in the range of X,

$$X_n \xrightarrow{P} X \Longrightarrow g(X_n) \xrightarrow{P} g(X);$$
$$X_n \xrightarrow{D} X \Longrightarrow g(X_n) \xrightarrow{D} g(X).$$

- If X_n converges and Y_n converges, what can we say about the convergence of (X_n, Y_n)? That is, does marginal convergence implies joint convergence? This is true for $\xrightarrow{P}$ but not true for $\xrightarrow{D}$ (apparently, their joint distribution isnt determined by the two marginals), unless X (or Y) is a constant (or the trivial case where the two sequences are totally independent).

- If $X_n \xrightarrow{P} X$ and $Y_n \xrightarrow{P} Y$, then $(X_n, Y_n) \xrightarrow{P} (X, Y)$ and consequently $g(X_n, Y_n) \xrightarrow{P} g(X, Y)$ for any continuous bivariate function g.

- If $X_n \xrightarrow{D} X$ and $Y_n \xrightarrow{D} c$, then $(X_n, Y_n) \xrightarrow{D} (X, c)$ and consequently $g(X_n, Y_n) \xrightarrow{D} g(X, c)$ for any continuous bivariate function g (**Slutsky Theorem**), e.g.,

$$X_n + Y_n \xrightarrow{D} X + c, \quad X_n Y_n \xrightarrow{D} cX, \quad Y_n^{-1} X_n \xrightarrow{D} c^{-1} X (c \neq 0).$$

Large Sample Properties of Estimators

For statistical inference such as estimation, we usually are only interested in convergence in distribution. Although *consistency* is defined via $\hat{\theta}_n \xrightarrow{P} \theta$, it's equivalent to $\hat{\theta}_n \xrightarrow{D} \theta$ since the target is a constant.

Most estimators we encounter in Stat410 are consistent estimators, i.e., we know

$$\hat{\theta}_n - \theta \xrightarrow{P} 0, \quad \text{or equivalently} \quad \hat{\theta}_n - \theta \xrightarrow{D} 0.$$

But we want to know *how fast* the difference $\hat{\theta}_n - \theta$ converges to zero. For example, both sequences, $a_n = \frac{1}{n}$ and $b_n = \frac{1}{\sqrt{n}}$, converge to zero, but $\{a_n\}$ goes to zero faster.

To find out the convergence speed, we scale up the difference $\hat{\theta}_n - \theta$ by an appropriate factor, e.g., $\sqrt{n}$, and show that $\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow{D} X$ then we can say that the convergence speed is $1/\sqrt{n}$, and for large sample size n, we can approximate $\hat{\theta}_n$ by $\frac{1}{\sqrt{n}} X + \theta$.

As I mentioned before, most estimators we encounter in Stat410 fall into the follow three categories, and we'll discuss how to show convergence in distribution for $n^r(\hat{\theta}_n - \theta)$ with an appropriate choice of r for each category.

1. $\hat{\theta} = \bar{X}$ or $\frac{1}{n}\sum_i \log X_i$, i.e., the sample mean of some iid random variables. Call CLT

$$\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow{D} \mathsf{N}(0, \sigma^2),$$

   where θ and σ^2 are the mean and variance of X_i (or $\log X_i$), respectively.

2. $\hat{\theta} = g(\bar{X})$, e.g., $\hat{\theta}_n = 1/\bar{X}$. Call CLT on $\bar{X}$, and then apply Δ-method.

3. $\hat{\theta}$ = order statistic, e.g., $Y_1 = \min_i X_i$ or $Y_n = \max_i X_i$. Then need to check the convergence of the CDF's F_n with an appropriate choice of n^r.

Limiting distributions of order statistics.

- Example 0 (Piazza Notes): $X_1, \ldots, X_n$ iid $\sim \mathsf{Unif}(0, \theta]$. Let $Y_n = \max_i X_i$ and we've known (from previous lectures) that $Y_n \xrightarrow{P} \theta$. For which values of r does

$$Z_n = n^r(\theta - Y_n)$$

converge in distribution? Note tha t $Z_n > 0$, so we know $F_n(z) = 0$, if $z < 0$. For any $z \geq 0$,

$$\begin{aligned} F_{Z_n}(z) &= \mathbb{P}\big(n^r(\theta - Y_n) \leq z\big) = \mathbb{P}\big(Y_n \geq \theta - \frac{z}{n^r}\big) \\ &= 1 - \Big(1 - \frac{z/\theta}{n^r}\Big)^n \to \begin{cases} 1, & \text{if } r < 1 \\ 1 - e^{-z/\theta}, & \text{if } r = 1 \\ 0, & \text{if } r > 1 \end{cases} \end{aligned}$$

So

  - if $r = 1$, $n(\theta - Y_n) \xrightarrow{D} \mathsf{Ex}(1/\theta)$, i.e., $\theta - Y_n$ goes to zero of order $1/n$;
  - if $r < 1$, then $n(\theta - Y_n) \xrightarrow{D} 0$;
  - if $r > 1$, then $n(\theta - Y_n)$ do not have a limiting distribution since the limit of F_n is not a valid CDF; they go to ∞ (i.e., over scale the difference).

- Similar to Example 1 (Piazza Notes): Let $\lambda > 0$ and let $X_1, \ldots, X_n$ be a random sample from a distribution with pdf

$$f(x; \lambda) = \lambda x^{-\lambda - 1}, \quad x > 1, \quad \text{zero elsewhere.}$$

  a) Let $Z_n = n(\min_i X_i - 1)$. Find the limiting distribution of of Z_n.

$$\begin{aligned} F_X(x) &= \int_1^x f(u; \lambda) du = 1 - x^{-\lambda} \\ F_{Z_n}(z) &= \mathbb{P}(Z_n \leq z) = \mathbb{P}(\min_i X_i \leq \frac{z}{n} + 1) = 1 - \mathbb{P}(\min_i X_i \geq \frac{z}{n} + 1) \\ &= 1 - \prod_{i=1}^n \mathbb{P}(X_i \geq \frac{z}{n} + 1) = 1 - \Big(\frac{z}{n} + 1\Big)^{-\lambda n} \to 1 - e^{-\lambda z}, \quad z > 0. \end{aligned}$$

  b) Let $W_n = \frac{\max_i X_i}{n^r}$. For which values of r does W_n converges in distribution? Find the limiting distribution.

$$\begin{aligned} F_{W_n}(w) &= \mathbb{P}(W_n \leq w) = \mathbb{P}(\max_i X_i \leq n^r w) \\ &= \prod_{i=1}^n \mathbb{P}(X_i \leq n^r w) = \Big(1 - \frac{1}{n^{r\lambda} w^\lambda}\Big)^n, \end{aligned}$$

which will converge to $w^{-\lambda}$ if $r\lambda = 1$ (or equivalently $r = 1/\lambda$):

$$\frac{\max_i X_i}{n^{1/\lambda}} \xrightarrow{D} W$$

where the cdf of W is $F_W(w) = e^{-w^{-\lambda}}$ and $w > 0$. For other values of r, you'll either get $W_n \xrightarrow{D} 0$ or W_n diverges to infinity (see P3 in HW5).

CLT and Δ-Method

An alternative definition of convergence in dist based on the mgf:

$$X_n \xrightarrow{D} X, \quad \text{if } M_{X_n}(t) \to M_X(t), \quad \text{for } |t| < h.$$

Examples

1. (Poisson approximation for Binomial) $X_n \sim \mathsf{Bin}(n, p_n = \lambda/n)$. Find the limiting distribution of X_n.

$$M_{X_n}(t) = \Big(1 - \frac{\lambda}{n} + \frac{\lambda}{n}e^t\Big)^n \to e^{\lambda(e^t-1)}.$$

So $X_n \xrightarrow{D} \mathsf{Po}(\lambda)$.

2. $X_n \sim \chi^2(n)$. Find the limiting distribution of $Y_n = X_n/n$.

$$M_{Y_n}(t) = \mathbb{E}(e^{tX_n/n}) = M_{X_n}(t/n) = \Big(1 - 2\frac{t}{n}\Big)^{-n/2} \to e^t.$$

So $Y_n \xrightarrow{D} Y$ where the mgf of Y is $M_Y(t) = \mathbb{E}e^{Yt} = e^t$. What's the distribution of Y?

3. $X_n \sim \chi^2(n)$. Find the limiting distribution of $Z_n = (X_n - n)/\sqrt{2n}$.

$$\begin{aligned}
M_{Z_n}(t) &= e^{-t\sqrt{n/2}} \times M_{X_n}(t/\sqrt{2n}) = e^{-t\sqrt{n/2}}\Big(1 - 2\frac{t}{\sqrt{2n}}\Big)^{-n/2} \\
&= \Big[e^{-t\sqrt{\frac{2}{n}}}\Big(1 - t\sqrt{\frac{2}{n}}\Big)\Big]^{-n/2} \\
e^{-t\sqrt{\frac{2}{n}}} &= 1 + t\sqrt{\frac{2}{n}} + \frac{t^2}{n} + o\Big(\frac{1}{n}\Big) \\
M_{Z_n}(t) &= \Big(1 - \frac{t^2}{n} + o\Big(\frac{1}{n}\Big)\Big)^{-n/2} \to e^{t^2/2},
\end{aligned}$$

when $|t|$ is small. So $Z_n \xrightarrow{D} \mathsf{No}(0, 1)$.

$$\begin{aligned} \text{Population} &: \quad \text{mean } \mu \quad \text{var } \sigma^2 \\ \text{Random Sample} &: \quad X_1, \cdots X_n \quad iid \\ \text{The sample mean} &: \quad \bar{X}_n = \frac{X_1 + \cdots + X_n}{n} \xrightarrow{P} \mu \quad \text{(WLLN)} \end{aligned}$$

(In fact, WLLN holds even if the variance σ^2 doesn't exist.)

What's the limiting distribution of $Z_n = \sqrt{n}(\bar{X}_n - \mu)/\sigma$? Without loss of generality, we can assume $\mu = 0$ in our derivation (why)

$$\begin{aligned} M_{Z_n}(t) &= \mathbb{E}\Big(e^{t(X_1+\cdots+X_n)/\sqrt{n}\sigma}\Big) = \Big[M_X(\frac{t}{\sqrt{n}\sigma})\Big]^n \\ M_X(\frac{t}{\sqrt{n}\sigma}) &= 1 + \frac{1}{2}\frac{t^2}{n} + o\Big(\frac{1}{n}\Big) \\ M_{Z_n}(t) &= \Big[1 + \frac{1}{2}\frac{t^2}{n} + o\Big(\frac{1}{n}\Big)\Big]^n \to e^{t^2/2} \end{aligned}$$

$$Z_n \xrightarrow{D} Z, \quad Z \sim \mathsf{No}(0,1)$$

Central Limit Theorem (CLT): $X_1, \ldots, X_n$ are iid with mean μ and variance σ^2.

$$\sqrt{n}(\bar{X}_n - \mu)/\sigma \xrightarrow{D} Z, \quad Z \sim \mathsf{No}(0,1).$$

That is, $\bar{X}_n \sim \mathsf{No}(\mu, \frac{\sigma^2}{n})$ approximately for large n.

Δ-Method (Theorem 4.3.9): $\sqrt{n}(X_n - \theta) \xrightarrow{D} \mathsf{No}(0, \sigma^2)$ and g is differentiable at θ and $g'(\theta) \neq 0$, then

$$\sqrt{n}(g(X_n) - g(\theta)) \xrightarrow{D} \mathsf{No}(0, \sigma^2[g'(\theta)^2]).$$

Intuition: by CLT,

$$X_n \sim \mathsf{No}(\theta, \frac{\sigma^2}{n}) \text{ approximately for large } n.$$

When x is close to θ,

$$g(x) \approx g(\theta) + g'(\theta)(x - \theta).$$

Therefore, if $g'(\theta) \neq 0$,

$$g(X_n) \sim \mathsf{No}(g(\theta), \frac{\sigma^2}{n}[g'(\theta)]^2) \text{ approximately for large } n.$$

Examples

1. Example 1.5 (Piazza Notes). Let $X_1, \ldots, X_n$ be a random sample from $\mathsf{Geo}(p)$, where $p > 0$. Find the limiting distribution[1] $\hat{p} = \tilde{p} = 1/\bar{X}$.

$$\mathrm{Var}(X) = (1-p)/p^2. \text{ By CLT, } \sqrt{n}(\bar{X} - 1/p) \xrightarrow{D} \mathsf{No}(0, \frac{1-p}{p^2}).$$

Since $g(x) = 1/x$ is differentiable at $1/p$ when $p > 0$ and $g'(1/p) = -p^2 \neq 0$,

$$\sqrt{n}\Big(g(\bar{X}) - g(1/p)\Big) \xrightarrow{D} \mathsf{No}\Big(0, (-p^2)^2 \frac{1-p}{p^2}\Big),$$

that is,

$$\sqrt{n}\big(\hat{p} - p\big) \xrightarrow{D} \mathsf{No}\big(0, p^2(1-p)\big).$$

2. Example 1 (Piazza Notes) Let $X_1, \ldots, X_n$ be a random sample from a distribution with pdf

$$f(x;\theta) = \frac{1}{\theta} x^{1/\theta - 1}, \quad 0 \le x \le 1, \quad \text{zero elsewhere } (\theta > 0).$$

a) Recall that the method of moments estimator

$$\tilde{\theta} = \frac{1 - \bar{X}}{\bar{X}} = \frac{1}{\bar{X}} - 1$$

is a consistent estimator of θ, since by WLLN, $\bar{X} \xrightarrow{P} \mu = \frac{1}{1+\theta}$ and $g(x) = \frac{1-x}{x}$ is continuous at $\frac{1}{1+\theta}$.

Find the limiting distribution of $\tilde{\theta}$.

$$\sigma^2 = \mathrm{Var}(X) = \mathbb{E}(X^2) - \mu^2 = \frac{\theta^2}{(1+2\theta)(1+\theta)^2}.$$

By CLT, $\sqrt{n}(\bar{X} - \mu) \xrightarrow{D} \mathsf{No}(0, \sigma^2)$. Since $g(x) = \frac{1-x}{x}$ is continuous at $\mu = \frac{1}{1+\theta}$ and $g'(\mu) = -(1+\theta)^2 \neq 0$,

$$\sqrt{n}(\tilde{\theta} - \theta) \xrightarrow{D} \mathsf{No}\Big(0, \frac{\theta^2(1+\theta)^2}{(1+2\theta)}\Big).$$

b) Recall that the MLE

$$\hat{\theta} = -\frac{1}{n}\sum_{i=1}^{n} \ln X_i$$

is a consistent estimator of θ, since by WLLN, $\hat{\theta} \xrightarrow{D} \mathbb{E}(\ln X) = -\theta$. Find the limiting distribution of $\hat{\theta}$.

$-\ln X_i \sim \mathsf{Ex}(1/\theta)$ with mean θ and variance θ^2. By CLT,

$$\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{D} \mathsf{No}(0, \theta^2).$$

[1] We know that $\hat{p}$ converges to a constant p. I use the term "the limiting distribution" to refer to a CLT type result, that is, $\sqrt{n}(\hat{p} - a) \sim \mathsf{No}(0, b)$ and we need to specify the values for parameters a and b.

c) Construct a $100(1-\alpha)\%$ confidence interval for θ.

Approach I: for large n, $\hat{\theta} \sim \mathsf{No}\Big(\theta, \frac{\theta^2}{n}\Big)$.

$$\begin{aligned} \mathbb{P}\Big(-z_{\alpha/2} \le \frac{\hat{\theta}-\theta}{\theta/\sqrt{n}} \le z_{\alpha/2}\Big) &= 1-\alpha \\ \mathbb{P}\Big(\frac{\hat{\theta}}{1+\frac{z_{\alpha/2}}{\sqrt{n}}} \le \theta \le \frac{\hat{\theta}}{1-\frac{z_{\alpha/2}}{\sqrt{n}}}\Big) &= 1-\alpha \\ \Big(\frac{\hat{\theta}}{1+\frac{z_{\alpha/2}}{\sqrt{n}}}, \frac{\hat{\theta}}{1-\frac{z_{\alpha/2}}{\sqrt{n}}}\Big) &: \text{ approximate } 100(1-\alpha)\% \text{ CI for } \theta. \end{aligned}$$

Approach II (use plug-in estimates for the variance): Since $\sqrt{n}(\hat{\theta}-\theta)/\theta \xrightarrow{D} \mathsf{No}(0,1)$ and $\hat{\theta} \xrightarrow{P} \theta$, by Slutsky's Thm, we have

$$\sqrt{n}(\hat{\theta}-\theta)/\hat{\theta} \xrightarrow{D} \mathsf{No}(0,1).$$

So $\hat{\theta} \pm z_{\alpha/2}\frac{\hat{\theta}}{\sqrt{n}}$ is an approximate $100(1-\alpha)\%$ CI for θ.

Similarly we can obtain the approximate CI based on $\tilde{\theta}$,

$$\tilde{\theta} \pm z_{\alpha/2}\frac{\tilde{\theta}(1+\tilde{\theta})}{\sqrt{(1+2\tilde{\theta})n}}.$$

Application of the Slutsky's Thm: Suppose $\sqrt{n}\frac{1}{\sigma}(\hat{\theta}_n-\theta) \xrightarrow{D} \mathrm{N}(0,1)$ and suppose the variance $\sigma^2 = \sigma^2(\theta)$ depends on the unknown parameter θ. Let $\hat{\sigma}^2 = \sigma^2(\hat{\theta})$ be the plug-estimate of the variance, i.e., we replace θ in the expression of the variance by its estimate $\hat{\theta}$. We can use Slutsky Thm to show that

$$\sqrt{n}\frac{1}{\hat{\sigma}}(\hat{\theta}_n-\theta) \xrightarrow{D} \mathrm{N}(0,1).$$

This is because

$$\sqrt{n}\frac{1}{\hat{\sigma}}(\hat{\theta}_n-\theta) = \Big(\frac{\sigma}{\hat{\sigma}}\Big)\sqrt{n}\frac{1}{\sigma}(\hat{\theta}_n-\theta) = Y_n X_n,$$

and

$$X_n = \sqrt{n}\frac{1}{\sigma}(\hat{\theta}_n-\theta) \xrightarrow{D} \mathrm{N}(0,1), \text{ and } Y_n = \frac{\sigma}{\hat{\sigma}} \xrightarrow{P} 1, \Longrightarrow X_nY_n \xrightarrow{D} \mathrm{N}(0,1).$$

The convergence in prob of Y_n holds if $\hat{\theta}$ is consistent and $\sigma(\theta)$ is a continuous function of θ.